1. Candidate’s declarations:

I, Calum Brown, hereby certify that this thesis, which is approximately 38,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in October 2008 and as a candidate for the degree of Ph.D. in October 2008; the higher study for which this is a record was carried out in the University of St Andrews between 2008 and 2012.

Date ...... signature of candidate ..........

2. Supervisor’s declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ...... signature of supervisor ..........

3. Permission for electronic publication: (to be signed by both candidate and supervisor)

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

(iii) Embargo on both all of printed copy and electronic copy for the same fixed period of 2 years (maximum five) on the following ground(s):

publication would preclude future publication;

Date ...... signature of candidate ...... signature of supervisor ..........
Abstract

The use of spatial statistics to investigate ecological processes in plant communities is becoming increasingly widespread. In diverse communities such as tropical rainforests, analysis of spatial structure may help to unravel the various processes that act and interact to maintain high levels of diversity. In particular, a number of contrasting mechanisms have been suggested to explain species coexistence, and these differ greatly in their practical implications for the ecology and conservation of tropical forests. Traditional first-order measures of community structure have proved unable to distinguish these mechanisms in practice, but statistics that describe spatial structure may be able to do so. This is of great interest and relevance as spatially explicit data become available for a range of ecological communities and analysis methods for these data become more accessible.

This thesis investigates the potential for inference about underlying ecological processes in plant communities using spatial statistics. Current methodologies for spatial analysis are reviewed and extended, and are used to characterise the spatial signals of the principal theorised mechanisms of coexistence. The sensitivity of a range of spatial statistics to these signals is assessed, and the strength of such signals in natural communities is investigated.

The spatial signals of the processes considered here are found to be strong and robust to modelled stochastic variation. Several new and existing spatial statistics are found to be sensitive to these signals, and offer great promise for inference about underlying processes from empirical data. The relative strengths of particular processes are found to vary between natural communities, with any one theory being insufficient to explain observed patterns. This thesis extends both understanding of species coexistence in diverse plant communities and the methodology for assessing underlying process in particular cases. It demonstrates that the potential of spatial statistics in ecology is great and largely unexplored.
I owe thanks to many people for help and support during my Ph.D. Firstly my supervisors, Janine Illian, David Burslem and Richard Law, who gave me the fantastic opportunity to work on this project and offered me encouragement and assistance throughout. I have learnt a great deal from them and am very grateful indeed to them all. Special thanks go to Janine for her tremendously encouraging, if at times apparently unfounded, belief in my ability to complete this thesis. This work would not have been possible without generous funding and support from Microsoft Research and, latterly, from the Centre for Research into Ecological and Environmental Modelling. This also enabled me to attend several conferences at which I benefitted from the advice of many statisticians and ecologists. My supervisor at Microsoft, Drew Purves, also provided inspiration and guidance.

Thanks are due to all the Principal Investigators of rainforest plots who collaborated with me, and especially to I Fang Sun, Yu-Wen Pan and other staff and students at Tunghai University who made my visit to Taiwan so useful and memorable. I’m also grateful for technical support from a number of people, notably Tony Travis at the Aberdeen Rowett Institute of Nutrition and Health, Herbert Fruchtl in St Andrews and, especially, Phil LeFeuvre at CREEM who was always willing and able to solve any problem at a moment’s notice. Thanks to Rhona Rodger too for calmly handling a great diversity of problems which had mystified me.

I benefitted enormously from the friendly and stimulating atmosphere at CREEM and am grateful to all its members for creating such an ideal working environment - and for the plentiful supply of cake that made the period between coffee and lunch times so productive. Angelika Studeny was extremely generous with her time and mathematical expertise throughout, and provided numerous useful critiques, ideas and discussions. Glenna Evans and Cornelia Oedekoven have been consistently supportive and cheerful companions on the PhD rollercoaster, and Bruno Caneco a dependable source of much-needed amusement, advice and music.

Finally, I owe a huge debt of gratitude to my whole family for their support, kindness and interest in this work. Heartfelt thanks go to my father, Ken, for never letting me forget the difficult and important questions behind it all (or, indeed, some less obviously relevant ones), my mother, Maureen, for her steadfast encouragement and optimism, and my brother, Evan, for providing the most reliable and generous technical support possible. And to Penelope, my wife, for being an example for me to follow, for her patience when my Ph.D. threatened to overshadow the more important things in life, for escaping to the hills with me, and for her gentle and unwavering confidence in me.
# Table of Contents

Abstract .................................................................................................................................................. ii
Acknowledgements ................................................................................................................................... iv
Table of contents ..................................................................................................................................... vi
List of figures ........................................................................................................................................... ix
List of tables .......................................................................................................................................... xii

## Chapter 1 - Introduction ......................................................................................................................... 1
1.1 Species coexistence – history ......................................................................................................... 2
1.2 Niche theory ....................................................................................................................................... 3
1.3 Lottery models .................................................................................................................................. 3
1.4 Janzen-Connell effect ....................................................................................................................... 4
1.5 Heteromyopia ................................................................................................................................... 4
1.6 Neutral theory ................................................................................................................................... 5
1.7 Model tests and comparisons: first-order ......................................................................................... 5
1.8 Model tests and comparisons: second-order .................................................................................... 6
1.9 Thesis plan ......................................................................................................................................... 8

## Chapter 2 - Linking ecological processes with spatial and non-spatial patterns in plant communities ........................................................................................................................................ 9
2.1 Introduction ....................................................................................................................................... 10
2.2 Materials and methods ..................................................................................................................... 11
  2.2.1 STOCHASTIC PROCESS FOR MULTISPECIES SPATIAL PATTERNS ........................................ 11
  2.2.2 ECOLOGICAL MODELS FOR SPECIES COEXISTENCE .............................................................. 12
    2.2.2.1 Neutral model ......................................................................................................................... 12
    2.2.2.2 Spatial niche model ................................................................................................................ 12
    2.2.2.3 Lottery (temporal niche) model ............................................................................................... 13
    2.2.2.4 Janzen–Connell model .......................................................................................................... 13
    2.2.2.5 Heteromyopia model .......................................................................................................... 13
  2.2.3 NUMERICAL VALUES FOR SIMULATIONS ............................................................................... 14
  2.2.4 CROSS-PAIR OVERLAP DISTRIBUTION (xPOD) .................................................................... 15
  2.2.5 COMMUNITY EVENNESS AND DIVERSITY ............................................................................. 15
  2.2.6 STATISTICAL INFERENCE ......................................................................................................... 16
2.3 Results .............................................................................................................................................. 18
  2.3.1 ILLUSTRATIVE EXAMPLE ........................................................................................................ 18
  2.3.2 SPECIES ABUNDANCE DISTRIBUTIONS ................................................................................. 18
  2.3.3 CROSS-PAIR OVERLAP DISTRIBUTIONS ............................................................................. 19
2.4 Discussion .................................................................................................................. 26
2.5 Appendix 2.1: Definition of the pair correlation function ........................................ 29
2.6 Appendix 2.2: Sensitivity analysis of variation in dispersal kernels .......................... 30
2.6.1 RESULTS & DISCUSSION ....................................................................................... 30
Chapter 3 - Multispecies coexistence in tropical forests: spatial signals of topographic niche differentiation increase with environmental heterogeneity ........................................ 32
3.1 Introduction ............................................................................................................... 33
3.2 Methods .................................................................................................................... 34
3.2.1 DATA .................................................................................................................... 34
3.2.2 SPATIAL STATISTIC (CROSS-PAIR OVERLAP DISTRIBUTION) ..................... 38
3.2.3 ENVIRONMENTAL METRICS AND REGRESSION ............................................. 40
3.3 Results ...................................................................................................................... 41
3.4 Discussion ............................................................................................................... 44
3.5 Appendix 3.1: Calculation of cross-pair overlap distribution (xPOD) ......................... 47
3.6 Appendix 3.2: Additional plots ............................................................................... 48
Chapter 4 - Spatial statistics for the inference of underlying process in plant communities ........................................ 55
4.1 Introduction ............................................................................................................... 56
4.1.1 SPATIAL STATISTICS ......................................................................................... 56
4.1.2 SPATIAL STRUCTURE .......................................................................................... 57
4.2 Methods .................................................................................................................... 58
4.2.1 SPATIAL STATISTICS ......................................................................................... 58
4.2.1.1 Scattering ........................................................................................................... 58
4.2.1.2 Exposure .......................................................................................................... 58
4.2.2 ECOLOGICAL SIMULATIONS .............................................................................. 63
4.2.3 COMPARING SPATIAL STATISTICS .................................................................... 63
4.3 Results ...................................................................................................................... 64
4.3.1 SPECIES LEVEL .................................................................................................. 64
4.3.1.1 Scattering ......................................................................................................... 65
4.3.1.2 Exposure ......................................................................................................... 72
4.3.2 COMMUNITY LEVEL .......................................................................................... 87
4.3.2.1 Scattering ......................................................................................................... 87
4.3.2.2 Exposure ......................................................................................................... 92
4.4 Discussion ............................................................................................................... 102
4.4.1 SCATTERING ...................................................................................................... 102
4.4.2 EXPOSURE ......................................................................................................... 103
4.4.3 LEVEL OF INFORMATION .................................................................................... 106
4.4.4 GENERAL ........................................................................................................... 106
4.5 Conclusion .............................................................................................................. 109
List of figures

Figure 2.1. Realization of a spatial multispecies birth–death process, showing the statistics calculated on the multispecies spatial pattern at the end. ................................................................. 17

Figure 2.2. Species abundance distributions (SADs) obtained from different ecological models. ...... 20

Figure 2.3. Boxplots of Kolmogorov–Smirnov test statistics. ................................................................. 21

Figure 2.4. Cross pair overlap distributions (xPODs) obtained from different ecological models. ..... 23

Figure 2.5. Boxplots of Kolmogorov–Smirnov test statistics obtained from comparing pairs of cross pair overlap distributions within and between ecological models. ................................. 25

Figure A2.2.1. Cross pair overlap distributions (xPODs) obtained from neutral and niche ecological models with a range of dispersal kernels. ............................................................................. 31

Figure 3.1: The contribution of species pairs to the xPOD for Lienhuachih. ................................. 39

Figure 3.2: Topographies and cross-pair overlap distributions for Pasoh and Gutianshan. ............... 42

Figure 3.3: Significant relationships between cross-pair overlap distribution standard deviations and environmental metrics................................................................................................................. 43

Figure A3.1: Topographies and cross-pair overlap distributions for all plots. ........................................ 49

Figure A3.2: Relationships between xPOD standard deviations and elevation range in each plot...... 50

Figure A3.3: Relationships, fitted regression lines and their associated p-values for each of the environmental metrics against standard deviation of cross-pair overlap distributions. ......................... 51

Figure A3.4: Standard deviations of cross-pair overlap distributions against biogeographical variables............................................................................................................................................. 53

Figure A3.5: The chance relationship between latitude and elevation range of plots.......................... 53

Figure 4.1: The class and level to which each piece of information used in the construction of the measures belongs. ......................................................................................................................... 62

Figure 4.2: Extremes of scattering and exposure. .................................................................................. 63

Figure 4.3: Measure of interspecific segregation results for each model. ............................................. 65

Figure 4.4a: The degree of aggregation for species in each model......................................................... 66

Figure 4.4b: The degree of aggregation for species in each model......................................................... 67

Figure 4c: Histograms of values of the degree of aggregation for species in each model..................... 68

Figure 4.5a: The proportion of conspecific neighbours for species with abundances of 500 or more in each model......................................................................................................................... 69

Figure 4.5b: The proportion of conspecific neighbours for species in each model............................. 70
Figure 4.5c: Histograms of values of the proportion of conspecific neighbours for species in each model. ................................................................. 71

Figure 4.6: Cross-pair overlap distributions for species in each model. ................................................................. 72

Figure 4.7a: ISAR curves for species with abundances of 500 or more in each model expressed in terms of species numbers and proportion of total community size. .............................................. 74

Figure 4.7b: ISAR result boxplots for species in each model. ................................................................. 75

Figure 4.7c: Histograms of ISAR values for species in each model. ................................................................. 76

Figure 4.8a: Mingling index for species in each model. ............................................................................. 78

Figure 4.8b: Mingling index boxplots for species in each model. ................................................................. 79

Figure 4.8c: Histograms of mingling index values for species in each model. ................................................................. 80

Figure 4.9a: The degree of association for species in each model, at the level of species pairs. ........... 81

Figure 4.9b: Degree of association boxplots for species pairs in each model. ................................................................. 82

Figure 4.9c: Histograms of degree of association values for species pairs in each model. ................................................................. 83

Figure 4.10a: The degree of association for species in each model, at species level. ....................... 84

Figure 4.10b: Degree of association boxplots for species in each model. ................................................................. 85

Figure 4.10c: Histograms of degree of association values for species in each model. ................................................................. 86

Figure 4.11a: Prediction intervals (95%) for the degree of aggregation (summed) at community level. ............................................................................. 88

Figure 4.11b: Boxplots of degree of aggregation (summed) values. ................................................................. 88

Figure 4.12a: Prediction intervals (95%) for the mean degree of aggregation at community level. .... 89

Figure 4.12b: Boxplots of degree of aggregation mean values. ................................................................. 89

Figure 4.13: Boxplots of mean values of the measure of interspecific segregation for each model. .. 90

Figure 4.14a: Prediction intervals (95%) for the mean proportion of conspecific neighbours at community level. ............................................................................. 91

Figure 4.14b: Boxplots of mean values of the proportion of conspecific neighbours for each model. 91

Figure 4.15a: Prediction intervals (95%) for the spatial Simpson index. ................................................................. 92

Figure 4.15b: Boxplots of spatial Simpson indices for each model. ................................................................. 93

Figure 4.16a: Prediction intervals (95%) for the spatial Simpson index. ................................................................. 93

Figure 4.16b: Boxplots of spatial Simpson indices. ............................................................................. 94
Figure 4.17a: Prediction intervals (95%) for the cross-pair overlap distribution. ........................................95

Figure 4.17b: Boxplots of means and standard deviations of the cross-pair overlap distribution ...... 96

Figure 4.18a: Prediction intervals (95%) for the mean ISAR at community level, as a proportion of total community size.......................................................... 97

Figure 4.18b: Boxplots of mean values of the ISAR, as a proportion of community size ....................... 98

Figure 4.19a: Prediction intervals (95%) for the mean mingling index at community level............... 99

Figure 4.19b: Boxplots of mean values of the mingling index for each model. ........................................ 100

Figure 4.20a: Prediction intervals (95%) for the mean degree of association at community level.... 101

Figure 4.20b: Boxplots of mean values of the degree of association for each model......................... 101

Figure 4.21: A 3-D plot of results from each model based on mean (community-level) values of the mingling index, proportion of conspecific neighbours, and standard deviation of the xPOD.......... 108

Figure A4.1: The labelling of cells or quadrats used for the calculation of measures of beta-diversity. ........................................................................................................ 112

Figure A4.2: Example results for measures of beta-diversity, typical of the three types of curve produced by the 21 measures considered. ................................................................. 113

Figure A4.3: Boxplots of estimated values of parameter alpha in the power law which best fits results for 3 measures of beta diversity in each model. ..................................................... 114

Figure 5.1: Model selection criteria against model complexity.......................................................... 123

Figure 5.2: Training and test data likelihoods for the five models with the highest test data maximum likelihoods, with and without hierarchies. ................................................................. 124

Figure 5.3: Posterior 95% credible intervals for the species-specific intrinsic mortality parameter k0 with and without a hierarchy............................................................................. 124

Figure 5.4: Significant (non-zero) posterior 95% credible intervals for the neighbour diameter parameter. ............................................................................................................ 126

Figure 5.5 Significant (non-zero) posterior 95% credible intervals for the neighbour phylogenetic relatedness parameter................................................................................................ 126

Figure 5.6: Diameter (dbh) ranges for species with significant phylogenetic and neighbour-diameter effects. ............................................................................................................. 129

Figure 5.7: The range of Clark & Evans aggregation index values produced by species with significant phylogenetic and neighbour-diameter effects. ........................................................ 130

Figure A5.1: Test data likelihood values for the three most successful models:................................. 132

Figure A5.2: The form of the relationship between tree diameter at breast height (dbh) and mortality, for a dbh range of 1 – 50 cm. ...................................................................................... 132
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1.</td>
<td>Parameters for ecological models</td>
<td>14</td>
</tr>
<tr>
<td>Table 2.2.</td>
<td>Mean evenness and diversity of each model</td>
<td>19</td>
</tr>
<tr>
<td>Table 2.3.</td>
<td>Mean and 90% limit values of the Kolmogorov–Smirnov test statistic comparing pairs of SADs within and between contrasting models of ecological interactions, using species of all abundances</td>
<td>22</td>
</tr>
<tr>
<td>Table 2.4.</td>
<td>Mean and 90% limit values of the Kolmogorov–Smirnov test statistic comparing pairs of SADs within and between contrasting models of ecological interactions, using an abundance threshold of 500 individuals</td>
<td>22</td>
</tr>
<tr>
<td>Table 2.5.</td>
<td>Mean, standard deviations and skewness of cross pair overlap distributions for contrasting models of ecological interactions</td>
<td>24</td>
</tr>
<tr>
<td>Table 2.6.</td>
<td>Mean and 90% limit values of the Kolmogorov–Smirnov test statistic comparing pairs of xPODs within and between contrasting models of ecological interactions</td>
<td>24</td>
</tr>
<tr>
<td>Table A2.2.1.</td>
<td>Parameters for ecological models</td>
<td>30</td>
</tr>
<tr>
<td>Table 3.1.</td>
<td>Summaries of the plots included in the analyses</td>
<td>36</td>
</tr>
<tr>
<td>Table 3.2.</td>
<td>The names, identities and characteristics described by the environmental metrics calculated for each plot</td>
<td>41</td>
</tr>
<tr>
<td>Table A3.1.</td>
<td>p-values for single linear regressions of environmental metrics against moments of species abundance distributions (SAD) and the Shannon diversity index for each plot</td>
<td>54</td>
</tr>
<tr>
<td>Table 4.1.</td>
<td>Separate pieces of information used in spatial measures</td>
<td>60</td>
</tr>
<tr>
<td>Table 4.2a.</td>
<td>Definitions of measures of scattering</td>
<td>60</td>
</tr>
<tr>
<td>Table 4.2b.</td>
<td>Definitions of measures of exposure</td>
<td>61</td>
</tr>
<tr>
<td>Table 4.3.</td>
<td>Means, standard deviations and skews of the degree of aggregation for species in each model with an abundance threshold of 500</td>
<td>68</td>
</tr>
<tr>
<td>Table 4.4.</td>
<td>Means, standard deviations and skews of the proportion of conspecific neighbours for species in each model with an abundance threshold of 500</td>
<td>71</td>
</tr>
<tr>
<td>Table 4.5.</td>
<td>Means, standard deviations and skews of xPODs for species in each model, with an abundance threshold of 500</td>
<td>73</td>
</tr>
<tr>
<td>Table 4.6.</td>
<td>Means, standard deviations and skews of ISAR values for species in each model with an abundance threshold of 500</td>
<td>76</td>
</tr>
<tr>
<td>Table 4.7.</td>
<td>Means, standard deviations and skews of mingling index results for species in each model, with an abundance threshold of 500</td>
<td>80</td>
</tr>
<tr>
<td>Table 4.8.</td>
<td>Means, standard deviations and skews of degree of association results for species pairs in each model, with an abundance threshold of 500</td>
<td>83</td>
</tr>
</tbody>
</table>
Table 4.9: Means, standard deviations and skews of degree of association results for species in each model, with an abundance threshold of 500 ................................................................. 86

Table 4.10: Principal characteristics of spatial patterns generated under each model in terms of scattering, exposure, and implications for turnover of species with distance (Beta-diversity). ........ 105

Table A4.1: Definitions of measures of β-diversity ................................................................. 111

Table 5.1: The identities and allowed ranges of all model parameters ........................................ 120
CHAPTER 1 - Introduction
One of the major goals of modern conservation is the preservation of biological diversity, the variation in life at genetic, organism and species levels, at the global scale (Magurran 2004; IUCN 2011). Despite being a priority in national and international policy around the world (e.g. Scottish Executive 2004; JNCC 2011; UNEP 2012), this objective is perhaps most closely associated with the conservation of tropical rainforests. Estimated to contain more than half of Earth’s species of land plants and animals, these forests are under intense pressure from logging, hunting and the conversion of land for pastoral and arable agriculture (Wilson 1985; Skole & Tucker 1993; Whitmore 1998; Primack & Corlett 2005). Attempts to stem the ongoing loss of rainforests to anthropogenic activities like these are many and varied, but share a focus on the protection of relatively pristine areas intended to act as reservoirs of diversity (Turner 1996; Myers et al. 2011). Beyond this broad and large-scale approach, conservation efforts are supported by relatively little ecological theory or empirical evidence. In particular, attempts to maintain the extraordinary levels of species diversity found in the tropics are hampered by a fundamental lack of understanding of the processes responsible for generating and supporting this diversity (Connell 1978; Chesson 2000; Burslem et al. 2005). Despite having been an important focus of study throughout the history of ecology, species coexistence in tropical rainforests remains largely unexplained.

1.1 Species coexistence – history

Mechanisms that support the coexistence of species have been the subject of study throughout the history of ecology, and many modern explanations were anticipated by early work on this subject. Darwin famously warned against the interpretation of an ‘entangled bank’ of vegetation as a randomly assembled group of species, suggesting instead that it was the product of deterministic processes through which each species had evolved a unique role within the community (Darwin, 1859). Until a fuller understanding of genetics came to underpin evolutionary theory, however, explanations focused on the physiological ability of plants to live in particular environmental conditions. Eugenius Warming, another of ecology’s founding figures, influentially argued that both current and previous conditions dictated species’ observed distributions (Warming 1895). The greater the ability of plants to adapt to prevailing conditions, the more important these became for shaping species distributions (Warming 1895; Collins et al. 1986).

The rediscovery of Gregor Mendel’s genetic experiments (Mendel 1865) dispelled much of the scepticism about evolution by natural selection and suggested that adaptation through genetic variation was a general, rapid and powerful process (Osborn 1926). Darwin’s earlier portrayal of a community as a tightly packed matrix of ‘wedges’, representing species with different environmental requirements that shape and displace one another, also received support. If genetic variation allowed organisms to evolve according to environmental pressures, they could equally respond to the pressures of other, competing, members of their community (Darwin 1859, Kingsland 1985). Coexistence was therefore a matter of both the environmental and interactive ‘fit’ of species (von Humboldt & Bonpland, 1805).

This suggested that closely-related (and therefore ecologically similar) species would compete most strongly, a process that would drive the emergence of new adaptations (Cavender-Bares et al. 2009). Warming found evidence for this in his observations, where the forms and habitats of confamilials were often divergent, but convergent in heterofamilials (Warming 1895). In both cases, selection pressures determined by local environmental conditions seemed to be shaping community structure (Collins et al. 1986).
1.2 Niche theory

The work of Darwin, Warming and others led to the development of niche-assembly theories, in which coexistence was explained by species’ differing specialisations (initially in terms of habitat). Prominent among the early developers of this theory was Grinnell (e.g. 1917), who expressed a general consensus that “each animal occupies a definite area, that is, has a habitat or range, which is distinctive enough to be included among the characters of the species and described along with its habits and the features of its bodily structure” (p115). This environmental, spatial view of niches was extended by others (e.g. Elton, 1927) to include the functional role of each species, and so was defined with reference to the interactions within a community. As a result, competition increasingly became viewed as central to the niche concept. Gause’s (1934) competitive exclusion principle, building on the theories of Darwin (1859), Grinnell (1904) and, especially, Volterra (1926) and Lotka (1932), formalised this and suggested that coexistence of species required each to have its own unique environmental and interactive niche.

This niche concept was typified by Hutchinson (1958) who expressed the purely environmental, abiotic domain of a species as its ‘fundamental niche’ and its actual range, constrained by interactive, biotic factors as its ‘realised niche’. Each represented a hypervolume with dimensions determined by environmental variables and interactions, and each hypervolume was distinct, even if only because of competition within identical fundamental niches. This was supported by subsequent research into the stabilising effect of predation on competitive interactions that would otherwise result in the exclusion of species (Slobodkin 1961, 1964). Diamond’s (1975) community assembly rules held that a community was formed when species are “coadjusted in their niches and abundances, so as to fit with each other and resist invaders” (p343).

Empirical tests of this form of niche theory faced two difficulties. Firstly, Hutchinson’s formulation predicted that even where species appeared to share identical niches, they would in fact differ in some unobserved – and potentially unobservable – way. Secondly, the fundamental niche was by its nature unobservable except in the absence of biotic interactions, and therefore effectively impossible to investigate, confirm or disprove (Andrewartha 1958). These problems were particularly substantial because competition between species was not often apparent. Hutchinson himself noted the “extreme difficulty of identifying competition as a process actually occurring in nature” (Hutchinson, 1958, p418), and that competitive exclusion may be circumvented by migration and random or cyclical disturbances (also noted by e.g. Gleason 1939). As a result, the established view of community structure as an expression of deterministic niche assembly processes began to be questioned (e.g. Gleason 1939; Williams 1964).

1.3 Lottery models

The idea of niche-structured communities existing in equilibrium was challenged not only because of the lack of evidence of interactions necessary to maintain equilibrium, but also because communities themselves rarely appeared stable. Fluctuations in species abundances or community composition were regularly observed under environmental change, and the frequency with which such changes occurred suggested that they might play a fundamental role in their own right (Ayala 1972; Grubb 1977; Wiens 1977; Levins 1979; Pickett 1980; Chesson 1981). Sale (1977) and Connell (1978) were among the first to conclude that environmental disturbances (caused for example by climate cycles) could be credited with maintaining tropical diversity, arguing that they prevented a relatively depauperate equilibrium arrangement from developing.
This argument formed the basis of lottery models, which assume that temporal environmental variance benefits different species at different times through niche availability. The first of these models was developed by Sale (1977) and quickly extended by others (e.g. Chesson & Warner, 1981; Chesson & Huntly, 1988). They attracted attention in part because the lottery model was the only coexistence theory to include environmental stochasticity over time (Hatfield et al. 1996). It also did not depend upon small and cryptic differences between niches as the established niche-differentiation theory did, suggesting only that species differed in their response to environmental change. Even this, of course, depended upon the existence of numerous distinct niches, and so lottery models did still represent coexistence purely as a function of species-specific environmental specialisation.

1.4 Janzen-Connell effect

The difficulties associated with the interactive components of niche theory did not lead only to a focus on environmental effects. Deterministic models of species interactions continued to develop, and suggested that competition could indeed play a central role in structuring communities (Connell 1975; Sugihara 1980; Engen & Lande 1996a). Particularly influential was the work of Janzen (1970) and Connell (1970), who independently sought to explain observed low densities and unexpectedly uniform dispersions of tropical tree species. They invoked species-specific herbivory to produce overcompensating density-dependent mortality, in which young plants are unable to survive in the immediate vicinity of their parents. In addition to explaining observed spatial structure, this effect would prevent dominance by a small number of common species, promote spatial mixing of species and, crucially, support species coexistence (Connell et al. 1984).

Considerable empirical support was found for the Janzen-Connell effect. High local densities were shown to affect growth and mortality in tropical trees (Bella 1971; Diggle 1976), and additional agents of conspecific density-dependent mortality were identified in pests and pathogens (e.g. Augspurger 1983). It was expected that common species would suffer disproportionately because of the large populations of pests, predators and pathogens that they could support (Ridley 1930), and experimental studies that manipulated densities of isolated species lent support to this (e.g. Connell 1970; Augspurger 1983; Schupp 1992; Bagchi et al. 2010). On larger scales, however, evidence for Janzen-Connell effects is not compelling, with the majority of species showing no evidence of them (e.g. Hyatt et al. 2003). This has called into question the theory’s ability to explain coexistence in highly diverse communities (Lambers et al. 2002).

1.5 Heteromyopia

A recently-proposed variant of the Janzen-Connell effect, heteromyopia describes the tendency of conspecifics to interact over greater distances than heterospecifics (Murrell & Law 2003). This is relevant to coexistence because it suggests that large-scale rather than local density may affect species’ mortality rates. Depending upon the dispersal abilities of species-specific pests, predators and pathogens, this may be able to more accurately describe the previously noted disproportionate mortality expected in common species and the community compensatory trends that result (Ridley 1930; Connell et al. 1984; Amarasekare 2003; Queenborough et al. 2007).

Heteromyopia is harder to test than the Janzen-Connell effect because the scale-dependent behaviour it describes is less suited to experimental manipulation. As a result, little empirical evidence currently exists for it. However, studies that consider large- or even community-scale species abundances in determining mortality rates are relevant. These are, again, inconclusive, with
compensatory trends consistent with heteromyopia being identified in some cases (Okuda et al. 1997; Webb & Peart 1999; Queenborough et al. 2007) but not in many others (He et al. 1997; Comita et al. 2010; Metz et al. 2010). The concept’s success in describing real community dynamics and explaining species coexistence is therefore uncertain.

1.6 Neutral theory

The difficulties in providing adequate explanations of coexistence using the established theories outlined above prompted the development of neutral theories intended to act as null models of coexistence (Caswell 1976; Hubbell 1979, 2001; Bell 2001). They too have deep roots in ecology. Darwin, despite arguing that deterministic processes were responsible for coexistence, acknowledged that the number and complexity of these could make community dynamics appear random (Darwin 1859). Hutchinson, even while developing what would become an archetypal expression of niche theory, similarly felt that “individual and unpredictable complexities in the determination of the niche boundaries...[mean] that in any overall view, the process would appear random” (1959, p154).

This was clearly not a contentious observation in itself, but neither did it provide a robust basis for testing the extent to which real communities were structured by non-random processes. This would instead come from the neutral model of population genetics, in which the spread of neutral mutations through a population was used as a null model for evolution by natural selection (e.g. Feller, 1951; Moran, 1958). The papers of Karlin & McGregor (1967) and Kimura (1968) are generally regarded as the first true neutral models of population genetics. These and others (e.g. Kimura, 1986; Hudson et al., 1987) can largely be adapted to ecology simply by changing the names of their variables so that genes, for example, become species (Chave 2004).

The first fully neutral model in ecology was published by Caswell (1976; based on Ewens 1972). It was able to produce relative abundance distributions that matched the log-series distribution that was among those considered empirically accurate (Fisher et al., 1943; Taylor et al. 1976). It was also able to predict very high levels of diversity, challenging the perceived importance of competition. The model was adapted by Hubbell (1979), who relaxed Caswell’s assumption of an infinite species pool and defined neutrality as ecological equivalence at the individual level, precluding any species-specific behaviour. He found that a wide range of relative abundance patterns could be produced and matched to observations by this theory.

Following this, a number of models were developed which included some neutral behaviour (e.g. Goldberg & Werner 1983; Graves & Gotelli 1983; Leigh et al. 1993; Terborgh et al. 1996), but most found that at least some non-neutral processes were necessary to explain empirical data. Hubbell’s updated neutral theory (1997; 2001) answered these criticisms by introducing a metacommunity in which speciation could occur and from which species could immigrate to the modelled community. This model was able to match a wide range of empirical species abundance distributions and has since become a focus of research, with 3212 papers citing Hubbell’s 2001 book (Google Scholar, 6/03/12).

1.7 Model tests and comparisons: first-order (non-spatial)

Hubbell’s (2001) neutral theory was tested primarily by its ability to produce species abundance distributions (SADs) that matched the shape of those observed in real communities. SADs were of particular interest as a basic but informative description of community structure, and were already
established as a test of niche-assembly theories (Bulmer, 1974; Engen & Lande, 1996a, b; May, 1976). The fact that SADs take a similar ‘hollow-curve’ shape in almost all known ecological communities has often been interpreted as evidence of their suitability for model testing (McGill et al. 2007).

First-order descriptions of community structure of this kind (relating to non-spatial properties of groups of individuals) have generally been the focus of model validation and comparison in ecology. This is partly due to the fact that they represent manageable descriptions of complex systems (Williams 1964; May 1981), but also because they are interpretable as expressions of general underlying mathematical laws. The search for laws of this kind motivated a great deal of early research in ecology (e.g. Volterra 1926; Lotka 1932), and prompted research into purely mathematical descriptions of species abundance patterns (e.g. Fisher et al. 1943). Debates about the appropriate mathematical function for fitting SADs and species-area curves are ongoing (Preston 1948; Bramson et al. 1996; 1998; McGill 2003).

Attempts to link these mathematical descriptions to particular ecological processes soon began, however (e.g. Kendall 1948). In fact, the first descriptions of SADs were mechanistic, usually being based on niche theory (Raunkiaer 1909; Smith 1913; Motomura 1932; MacArthur 1957). These became increasingly complex, but their ability to match empirical patterns did not improve substantially (Bulmer, 1974; Sugihara, 1980; Tokeshi, 1990; Marquet et al., 2003), and Hubbell’s (2001) neutral theory, which generates a zero-sum multinomial distribution of abundances, was found to be at least as successful as any of these.

Nevertheless, research of this kind has often been criticised for the apparent lack of insight that it provides. Hurlbert (1971) was among the first to express scepticism about the interpretability of numerical values or indices generated by theoretical distributions. Lepš (1990) argued that species abundance distributions are “able neither to detect nor to measure the biological interaction within a community” (p7). Neutral models in particular have been criticised as exercises in “fitting a relatively flexible mathematical function to a limited set of rank-abundance relationships” (Ricklefs 2006, p186). Nevertheless, the ability of a parsimonious theory such as neutrality to match some community characteristics at least as well as more complex and supposedly realistic models does recommend it, at least, as a suitable null model of community structure (Hubbell 2001). It may therefore be used to develop testable predictions which distinguish neutrality from other proposed mechanisms of coexistence. What is undoubtedly true is that the SAD and other first-order descriptions of community structure do not provide useful tests for such predictions, being unable to falsify any of a large number of contrasting theories about community dynamics (Vallade & Houchmandzadeh 2003; Volkov et al. 2003; McGill et al. 2007).

1.8 Model tests and comparisons: second-order (spatial)

While first-order descriptions of plant community structure have been the focus of model development and testing, research into spatial structure is equally well-established. Its importance has long been recognised (e.g. von Humboldt & Bonpland, 1805; de Candolle 1821), and rigorous descriptions of second-order structure (relating to the spatial properties of pairs of individuals) were developed alongside those of first-order structure discussed above (Watt 1947; Clark & Evans 1954; Hurlbert 1971). In recent years, however, the majority of these have come from statistical spatial point process theory, which, although partly motivated by ecology, has developed largely independently of it (Stoyan & Penttinen 2000).

To some extent, measures of β-diversity (differences in levels of diversity between sites) have been an alternative focus of research into the spatial structure of plant communities. Developed by
Whittaker (1960; 1972) and subsequently by many others (e.g. Legendre & Legendre 1998; Nekola & White 1999; Koleff et al. 2003; Legendre et al. 2005), such measures continue to be used in studies of coexistence mechanisms and other general processes (Terborgh et al. 1996; Chave & Leigh 2002; Condit et al. 2002; Dornelas et al. 2006). They effectively span the classes of first- and second-order measures, containing information on spatial structure that is aggregated to some sub-global level. This makes them particularly suitable where spatial signals are of interest but practical or computational limits prevent the use of fully spatially-explicit data. Such limits are increasingly absent, however, resulting in a shift away from measures of $\beta$-diversity in studies of community structure (e.g. Condit et al. 2000).

Spatial point process methodology, in contrast, allows for the analysis and description of patterns formed by the precise locations of individuals in space. Mathematically, these patterns may be described as realisations of a random variable referred to as a spatial point process (Illian et al. 2008). Because the formulation of a process controls the characteristics of the resulting patterns, it can be used both descriptively and for modelling. The descriptive use of spatial point process methods has been dominant in ecology, and is the focus of this thesis. Nevertheless, a wide range of spatial point process models are available (Cox 1955; Stoyan & Stoyan 1994; Møller et al. 1998; Van Lieshout 2000; Illian et al. 2008) and it is increasingly possible to fit these to ecological data with highly informative results (e.g. Baddeley & Turner 2005; Guan 2006; Rue et al. 2009; Illian et al. in press).

In some cases, the application of statistics originating in spatial point process theory to forest ecology has occurred alongside their development (Matérn 1960; Ripley 1976; 1977). They have generally been used to formally describe patterns of interest such as the associations between different species (Ogata & Tanemura 1985; Wiegand et al. 2007a), clustering in mortality (Sterner et al. 1986; Queenborough et al. 2007), patterns of colonisation (Salonen et al. 1992) and to try and separate the effects of local dispersal, species interactions and environmental niche differentiation (Tuomisto et al. 2003; Hardy & Sonké 2004; Wiegand et al. 2007b).

Although these applications have shown that spatial statistics can suggest underlying ecological processes, their use in formal inferential tests of these has remained rare. Indeed, this has often been discouraged by the possibility that different processes can generate similar or identical spatial patterns (Baddeley & Silverman 1984; Shipley & Keddy 1987; Stoyan & Penttinen 2000; Coomes et al. 1999). Lepš (1990), for example, argued that “mechanisms can be suggested on the basis of observed patterns, but they cannot be tested” (p9).

This is also true of statistical analysis in general, of course, in which correlations between variables (often describing a pattern of some form) are described for the purpose of inference and prediction (Austin 2002). In ecology, the use of simple (first-order) patterns for inference about putative fundamental laws has been widespread and particularly controversial (Andrewartha 1958; Hajnal 1958; Prigogine 1980; Cramer 1993). Macroeological patterns are often especially amenable to consistent statistical descriptions that resemble general laws, but considerable disagreement exists over whether these contain any real ecological information (Lawton 1999; Turchin 2001; Colyvan & Ginsburg 2003). The use of SADs to verify the neutral model, for example, has been described as a “spectacular warning of the dangers of inferring process from pattern” (Chave & Leigh 2002, p164).

Spatial statistics are clearly useful in this context as they extend the quantity of information about community structure that is available to inferential studies of the kind that already occur. The inclusion of further information can only be of benefit to such studies. Unlike traditionally-used first-order statistics, spatial statistics may enable the development of contrasting and testable predictions from different theories of coexistence (Amarasekare 2003). While unique links between process and pattern remain impossible, second-order spatial structure is therefore a more powerful indicator of
underlying process than the first-order patterns traditionally used (Illian et al. 2008; Law et al. 2009). Given this, a systematic investigation of the informative potential of spatial structure is required.

1.9 Thesis plan

This thesis seeks to examine the scope for detecting and distinguishing underlying ecological processes of the kind thought to explain species coexistence in tropical rainforests using spatial statistics. We will investigate the links between particular processes and the spatial structure of forest communities using established and newly developed summaries of spatial structure. We do so in order to develop robust methods of inferring underlying processes from resulting spatial patterns, with the ultimate aim of contributing to a resolution of the debate about the mechanisms of species coexistence. The principal theorised coexistence mechanisms described above have all proved able to match several observed first-order characteristics of ecological communities, and this has hampered efforts to determine which of them are accurate. The degree to which they can be distinguished by their spatial signals has not previously been systematically assessed, however. In Chapter 2 we attempt to do this using a stochastic individual-based model to generate spatially explicit data under each ecological process in isolation.

In Chapter 3, we go on to look for spatial signals in empirical data from a range of tropical rainforest plots, using the spatial statistic developed in Chapter 2. Different spatial predictions are generated by neutral and niche theories, with the structure of niche-assembled communities expected to vary systematically with the environment and the structure of neutral communities expected not to. We therefore test a prediction derived from niche theory that increasing topographical heterogeneity in tropical rainforests should lead to an increasing spatial spread of species along elevational gradients.

No single summary statistic is expected to be sensitive to all forms of spatial structure, and so we compare the performance of several established spatial statistics in Chapter 4. We do this on the basis of the simulated data in Chapter 2, where each process is again known to be operating in isolation. Our aim here is twofold: to find which pieces of information and statistical summaries of spatial structure are best able to distinguish mechanisms of coexistence, and what the full spatial implications of each of these mechanisms are. We do not consider an exhaustive range of spatial statistics but only those that are representative of widely used types, relying on particular spatial measurements. Those that perform best in this comparison are expected to form the basis for a probability-based test for underlying process.

In Chapter 5 we focus on the operation of one particular process, the Janzen-Connell effect, in a real rainforest community. Having a range of potential forms and extents, the scope for identifying this effect by its spatial signals is uncertain, and so the aim of this chapter is to investigate the strength and scale of these signals. We use hierarchical Bayesian models of empirical mortality data with respect to small-scale spatial structure so that the likelihood of identifying spatial signals can be estimated.

Finally, having established the theoretical ability of spatial statistics to distinguish coexistence mechanisms, looked for signals of these mechanisms in empirical data, attempted to develop a robust method of inferring underlying process from resulting spatial pattern and assessed the likelihood of encountering particular processes, we consider the implications of our findings for studies of rainforest ecology using spatial statistics. This thesis is intended to contribute to the development of methods capable of resolving the debate over mechanisms of species coexistence, and we discuss the potential for this resolution in light of the work presented here.
CHAPTER 2 - Linking ecological processes with spatial and non-spatial patterns in plant communities

The work presented in this chapter has been published in:


R. Law, J.B. Illian and D.F.R.P. Burslem supervised the work presented here and published in the above paper.
A great deal of research in ecology tries to infer ecological processes from patterns observed in nature, and this is one objective of this thesis. In community ecology, the species abundance distribution (SAD) has received particular attention, as discussed in Chapter 1 (also see e.g. McGill et al. 2007). A SAD describes the absolute or relative abundances of species in a community and is found to conform to a near-universal ‘hollow curve’ shape, comprising a small number of common species and a large number of rare ones. There is no obvious a priori reason to expect this shape, and the detailed features of SADs have therefore been used to discriminate between underlying processes, such as those involved in species coexistence. This work began with theories of niche assembly (e.g. Motomura 1932; MacArthur 1957; Tokeshi 1990), and more recently comparing SADs has become a key tool for validating the neutral theory against observed data from ecological communities (Hubbell 1979; 1997; 2001; Chapter 1).

As discussed above, a single ecological process can produce rather variable SADs (Magurran 2005; Williamson & Gaston 2005; Volkov et al. 2005; Chapter 1), making the detection of processes from empirically derived SADs difficult (McGill et al. 2007). This is not surprising: a SAD is, after all, just a description of species’ relative abundances averaged over space. Processes affecting coexistence rely on spatial proximity of individuals, especially in sessile organisms, and SADs convey no information on spatial structure. In the context of a spatial analysis of communities, a SAD would be said to be a first-order measure (Illian et al. 2008).

A motivating assumption of this thesis is that spatial correlations ought to provide a more sensitive indicator of ecological interactions among plant species because of the importance of interactions as drivers of spatial pattern in plant communities (Bolker & Pacala 1997; Murrell & Law 2003; Wiegand et al. 2007a). There is a long history in plant ecology of using spatial pattern to gain insight into ecological processes (e.g. Watt 1947; Clark & Evans 1954; Sterner et al. 1986), and indeed this was one motivation for the development of spatial point process methods (Matérn 1960; Ripley 1977; Stoyan & Penttinen 2000). It would be unrealistic to expect a unique mapping from a spatio-temporal process to a spatial pattern because of the array of biotic and abiotic factors at play (e.g. Baddeley & Silverman 1984; Lepš 1990), but it is reasonable to ask whether an analysis that makes use of spatial structure is a better discriminator among ecological mechanisms than one based on SADs that ignores this information.

Ecologists do often have far more information at their disposal than just that needed to construct SADs. For example, several complete spatial censuses exist for tropical rainforest trees on the 50-ha plot at Barro Colorado Island (BCI) in Panama (Hubbell et al. 2005) and numerous other sites (Losos & Leigh 2004). In the search for evidence about underlying processes on these plots, it should be possible to go beyond SADs (e.g. Hubbell 2001; Volkov et al. 2003; Etienne & Ollf 2005; He 2005) to second-order measures such as spatial correlations that make use of this spatial information. The potential of these has been recognized in studies of the roles of seed dispersal and habitat heterogeneity (Condit et al. 2000; John et al. 2007), the aggregations produced by neighbourhood recruitment and mortality (Hubbell et al. 2001; Uriarte et al. 2005), and spatial patterns in diversity (Weigand et al. 2007b). It is also further investigated in Chapters 3 and 4 of this thesis. Elsewhere, temporal patterns have been used to discriminate between neutral and non-neutral mechanisms for maintaining the structure and diversity of communities (e.g. Clark & McLachlan 2003; McGill et al. 2005), as have comparisons between different spatial scales (Gilbert & Lechowicz 2004; Dornelas et al. 2006; McGill et al. 2006). However, a systematic analysis of the spatial signatures generated by different kinds of species interaction has not previously been attempted.

This chapter evaluates the effectiveness of SADs and measures of spatial correlations to discriminate between multispecies spatial patterns that make different underlying assumptions about ecological
interactions (as described in Chapter 1). Our baseline was the neutral model, with its assumption of per capita ecological equivalence between species (Hubbell 2001). Two niche models were included: a conventional niche model in which species favour specific environmental conditions that are defined spatially (e.g. Grinnell 1917; Hutchinson 1958; Zillio & Condit 2007), and a lottery model in which temporal environmental variance favours different species at different times (Sale 1977; Chesson & Warner 1981; Chesson & Huntly 1988). The Janzen–Connell hypothesis, according to which young plants suffer increased mortality in the neighbourhood of their parents, was also implemented (Janzen 1970; Connell 1970), as was a purely spatial heteromyopia model in which interspecific competition occurs over shorter distances than intra-specific competition (Murrell & Law 2003).

We generated multispecies spatial patterns through realisations of spatio-temporal stochastic processes (stochastic individual-based models) using the different underlying models of ecological interactions. At first order, we computed SADs on these spatial patterns. At second order, we computed a new community-level measure of species segregation, built from spatial pair-correlation functions, referred to as the cross-pair overlap distribution (xPOD). The xPOD will be used repeatedly throughout this thesis.

First-order signals of the modelled ecological interactions were expected to be limited given the inherent variability of SADs. Some differences in community diversity and evenness were anticipated, but were difficult to predict because the relative strength of each form of interaction in promoting coexistence has not previously been assessed. Second-order spatial signals were expected to be substantially stronger, and to take a more predictable form for each model. In particular, the spatial niche model was expected to increase the segregation among species, while the Janzen–Connell model was predicted to constrain conspecific clumping and so have the opposite effect (Chapter 5 further investigates the spatial signals of the Janzen-Connell effect).

### 2.2 Materials and methods

#### 2.2.1 STOCHASTIC PROCESS FOR MULTISPECIES SPATIAL PATTERNS

Multispecies spatial patterns were obtained from realisations of a stochastic individual-based model (IBM) of a plant community based on a method in Law & Dieckmann (2000). This model will be used again in Chapter 4 and the description provided here is also relevant there. In this setting, individuals occur at discrete points \( \mathbf{x} = (x_1, x_2); x_1, x_2 \in [0,1] \) in a continuous two-dimensional space. The space comprises an arena of unit area with periodic boundaries, so forming a torus and preventing the inward propagation of edge effects. The spatial pattern \( p_i(\mathbf{x}, t) \) of a species \( i \) at time \( t \), \( t \in \mathbb{R}^+ \), comprises the locations of all individuals of species \( i \), and the multispecies pattern \( p(\mathbf{x}, t) \) is the union of all these single-species patterns. Birth and death events take place in continuous time, together with occasional arrival of new species, so the multispecies spatial pattern changes at every event. The effect of particular assumptions about ecological interactions on the multispecies spatial pattern is investigated after a large number of birth and death events have taken place.

For simplicity, the birth process is common to all species, independent of location in the arena, and comprises an intrinsic probability per unit time \( b \) of producing an offspring, and a function \( m(\mathbf{x} - \mathbf{x'}) \) giving the probability of the offspring being located at \( \mathbf{x'} \) for a parent at \( \mathbf{x} \). Thus the probability per unit time \( B(\mathbf{x}, \mathbf{x'}) \) of a parent at \( \mathbf{x} \) producing an offspring at \( \mathbf{x'} \) is

\[
B(\mathbf{x}, \mathbf{x'}) = b \, m(\mathbf{x} - \mathbf{x'}) \quad \text{eqn 2.1}
\]
In a community of finite size, there would be a gradual erosion of species diversity as extinctions take place. To counter this, a low constant probability per unit time of immigration of a new species is assumed, and the location of the new individual is chosen uniformly at random in the arena.

The death process is designed to allow various models of interactions among plants and is therefore more intricate. The probability per unit time of death $D_i(x,p)$ of an individual of species $i$ at point $x$ is:

$$D_i(x,p) = d_i + \sum_j d'_i \int w_{ij}(x'-x)[p_j(x',t) - \delta_{ij}\delta_i(x')]dx'$$  \hspace{1cm} \text{eqn 2.2}$$

Here $d_i$ is an intrinsic probability per unit time of death of an individual of species $i$. The term inside the integral takes a neighbour of species $j$ located at $x'$, and attaches weight $w_{ij}(x'-x)$, which depends on the displacement $x' - x$ of the neighbour from the target individual of species $i$ located at $x$, allowing the distance over which individuals interact to depend on the species identity of the target and neighbour. The integral adds up the effect of all neighbours of species $j$ on the target, and is given a weight $d'_ij$, so that interaction strength can also depend on the species identities. The summation adds the effects over all neighbour species $j$. In the case where $j = i$, the neighbours are conspecifics and the product $\delta_{ij}\delta_i(x')$, a Kronecker delta and Dirac delta function, is needed so that the target individual is not counted as a neighbour of itself.

Spatial structure in the community comes ultimately from local dispersal and local competition, assumed to be bivariate Gaussian functions

$$m(x-x') = \frac{1}{2\pi\sigma_b^2} \exp\left(-\frac{|x-x'|^2}{2\sigma_b^2}\right), \hspace{1cm} \text{eqn 2.3}$$

$$w_{ij}(x'-x) = \frac{1}{2\pi\sigma_{dij}^2} \exp\left(-\frac{|x-x'|^2}{2\sigma_{dij}^2}\right), \hspace{1cm} \text{eqn 2.4}$$

normalized to make the volume = 1. The parameters $\sigma_b$ and $\sigma_{dij}$ give the width of the distributions; small values indicate, respectively, that offspring are dispersed close to their parent and that competition occurs with close neighbours.

### 2.2.2 ECOLOGICAL MODELS FOR SPECIES COEXISTENCE

**2.2.2.1 Neutral model**

This is the simplest model and provides a standard against which to test the others. It was implemented by making the parameters $b$, $d$, $\sigma_b$ and $\sigma_{dij}$ constant and identical for every individual irrespective of species (Section 1.6; Table 2.1).

**2.2.2.2 Spatial niche model**

This differs from the neutral model in that species are sensitive to a spatially defined environmental variable (Section 1.2). We use a simple algorithm to construct a landscape based on a circular distribution divided into quadrants, producing a ‘peak’ in the centre of the arena $x = (0.5,0.5)$ and identical values at the periodic boundaries. The circular distribution in each quadrant is truncated at half its radius and the range within the arena scaled to the interval [0,1]. The value or ‘height’ of the environment at location $x$, $x = (x_1,x_2)$ is therefore:
This produces an environment in which the spatial extent of each specific value, and hence the size of each potential niche, is approximately equal, ensuring that any spatial signal produced is attributable to the niche process rather than environmental availability (see Fig. 2.1d).

The spatial niche model requires species to respond to the environment in different ways, so the niches of species are set by environment-dependent intrinsic death rates. Each species \( i \) is assigned a uniformly distributed random number in the range \([0,1]\) to give it a preferred niche \( h^0_i \). An individual in this optimal environment has a death probability \( d_i \) as in the neutral model. The dependence of the death rate on the environment is made explicit here by writing it as \( d_i(x) \) for an individual of species \( i \) located at a point \( x \) where the environment has a 'height' \( h_x \). The value of \( d_i(x) \) is assumed to have an inverted Gaussian shape around the birth rate \( b \), determined by the deviation \( h_x - h^0_i \) from the species' preferred niche, so that \( d_i(x) \) is minimized at \( h^0_i \):

\[
d_i(x) = b - \frac{(b - d_i)}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(h_x - h^0_i)^2}{2\sigma_n^2}\right) \tag{eqn 2.6}
\]

This means that, as the environment departs further from the preferred niche, the death rate of species \( i \) increases, eventually becoming the same as its birth rate and preventing it from becoming established. The parameter \( \sigma_n \) is a niche-width parameter common to all species. Thus, as \( \sigma_n \) becomes large, species do not perceive the landscape and the dynamics are as in the neutral model. Decreasing \( \sigma_n \) makes the success of each species increasingly dependent upon it being within its preferred niche \( h^0_i \).

2.2.2.3 Lottery (temporal niche) model
This model differs from the neutral model in that the intrinsic death rates \( d_i \) of species are drawn from a uniform distribution centred on that of the neutral model. The environment has a low, constant probability per unit time of changing, set at \(5 \times 10^{-5}\). When the environment changes, new values of \( d_i \) for each species are drawn from the uniform distribution. As the range of the uniform distribution tends to zero, the behaviour tends to the neutral model; increasing the range makes the lottery effect stronger.

2.2.2.4 Janzen–Connell model
This model requires local dispersal of propagules, applied here when the parameter of the dispersal function is sufficiently small. It also requires a higher death rate in the presence of conspecific neighbours than in the presence of heterospecific ones, due to host-specific enemies (Janzen 1970; Connell 1970; Section 1.4). We do not model host-specific enemies, and so the Janzen–Connell mechanism is introduced by making the neighbourhood conspecific death term \( d'_i \) larger than the heterospecific one \( d''_i \). In the natural environment, a distinction also would be made between the parent and the offspring, but we do not do this because age and size of individuals are not specified under the simple conditions of the stochastic process.

2.2.2.5 Heteromyopia model
This is a spatial mechanism for coexistence which requires shorter distances for interactions between species than within species. We introduce this mechanism by making the width parameter \( \sigma_{d''_i} \) for interaction distances with conspecifics larger than the width parameter \( \sigma_{d'_i} \) for interactions with heterospecifics.
2.2.3 NUMERICAL VALUES FOR SIMULATIONS

Each realization starts with 50 species of 100 individuals (5000 individuals in total), distributed independently and uniformly at random in the arena. The spatial pattern is updated by birth and death events using the Gillespie algorithm (Gillespie 1977), following the rules of the stochastic process defined above with parameter values as in Table 2.1. The ‘Mersenne-twister’ pseudo-random number generator (Matsumoto & Nishimura 1998) is used throughout. A small amount of large-scale dispersal is introduced by placing newborn individuals of existing species uniformly at random in the arena with a low probability of 0.01. In addition to more accurately describing the variable dispersal mechanisms employed by rainforest trees, this element of random dispersal enables species to colonize distant areas of preferred environment in the spatial niche model.

Parameter values in the different models are controlled to generate simple and clearly distinguishable departures from neutrality. Many of the parameters do not affect spatial structure and are used to set the spatial and temporal scale of the simulations; these include the birth and death rates \( b \) and \( d \), and the density-dependent death rates \( d_{ii}' \) and \( d_{ij}' \). Parameters controlling dispersal and density-dependence kernels are set to allow fine-scale behaviour but also some spatial mixing on the scale of the arena. These are common to all species under neutrality and are varied only where necessary in other models (Table 2.1).

<table>
<thead>
<tr>
<th>Model</th>
<th>( b )</th>
<th>( d_{i} )</th>
<th>( d_{ii}' )</th>
<th>( d_{ij}' )</th>
<th>( \sigma_{b} )</th>
<th>( \sigma_{d_{ii}} )</th>
<th>( \sigma_{d_{ij}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>0.4</td>
<td>0.2</td>
<td>( 4\times10^{-6} )</td>
<td>( 4\times10^{-6} )</td>
<td>0.01</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>spatial niche</td>
<td>0.4</td>
<td>(0.2-0.4)</td>
<td>( 4\times10^{-6} )</td>
<td>( 4\times10^{-6} )</td>
<td>0.01</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>lottery</td>
<td>0.4</td>
<td>( U(0.1,0.3) )</td>
<td>( 4\times10^{-6} )</td>
<td>( 4\times10^{-6} )</td>
<td>0.01</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>J-C</td>
<td>0.4</td>
<td>0.2</td>
<td>( 1.6\times10^{-5} )</td>
<td>( 4\times10^{-6} )</td>
<td>0.01</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>heteromyopia</td>
<td>0.4</td>
<td>0.2</td>
<td>( 4\times10^{-6} )</td>
<td>( 4\times10^{-6} )</td>
<td>0.01</td>
<td>0.020</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 2.1. Parameters for ecological models.

Boldface values indicate differences from the neutral model. J–C denotes the Janzen–Connell model.

* Defined as:
  \( b \) intrinsic birth probability per unit time;
  \( d_{i} \) intrinsic death probability per unit time;
  \( d_{ii}' \) weighting for within-species density-dependent deaths;
  \( d_{ij}' \) weighting for between-species density-dependent deaths (\( j \neq i \));
  \( \sigma_{b} \) dispersal kernel standard deviation;
  \( \sigma_{d_{ii}} \) within-species density-dependent death kernel standard deviation;
  \( \sigma_{d_{ij}} \) between-species density-dependent death kernel standard deviation;

† Environment \( h_{x} \) defined in text; preferred niche for species \( i \): \( h_{i}^{0} \in [0,1] \); niche width: \( \sigma_{n} = 0.2 \).

A low probability per unit time of an immigrant birth, 0.001, is allowed to counter extinctions of species that happen naturally as a result of the birth–death process in the finite arena. Immigrant species are drawn from an infinite pool, and consequently immigrations by existing species do not occur. In the lottery and niche models, immigrant species are randomly assigned death rates or
preferred niches, as above; in the other models their characteristics match those of existing species, incorporating Janzen–Connell or heteromyopia effects where appropriate. After $10^6$ events the realisation is stopped and the emergent multispecies spatial pattern used to test for differences between ecological models of interactions. We checked to make sure that this allows sufficient time for the realisation to reach a stationary distribution in terms of both the total number of individuals and the number of species (e.g. Figs. 2.1a and 2.1b). The characteristics of the spatial patterns were also found to be stable by this point, with no systematic temporal changes occurring. For each model, 10 independent spatial patterns are generated, and are referred to as replicates below.

### 2.2.4 CROSS-PAIR OVERLAP DISTRIBUTION (XPOD)

We define a new community-level measure of the spatial overlap of species, based on the distribution of spatial overlaps of pairs of species. This measure will be used again in Chapters 3 and 4, where the definition will be repeated and further details given as necessary. A measure aggregated to the community level is needed as it would be unmanageable to work with all species pairs separately in a multispecies community. The measure is built from the cross pair correlation function $g_{ij}(r)$ (Diggle 2003; Illian et al. 2008; Appendix 2.1) of each species pair $i, j$, which is reduced to a scalar quantity, the area $A_{ij}$ obtained by integrating the log of the function. This describes the average overlap of species $i$ and $j$ up to a fixed distance $R$:

$$A_{ij} = \int_0^R \log(g_{ij}(r))dr \quad \text{eqn 2.7}$$

An $A_{ij}$ close to zero implies that the two species are close to independent on average, up to $R$. A positive $A_{ij}$ implies that they tend to occur together, and a negative one that they are separated in space. Taking $A_{ij}$ of all non-self combinations of $i$ and $j$, gives a cross-pair overlap distribution (xPOD) for the community as a whole. (We do not consider the self-pair overlap distribution, i.e. the case $i = j$, because it is the between-species spatial structure that is of concern here.) At the community level, an xPOD with predominantly positive values of $A_{ij}$ implies an overall tendency for species to occur together, and one with predominantly negative values implies a tendency for species to be separated.

Estimation of the cross-pair correlation functions $g_{ij}(r)$ is based on the method given in Law et al. (2009). Following Baddeley & Turner (2005), an upper limit for estimation of $g_{ij}(r)$ is set at $R = 0.25$, so that spatial behaviour of interest is not averaged out over larger areas. We chose this upper limit as being substantially greater than the spatial scale of the dispersal and competition kernels, but substantially smaller than the spatial scale of the arena. Tests with an upper limit of $R = 0.15$ gave similar results, but variation in observed behaviour with spatial scale could occur and warrants further investigation.

Estimation of $g_{ij}(r)$ and $A_{ij}$ requires discretization of $r$, for which we use 10 bins of equal width. Comparison with results generated by up to 100 bins suggested that 10 are sufficient to capture the spatial behaviour without being unduly influenced by small-scale noise. The xPODs are constructed for species pairs in which both species have at least 500 individuals, so that random noise in spatial pattern from small samples does not mask the signal for the ecological interaction.

### 2.2.5 COMMUNITY EVENNESS AND DIVERSITY

In addition to first- and second-order summaries of community structure, we derive mean indices of community evenness and diversity for each model. The indices used are Pielou’s evenness index
(Pielou 1966) and Shannon’s diversity index (Shannon & Weaver 1949). The larger the positive values of these, the greater the evenness and diversity of the community, respectively.

2.2.6 STATISTICAL INFERENCE

To compare SADs and xPODs generated by each simulation, we use the Kolmogorov–Smirnov (K–S) goodness-of-fit test statistic (Massey 1951), a nonparametric and distribution-free measure of the similarity of distributions. The comparison is done separately for the two measures so that their capacity to distinguish ecological models can be evaluated, and, in the case of SADs, at abundance thresholds of both 1 and 500 to ensure fair comparison with the xPOD. The K–S test statistic gives a measure of similarity of pairs of distributions; its value is calculated for every pair of SADs and xPODs in the study, giving 45×5 values within ecological models and 100×10 values between ecological models. At this broad level, the distribution of K–S values shows the relative magnitude of random differences within ecological models and systematic differences between ecological models. Comparisons within and between particular ecological models are also made by disaggregating the K-S values down to single ecological models. This is not intended as a method of identifying the underlying model, however, but simply as an expression of the magnitude and consistency of visible differences between distributions.

We additionally quantify random variability within models using prediction intervals for SADs and xPODs. These intervals provide, with a confidence of 95%, the range within which future observations will be found, given the observations already made. We chose prediction intervals instead of confidence intervals because the range of possible results is of more interest than the potential value of a mean.
Figure 2.1. Realization of a spatial multispecies birth–death process, showing the statistics calculated on the multispecies spatial pattern at the end.

(a) Total number of individuals with time; (b) total number of species with time; (c) species abundance distribution; (d) the spatial pattern of two species extracted from the multispecies pattern on a contour plot of the underlying environment; (e) logarithm of the pair correlation function of the two species in (d); $A_{ij}$, the sum of the shaded areas; (f) cross pair overlap distribution, the histogram of $A_{ij}$s.
2.3 Results

2.3.1 Illustrative Example

To help understanding of the results that follow, we start by showing the statistics estimated from a single multispecies spatial pattern. To do this, we use the stochastic IBM with a spatial niche model. By the end of the simulation, when the spatial pattern was extracted for analysis, both the number of individuals and the number of species were close to steady state, as in all other realizations (Figs. 2.1a and 2.1b). This behaviour over time is typical of all the simulations of the IBM carried out. The SAD calculated from the emergent multispecies spatial pattern had the shape typically observed in these distributions, i.e. a relatively small number of common species and many rare ones (Fig. 2.1c).

To construct the xPOD from the multispecies spatial pattern, the spatial patterns of species were extracted in pairs. For the sake of illustration, the pair in Fig. 2.1d is chosen as one that exhibits clear spatial segregation. The correlation function $g_{ij}(r)$ of this pair thus has values less than one at short distances, and rises to close to one as radius increases and the density of pairs approaches the spatial average. Corresponding to this, $\log(g_{ij}(r))$ is negative at short distances and increases to around zero as radius increases (Fig. 2.1e). The integral over $r$, $A_{ij}$, sums the shaded areas in Fig. 2.1e, keeping track of the sign; this is also negative, and provides one datum for constructing the xPOD. Repeating this calculation for all pairs of species above the threshold abundance gives a distribution of $A_{ij}$s - the xPOD of the spatial pattern (Fig. 2.1f). Although the xPOD has a centre near to zero, it has a tail of negative values because the spatial niche model leads to segregation of species with dissimilar niche requirements (see Chapter 3).

2.3.2 Species Abundance Distributions

Irrespective of the ecological model, SADs had the general property of a small number of abundant species and a large number of rare ones (Fig. 2.2). The most frequent abundance class was for species with just one individual, and these mainly comprised recent immigrants that would subsequently have become extinct.

Importantly, there were limited differences between the SADs arising from different ecological models (Fig. 2.2). Some small differences were apparent in community evenness, with the lottery model producing the least even and diverse communities on average, and the Janzen–Connell model producing the most even and diverse (Table 2.2). However, these differences were not consistently detected by the indices used and were small in comparison to the random variation in shape produced within each model, which was especially large in the lottery model.

This impression is confirmed by the distributions of K–S test statistics (Fig. 2.3). These indicate that variation between SADs was similar irrespective of whether the pairs were drawn from replicate realisations within ecological models or drawn from different ecological models. Disaggregating down to particular ecological models still led to mean values of the K-S test statistic for pairs of SADs from replicates within single models as large as those for pairs of SADs taken from different models (Table 2.3).

Species of low abundance were one cause of variation between SADs, and the mean values of the K–S test statistic for pairs of SADs within some models were reduced when the SADs were restricted to species with at least 500 individuals; the same abundance threshold applied to the xPODs (Table 2.4). Although variation among SADs was greater between than within ecological models when using this threshold, the magnitude of the difference was small. Hence, distributions generated using the same model were not consistently more similar than those generated by different models.
The use of this abundance threshold did not therefore affect our ability to distinguish underlying models using SADs.

<table>
<thead>
<tr>
<th></th>
<th>Lottery</th>
<th>Niche</th>
<th>Heteromyopia</th>
<th>Neutral</th>
<th>Janzen-Connell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pielou</td>
<td>0.43</td>
<td>0.72</td>
<td>0.71</td>
<td>0.71</td>
<td>0.81</td>
</tr>
<tr>
<td>Shannon</td>
<td>2.59</td>
<td>3.84</td>
<td>4.41</td>
<td>4.44</td>
<td>5.27</td>
</tr>
</tbody>
</table>

**Table 2.2. Mean evenness and diversity of each model.**

Indices used are Pielou’s evenness index and Shannon’s diversity index.

### 2.3.3 CROSS-PAIR OVERLAP DISTRIBUTIONS

In contrast to the SADs, xPODs produced by different ecological models were consistently distinct, and their prediction intervals showed systematic differences between models that far outweighed random variation between replicates within models (Fig. 2.4). The degree of overlap between species at the community level was ordered: spatial niche < lottery < heteromyopia/neutral < Janzen–Connell. This can be seen in the first three moments of the xPODs showing an increasing (though always negative) mean value, a declining standard deviation, and a broadly declining skewness (Table 2.5). The only exceptions were the neutral and heteromyopia xPODs which were almost entirely coincident.

Goodness-of-fit test statistics showed that xPODs generated by the same model were far more alike than the equivalent SADs, while those produced by different models were far more dissimilar (Fig. 2.5). Disaggregating down to particular ecological models, the neutral, niche and lottery models produced distinct results, while the results of the neutral and heteromyopia models were the most similar. Variability in lottery model xPODs occurred where the number of abundant species was low and the distribution correspondingly sparse.
Figure 2.2. Species abundance distributions (SADs) obtained from different ecological models.

An example of a SAD from a single realisation (dashed line) and 95% prediction intervals for the SADs based on 10 realisations (solid lines) are shown. Abundances are expressed as the logarithm of abundance relative to total community size in order to facilitate comparison.
Figure 2.3. Boxplots of Kolmogorov–Smirnov test statistics.

Values are obtained from comparing pairs of species abundance distributions within and between ecological models (producing 45 and 100 values respectively). Results shown are from comparisons: (a) within neutral; (b) neutral–niche; (c) neutral–lottery; (d) neutral–Janzen-Connell; (e) neutral–heteromyopia. Whiskers extend to the largest and smallest values within 1.5 times the interquartile range and outliers are not shown. Low values of the test statistic indicate similarity of distributions.
Table 2.3. Mean and 90% limit values (in brackets) of the Kolmogorov–Smirnov test statistic comparing pairs of SADs within and between contrasting models of ecological interactions, using species of all abundances. 90% limits are used to summarise the range of test statistic values in each comparison.

Mean values based on 45 pairs of SADs within ecological models and 100 pairs between models. Within-model means are in bold. J–C denotes Janzen–Connell.

<table>
<thead>
<tr>
<th>Model</th>
<th>Neutral (0.194 (0.223 / 0.156))</th>
<th>Niche (0.199 (0.229 / 0.156))</th>
<th>Lottery (0.212 (0.348 / 0.183))</th>
<th>J-C (0.211 (0.249 / 0.169))</th>
<th>Heteromyopia (0.206 (0.233 / 0.175))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>0.194 (0.223 / 0.156)</td>
<td>0.199 (0.229 / 0.156)</td>
<td>0.212 (0.348 / 0.183)</td>
<td>0.211 (0.249 / 0.169)</td>
<td>0.206 (0.233 / 0.175)</td>
</tr>
<tr>
<td>Niche</td>
<td>0.271 (0.297 / 0.255)</td>
<td>0.259 (0.348 / 0.201)</td>
<td>0.201 (0.468 / 0.197)</td>
<td>0.208 (0.438 / 0.201)</td>
<td>0.208 (0.233 / 0.182)</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.264 (0.348 / 0.266)</td>
<td>0.317 (0.468 / 0.197)</td>
<td>0.242 (0.348 / 0.201)</td>
<td>0.230 (0.258 / 0.204)</td>
<td>0.207 (0.233 / 0.175)</td>
</tr>
<tr>
<td>J-C</td>
<td>0.135 (0.153 / 0.123)</td>
<td>0.235 (0.153 / 0.123)</td>
<td>0.208 (0.233 / 0.182)</td>
<td>0.207 (0.233 / 0.175)</td>
<td>0.207 (0.233 / 0.175)</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4. Mean and 90% limit values (in brackets) of the Kolmogorov–Smirnov test statistic comparing pairs of SADs within and between contrasting models of ecological interactions, using an abundance threshold of 500 individuals. 90% limits are used to summarise the range of test statistic values in each comparison.

Within-model means are in bold. J–C denotes Janzen–Connell.
Figure 2.4. Cross pair overlap distributions (xPODs) obtained from different ecological models.

The shape of an example xPOD from a single realization (dashed lines) and 95% prediction intervals for xPODs based on 10 realizations (solid lines) are shown. Frequencies of areas are given relative to the total number of areas in each distribution.
Table 2.5. Mean, standard deviations and skewness of cross pair overlap distributions for contrasting models of ecological interactions.

J–C denotes Janzen–Connell.

<table>
<thead>
<tr>
<th></th>
<th>Niche</th>
<th>Lottery</th>
<th>Heteromyopia</th>
<th>Neutral</th>
<th>J-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.267</td>
<td>-0.107</td>
<td>-0.055</td>
<td>-0.053</td>
<td>-0.050</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.319</td>
<td>0.132</td>
<td>0.043</td>
<td>0.040</td>
<td>0.026</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.556</td>
<td>-0.968</td>
<td>-0.434</td>
<td>-0.329</td>
<td>-0.262</td>
</tr>
</tbody>
</table>

Table 2.6. Mean and 90% limit values (in brackets) of the Kolmogorov–Smirnov test statistic comparing pairs of xPODs within and between contrasting models of ecological interactions. 90% limits are used to summarise the range of test statistic values in each comparison.

Within-model means are in bold. J–C denotes Janzen–Connell.
Figure 2.5. Boxplots of Kolmogorov–Smirnov test statistics obtained from comparing pairs of cross pair overlap distributions within and between ecological models.

(a) Within neutral; (b) neutral–niche; (c) neutral–lottery; (d) neutral–Janzen-Connell; (e) neutral–heteromyopia. Whiskers extend to the largest and smallest values within 1.5 times the interquartile range and outliers are not shown. Low values of the test statistic indicate similarity of distributions.
2.4 Discussion

Attempts to infer underlying process from pattern in ecology have tended to rely on first-order community characteristics such as the SAD, which is the chief empirical test of many theories of species coexistence and the focus of a great deal of theoretical study in its own right (Fisher et al. 1943; Hubbell 2001; Volkov et al. 2003; McGill et al. 2007; Chapter 1). Relatively little work has investigated the potential of second-order spatial information for similar inference, and we are aware of none that has systematically assessed the first- and second-order signals of modelled ecological processes.

The SADs produced by the models considered here show, on visual inspection, a great deal of overlap and considerable variation within those produced by any one model. There may be certain regions of the distributions where the models are distinct from one another, particularly in terms of community evenness. The lottery model produced the least even communities and the Janzen–Connell model the most even and diverse ones. These differences are small, however, and do not prove reliably distinguishable by the statistical tests used here. Differences in overall species diversity are attributable to model design, especially in the niche model where fecundity is depressed outside species’ optimum environment.

The Kolmogorov–Smirnov test indicates that random variations in species abundances outweigh similarities engendered by ecological process, as the test statistic takes very similar values for single-model pairs and cross-model pairs. On the basis of the values found here, SADs produced by the same model are, at best, only slightly more similar to one another than they are to those produced by different – or even contradictory – processes. The variations we find in SADs cannot, therefore, be classified as random or systematic by magnitude alone.

Truncating the SADs at an abundance of 500 slightly depresses within-model variation relative to between-model variation. Further analysis indicates that a threshold of 100 is better yet; this suggests that it may be possible to find an optimum section of the distribution in which random variation is minimized. Nevertheless, random and systematic differences between distributions, as captured by the Kolmogorov–Smirnov test statistic, remain of the same order in the threshold values that we have investigated (1, 100 and 500).

Previous studies of the SAD have generally concentrated either on its mathematical description (e.g. Fisher et al. 1943; Preston 1948; Dewdney 2003; McGill 2003; Volkov et al. 2003) or on comparing the shapes of the distribution generated by theoretical and empirical data (e.g. Motomura 1932; Tokeshi 1990; Hubbell 2001). It has also been argued that potential exists for distinguishing disturbance factors (Mouillot & Lepretre 2000), habitat complexity (Hurlbert 2004) and community stability (Thibault et al. 2004) by comparing SADs. While the SAD may be of value in studies of processes related to these, we find little justification for the assumption that the coexistence mechanisms that we modelled can be distinguished in its shape (Hubbell 2001; Etienne & Ollf 2005; Marquet et al. 2003).

The xPODs of each model are more reliably different from one another than the SADs. This would be expected because an xPOD is essentially an accessible summary of the spatial relationships between species, and should reflect different models of species interactions in so far as they generate different spatial patterns. There are large differences between results from each model, with the exception of the neutral and heteromyopia models, and prediction intervals demonstrate that random within-model variation does not erode the distinction between models. This holds true even for the lottery model, in which random variation is large. At these parameter settings, the ecological processes modelled here are eminently distinguishable by their spatial signals (see also Chapter 4).
The neutral and heteromyopia models produce xPODs with slightly negative means and small variances, the lottery and niche models give more negative means and skew distributions with greater variances, and the Janzen–Connell model a smaller negative mean and smaller variance. These signals broadly conform to expectations. The negative mean of the xPOD under neutrality is due to separation of species by density-dependent mortality, which is balanced within species by local dispersal. In the niche model, a greater spread of (mainly negative) values is directly attributable to the introduction of spatial niches which separate species according to their response to the underlying environment. While the spatial extent of each niche is approximately equal, the strength of niche separation itself is weak; niches are randomly assigned and so may overlap or entirely coincide. Stronger niche separation, corresponding to competition between species, would be expected to increase the spread of the xPOD further (and has been found to do so in trial results).

The lottery model lacks a mechanism for such strong separation of species, but the temporal variations in species fecundity allow dominant species to mingle with others and force them apart, so producing the observed spread of values. This effect is likely to be dependent upon the frequency of variation in fecundity (O’Malley et al. 2010). The Janzen–Connell effect counteracts the tendency of species to separate slightly under neutrality by suppressing conspecific clumping (Chapter 5). There remains no mechanism to encourage the mingling of species, however, so the distribution produced has a smaller spread, as expected. Although the Janzen–Connell effect modelled here is not explicitly overcompensating, it does effectively prevent the survival of individuals in the immediate vicinity of their parent at equilibrium density, as required by the original theory and observed in some tropical tree species (Janzen 1970; Freckleton & Lewis 2006; Bagchi et al. 2010).

The heteromyopia model produces distributions indistinguishable from those of the neutral model, with identical means and variances. This may be because density-dependent mortality within species is spread over a larger area, so allowing increased clumping at small scales which counterbalances any repression of clumping at medium scales. This conformity of spatial patterns between models illustrates the absence of unique links between process and pattern. Nevertheless, differences between the spatial consequences of neutrality and heteromyopia are further investigated in Chapter 4.

Goodness-of-fit tests show that xPODs have far more success in distinguishing models than SADs. xPODs produced by different models are highly dissimilar, except those produced by the neutral and heteromyopia models, while distributions drawn from the same model are far more alike, producing smaller and more restricted values of the test statistic (Fig. 2.5, Table 2.6). The sole exception is the lottery model, in which random fluctuations in fecundities give rise to substantial variation in spatial pattern. Despite this, the difference between these and the comparable results for SADs (Fig. 2.3) is striking; the fingerprint of each model emerges from the background noise clearly and consistently in the case of the xPOD.

Our findings are subject to several caveats, the most important of which relates to the direct deduction of process from pattern. While we find that second-order patterns are significantly more informative than their first-order counterparts, it remains unreasonable to assume an inviolable causal link from underlying mechanisms (e.g. Baddeley & Silverman 1984; Lepš 1990). It must also be stressed that our results come from computer-based simulations, and that behaviour which we hold constant in the interests of clarity does not obey this requirement in real-world plant communities, potentially producing quite different outcomes (as similarly noted by e.g. Chave & Leigh 2002; Chave et al. 2002; Levin et al. 2003, but see Chapter 3). In particular, dispersal distances vary widely between species and have a dramatic impact on spatial structure (Levin et al. 2003). Changes in the relative scales of dispersal and environmental variation may have a similarly confounding effect (e.g. Lande et al. 1999; Wiegand et al. 2007a). Complex species-specific parameterization of dispersal kernels would be difficult to perform accurately and would undermine
our attempts to identify the spatial signals of isolated processes, and so requires further research (see Appendix 2.2).

We also make assumptions about the environment and the expression of niche differentiation in our spatial and temporal niche models which may affect observed behaviour. While greater separation of niches is expected to strengthen the signal detected in our niche model, greater niche overlap would have the opposite effect because species remain effectively neutral within shared niches. We constrain our environment to a single circular distribution of values to ensure that differences in niche availability do not affect our results. However, the signals of niche differentiation in patchier environments would remain similar to those observed here, at similar relative scales, due to spatial separation between species with differing niche requirements (Chapter 3). As in all models presented here, niche processes act through changes to the death term for the sake of consistency but are equivalent to the opposite adjustment to the birth term.

It is notable that the xPOD, as used here, still only makes use of a small fragment of the spatial information that is often available. For instance, marked point patterns from forests often contain information on spatial structure in relation to size, age and environment, and at different spatial scales. Moreover, there is spatial structure at third-order and beyond. Nevertheless, second-order information evidently helps to distinguish underlying process, and is more informative than SADs, the well-established first-order measure. Together with the ready availability of spatially explicit data, this suggests that much of the power of spatial analysis in ecology remains to be exploited.

While bearing the above caveats in mind, the differences between the spatial signatures found here are large and intuitive, especially between the neutral model and the spatial niche model. We therefore predict, as a hypothesis for empirical research, that plant communities in environments with more physical heterogeneity should have xPODs characterized by more negative means and larger variances than plant communities in environments with less physical heterogeneity. This hypothesis will be tested in Chapter 3, using data from tropical rainforest plots. If the neutral model is correct, the hypothesis will not be supported.
2.5 Appendix 2.1: Definition of the pair correlation function

The pair correlation function is defined mathematically as follows. Consider a small area $|dx|$ containing the point $x$ and another area $|dy|$ containing $y$, together with the number $n_i(dx)$ of individuals of type $i$ in $|dx|$, and the number $n_j(dy)$ of type $j$ in $|dy|$. The pair density $N_{ij}(x,y)$ is defined as

$$N_{ij}(x,y) = \lim_{|dx| \to 0, |dy| \to 0} \left\{ E \left[ n_i(dx) n_j(dy) \right] \right\},$$

eqn A2.1.1

where $E$ is the expected value (Diggle 2003:43). The pair correlation function $g_{ij}(x,y)$ is the dimensionless normalised form of the pair density, obtained by dividing by the intensity $\lambda_i(x)$ of $i$ in $dx$ and the intensity $\lambda_j(y)$ of $j$ in $dy$. If the spatial pattern is stationary and isotropic, the pair correlation depends only on the distance $r = ||x-y||$ between $x$ and $y$, given as $g_{ij}(r)$. 

2.6 Appendix 2.2: Sensitivity analysis of variation in dispersal kernels

In this Appendix we extend the neutral and niche models presented in Chapter 2 in order to more accurately describe the range of dispersal kernels found in tropical rainforests. For the sake of simplicity and to isolate the spatial signals of the modelled processes, all of the models in Chapter 2 included a single dispersal kernel that applied to all species (Table 2.1). This is known to be inaccurate, however, because dispersal mechanisms and distances vary greatly between species (e.g. Muller-Landau et al. 2008). Such variation is expected to have substantial effects on spatial structure, and may erode the differences we detect between models.

Here, we parameterise dispersal kernels in the neutral and niche models to vary randomly between species within an empirically realistic range centred on the value of 0.01 used in Chapter 2 (Table A2.2.1). This range is taken from a study of dispersal kernels in the rainforest plot at Barro-Colorado Island in Panama (Hubbell & Foster 2005) presented in Muller-Landau et al. (2008). The models and methods are otherwise as in Chapter 2.

2.6.1 RESULTS & DISCUSSION

The inclusion of variable dispersal kernels produces a smaller range of spatial overlaps in the xPOD for both the neutral and niche models (Fig. A2.2.1). The differences between the two are preserved, however, with the niche model still giving a far wider range of values than the neutral model. These results suggest that all species pairs have more random relative spatial distributions when dispersal kernels vary, presumably because the ability of some species to disperse more widely encourages greater overall mixing. As this does not erode the relative signal of niche differentiation, we conclude that the spatial signals identified in Chapter 2 are likely to be robust to the differences between species in dispersal distances found in real-world rainforest communities.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>b</th>
<th>d_i</th>
<th>d'_ii</th>
<th>d'_{ij}</th>
<th>σ_b</th>
<th>σ_{d_{ii}}</th>
<th>σ_{d_{ij}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td></td>
<td>0.4</td>
<td>0.2</td>
<td>4 \times 10^{-6}</td>
<td>4 \times 10^{-6}</td>
<td>(0.005-0.02)</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>spatial niche</td>
<td></td>
<td>0.4</td>
<td>(0.2-0.4)</td>
<td>4 \times 10^{-6}</td>
<td>4 \times 10^{-6}</td>
<td>(0.005-0.02)</td>
<td>0.005</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table A2.2.1. Parameters for ecological models.

Boldface values indicate differences from the neutral model

* Defined as:
  b intrinsic birth probability per unit time;
  d_i intrinsic death probability per unit time;
  d'_ii weighting for within-species density-dependent deaths;
  d'_{ij} weighting for between-species density-dependent deaths;
  σ_b dispersal kernel standard deviation;
  σ_{d_{ii}} within-species density-dependent death kernel standard deviation;
  σ_{d_{ij}} between-species density-dependent death kernel standard deviation;
  † Environment h_x defined in text; preferred niche for species i: h_i^0 \in [0,1]; niche width: σ_n = 0.2.
Figure A2.2.1. Cross pair overlap distributions (xPODs) obtained from neutral (right) and niche (left) ecological models with a range of dispersal kernels as in Table A2.2.1.

The shape of an example xPOD from a single realisation (dashed lines) and 95% prediction intervals for xPODs based on 10 realisations (solid lines) are shown. Frequencies of areas are given relative to the total number of areas in each distribution.
CHAPTER 3 - Multispecies coexistence in tropical forests: spatial signals of topographic niche differentiation increase with environmental heterogeneity

Data used in this Chapter were provided by:

3.1 Introduction

In this chapter we test a hypothesis developed in Chapter 2, that the spatial structure of tropical rainforests should vary systematically and detectably with environmental heterogeneity if niche differentiation exists between species. In Chapter 2, we demonstrated that the spatial signal of niche differentiation was detectable using the cross-pair overlap distribution (xPOD). Here, we apply this statistic to empirical data, taken from tropical rainforest plots from around the world.

As discussed above, niche differentiation is an important subject for investigation because the specialisation of plant species to particular environmental conditions is one potential mechanism supporting the coexistence of hundreds of species in tropical rainforests (Ashton 1969; Grubb 1977; Ricklefs 1977; Chesson 2000). While such niche differentiation is known to occur in numerous specific cases (e.g. Clark et al. 1998; Svenning 2001; Schulman et al. 2004; Tuomisto 2006; Keppel et al. 2011), its role in maintaining diversity at large scales remains unclear (Wright 2002; Gravel et al. 2011). In particular, the neutral theory’s ability to accurately predict community-scale patterns such as Species-Area Relationships (SARs) and Species Abundance Distributions (SADs) without recourse to any functional differences between species suggests that coexistence is instead a function of neutral and ongoing ‘drift’ of species (Bell 2000; Hubbell 1997; 2001; Horvat et al. 2010; Chapters 1 and 2).

As we have shown, however, the relationship between processes and patterns such as SADs and SARs is not one-to-one and their shape cannot be used to directly infer underlying processes (McGill et al. 2007; Chisholm & Pacala 2010). Neutral and niche effects, singly or in combination, have been shown to produce patterns indistinguishable from one another (e.g. Chave 2004; Purves & Pacala 2005; Ruokolainen et al. 2009; Du et al. 2011), although differences may exist at particularly small or large spatial scales (Jabot & Chave 2011). The importance of niche differentiation in community dynamics therefore remains unknown.

In recent years, the use of summary statistics that include information on spatial structure has become increasingly widespread in ecology (Law et al. 2009). These provide a more promising basis for the discrimination of ecological processes from their resulting patterns, both in empirical (Condit et al. 2002; Seidler & Plotkin 2006) and simulated settings (Rangel & Diniz-Filho 2005; Chapter 2). The possibility of resolving the neutral-niche debate has driven their development and application, but a resolution has not yet been realised (Hardy & Sonké 2004; Leibold & McPeek 2006; Chase & Myers 2011; Weiher et al. 2011). One reason for this is that limited dispersal – in theory a potentially neutral process in that it may occur identically across species – can produce forms of clustering and species turnover that correlate spuriously with environmental variation, implying niche differentiation where none in fact exists (Harms et al. 2001; Bin et al. 2010). This problem is exacerbated by the fact that only a small minority of species’ dispersal kernels have been estimated (but see e.g. Muller-Landau et al. 2008), and that niche effects, if present, would depend upon the magnitude and scale of environmental variation in ways that are hard to predict (Rejou-Mechain & Hardy 2011). When both occur together, disentangling the two is extremely difficult. Attempts to do so generally rely upon an assumed separation of scales between the effects, often involving the selection of an arbitrary radius at which environmental variation is thought to replace dispersal limitation as the dominant spatial effect (Karst et al. 2005; Wiegand 2007). A variation on this technique is the identification of specific habitats or habitat types, within which niche effects are assumed to be negligible (Gilbert & Lechowicz 2004; Rejou-Mechain & Hardy 2011).

At the level of particular species or groups of species, evidence of niche differentiation is strong. Numerous studies have identified associations of tropical forest tree species to environmental conditions, in terms of elevation, soil chemistry and water content, exposure, humidity, temperature, and light regimes (Lieberman et al. 1996; Kobe 1999; Debski et al. 2002; Gunatilleke et
3.2 Methods

3.2.1 Data

Data for all free-standing woody plants ≥ 1 cm diameter at breast height (dbh) mapped to a precision of 0.1 m, and topography surveyed to a precision of 20 m, were obtained for 12 plots in the network coordinated by the Center for Tropical Forest Science (Losos & Leigh 2004). The plots were Barro-Colordo Island (BCI) in Panama, Dinghushan, Gutianshan and Xishuangbanna in China, Fushan and Lienhuachih in Taiwan, Luquillo in Puerto Rico, Mo Singto in Thailand, Mudumalai in India, Pasoh in
Malaysia, Sinharaja in Sri Lanka and Yasuni in Ecuador (Table 3.1). Topography was used to measure environmental variation because other data, relating for example to soil chemical and physical properties, are not available for all sites. Plots were selected to sample a wide range of topographical characteristics within the tropical forest biome, but without further restrictions on geographical location, altitude or climate. We set a lower limit of 16 ha on plot size in order to maximise the number of sites available for study, while ensuring that a minimum of 100 species pairs met our diameter and species abundance thresholds to produce a clear cross-pair overlap distribution.

The twelve plots vary between 16 ha and 50 ha in size, and between 25 m (Pasoh) and 269 m (Gutianshan) in elevation range. Plots also differed in terms of longitude, latitude, climate, disturbance regimes, species diversity and stem density, amongst other characteristics. These substantial and important differences should have significant impacts on the ecological processes occurring in the forests, including those that affect spatial patterns. However, our hypothesis is that the signal of environmental heterogeneity introduced into the relative locations of species by niche differentiation is universal, and so should be broadly predictable in comparisons of plots regardless of their individual characteristics.

Prior to analysis, each dataset was formatted consistently and differences in the measurement or recording of data accounted for. Only main stems of living woody plants with accurate coordinates, diameters and species identities were included. We also used elevation data gathered at the corners of 20 x 20 m sub-plots for the derivation of environmental metrics, as this was the finest resolution of the topographic survey scale that was available consistently for all plots.
<table>
<thead>
<tr>
<th>Location</th>
<th>Size (ha)</th>
<th>Mean elevation (m)</th>
<th>Elevation range (m)</th>
<th>Rainfall (annual mean; monthly range, mm)</th>
<th>Temp. range (monthly averages,°C)</th>
<th>No. of trees; no. of species</th>
<th>Forest type</th>
<th>Disturbance</th>
<th>Census year</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasoh</td>
<td>Malaysia 2°58’N, 102°18’E</td>
<td>50</td>
<td>97</td>
<td>25</td>
<td>1571; 94-224</td>
<td>21.9 - 34.6</td>
<td>378,186; 822</td>
<td>lowland mixed dipterocarp forest</td>
<td>Low; small (&lt;1ha) windthrows</td>
<td>1995</td>
</tr>
<tr>
<td>Yasuni</td>
<td>Ecuador 00°41’S, 76°23’W</td>
<td>25</td>
<td>228</td>
<td>32</td>
<td>3081; 174-412</td>
<td>21.2 - 36.6</td>
<td>173,172; 1088</td>
<td>evergreen lowland wet forest</td>
<td>low; occasional wind storms</td>
<td>1999</td>
</tr>
<tr>
<td>BCI</td>
<td>Panama 9°9’N, 79°51’W</td>
<td>50</td>
<td>144</td>
<td>40</td>
<td>2551; 23-364</td>
<td>22.8 - 23.7</td>
<td>208,387; 299</td>
<td>semideciduous lowland moist forest</td>
<td>low; rare windstorms, droughts</td>
<td>2010</td>
</tr>
<tr>
<td>Luquillo</td>
<td>Puerto Rico; 18°19’N, 65°49’W</td>
<td>16</td>
<td>378</td>
<td>93</td>
<td>3548; 203-401</td>
<td>18.7 - 22.0</td>
<td>76,515; 140</td>
<td>tropical montane / subtropical wet forest</td>
<td>medium; regular windstorms &amp; hurricanes</td>
<td>2000</td>
</tr>
<tr>
<td>Fushan</td>
<td>Taiwan, 24°45’N 121°33’E</td>
<td>25</td>
<td>675</td>
<td>133</td>
<td>4271; 149-608</td>
<td>11.8 - 24.0</td>
<td>114,354; 110</td>
<td>moist subtropical mixed evergreen</td>
<td>high; frequent typhoons, wind</td>
<td>2003</td>
</tr>
<tr>
<td>Location</td>
<td>Country</td>
<td>Latitude, Longitude</td>
<td>Size</td>
<td>Height</td>
<td>Age</td>
<td>Dominant Forest Type</td>
<td>Damage &amp; Landslides</td>
<td>Year</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>------</td>
<td>--------</td>
<td>-----</td>
<td>----------------------</td>
<td>---------------------------------------------</td>
<td>------</td>
<td>------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Mudumalai</td>
<td>India; 11°35'N, 76°31'E</td>
<td>50 x 1035 x 130 x 1200</td>
<td>14.2 - 20.1</td>
<td>30,786; 73</td>
<td>dry/moist deciduous forest</td>
<td>medium; fires, browsing and some cyclones</td>
<td>1988</td>
<td>Sukumar et al. 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sinharaja</td>
<td>Sri Lanka; 6°24'N, 80°24'E</td>
<td>25 x 471 x 150 x 5016; 171-695</td>
<td>19.3 - 26.8</td>
<td>378,186; 206</td>
<td>mixed dipterocarp forest</td>
<td>low; evidence of rare but substantial wind damage</td>
<td>2002</td>
<td>Gunatileke et al. 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xishuangbanna</td>
<td>China; 21°36'N, 101°34' E</td>
<td>20 x 765 x 150 x 1493; 20-320</td>
<td>15.6-25.3</td>
<td>95,940; 469</td>
<td>tropical seasonal rainforest</td>
<td>low</td>
<td>2007</td>
<td>Cao et al. 2006; Lin et al. in press</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lienhuachih</td>
<td>Taiwan, 23°34'N, 120°52' E</td>
<td>25 x 762 x 178 x 2285; 21-483</td>
<td>14.8-25.2</td>
<td>135,268; 144</td>
<td>subtropical evergreen broadleaved forest</td>
<td>medium; typhoons &amp; landslides</td>
<td>2008</td>
<td>Chang 2010; Lu et al. 2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dinghushan</td>
<td>China, 23°09'N, 112°30' E</td>
<td>20 x 338 x 246 x 1985; 40-320</td>
<td>12.6 - 28.0</td>
<td>71,617; 210</td>
<td>subtropical evergreen broadleaved</td>
<td>medium; regular storms &amp; typhoons</td>
<td>2005</td>
<td>Zhou &amp; Yan 2001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gutianshan</td>
<td>China, 29°15'N, 118°07' E</td>
<td>24 x 580 x 269 x 1787; 70-330</td>
<td>4.7–27.6</td>
<td>140,700; 159</td>
<td>subtropical evergreen broadleaved</td>
<td>medium; rare but destructive storms</td>
<td>2005</td>
<td>Du et al. 2009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Summaries of the plots included in the analyses
3.2.2 SPATIAL STATISTIC (CROSS-PAIR OVERLAP DISTRIBUTION)

To summarise spatial pattern in the different plots, we use the cross-pair overlap distribution (xPOD; Chapter 2). This measure describes the spatial overlap of all species in a community as a histogram of scalar quantities $A_{ij}$, each of which is the area under the log of the cross-pair correlation function for one pair of species. (The cross-pair correlation function describes how the spatial relationship of pairs of species varies with radius, taking a value of one (zero on the log scale) when they are distributed independently of one another, and taking higher or lower values when they are more or less aggregated than this (Law et al. 2009)) (tests of the significance of the departure from independence are possible (e.g. Wiegand & Moloney 2004; Wiegand et al. 2007b)). The area is computed as

$$A_{ij} = \int_{0}^{R} \log(g_{ij}(r))dr$$, eqn 3.1

where $r$ is radius, $R$ is the maximum radius over which the calculation is performed and $g_{ij}$ is the cross-pair correlation function of species $i$ and $j$. In Chapter 2, the xPOD was defined for use over a radius $R$ of one quarter of minimum plot dimension following the recommendation of Baddeley & Turner (2005). In order to avoid violating this in the smallest plots, we set a value of $R$ of 80 m here. An $A_{ij}$ value of zero implies that the two species are randomly distributed relative to one another, a positive value indicates that they overlap in space, and a negative value indicates that they are separate. The xPOD is the frequency distribution of these $A_{ij}$s over an assemblage of species and provides information about the consistency of behaviour across species pairs. A distribution with a small range suggests that all species have a similar degree of overlap, while a distribution with a large range suggests that some species pairs overlap to a far greater or lesser extent than others. The xPOD is calculated here so that edge effects do not occur (see Appendix 3.1). In measuring the spatial overlap of species pairs, the xPOD is also normalised to differences in overall density and diversity and is not affected by species characteristics.

The choice of spatial summary statistic is crucial if meaningful information is to be retained at the level of a community, as in this study. We use the xPOD because it has been shown to discriminate reliably between neutral and niche processes in the simulation study presented in Chapter 2. Simulated communities generated under the assumptions of the neutral model possessed a distribution of values of $A_{ij}$ with a slightly negative mean and small spread, indicative of the consistent separation between species produced by dispersal limitation and density dependent mortality. Under niche differentiation, however, distributions with a more negative mean and considerably greater spread were produced, as species separated out along a geographically-defined environmental gradient. Those species with similar environmental niches were found to overlap, while those with very different niches were highly separate in space. Because the principal difference between xPODs produced by neutral and niche simulation models was in their spread, we summarise the xPODs of each forest plot by their standard deviation. This allows us to generate a single value for each site characterising the spatial behaviour of species within it, which can be compared directly to environmental metrics.

Figure 3.1 illustrates the links between spatial pattern and xPOD mean and spread. Figures 3.1a and 1b show the locations of all individuals of two species at the Lienhuachih plot in Taiwan. These are clearly separate in space, and so their log cross-pair correlation function is below zero (Fig. 1e). In Figures 3.1c and 3.1d, in contrast, two other species that overlap in space are shown, and their log cross-pair correlation function is above zero (Fig. 3.1f). As a result, the areas under these functions are negative and positive respectively. When combined in a frequency distribution of values representing all possible species-pairs on the Lienhuachih plot, the respective positions of these
scalar quantities are apparent, and are indicated in the complete xPOD for Lienhuachih in Figure 3.1g.

Figure 3.1: The contribution of species pairs to the xPOD for Lienhuachih.

The species shown in (a) and (b) are separate in space, while the species in (c) and (d) are highly coincident, both occurring along ridges. Plot (e) shows the log of the cross-pair correlation function for species (a) and (b), which is below the random line at all radii, indicating their separation; plot(f) shows the corresponding plot for species (c) and (d), which is above the random line at all radii. These are summarised in the xPOD of the whole plot (g) by the area under the line relative to the random line at $x = 0$, and these areas appear in the final plot where indicated.
The xPOD was defined for use with a minimum abundance threshold of 500, so that uninformative signals from rare species (which, by their nature, tend to be highly separate in space) did not dominate the result. We observe this convention but also calculate xPODs at a lower abundance threshold of 100 individuals, in order to ensure that sufficient species are included to produce clear distributions. Minimum diameter at breast height (dbh) thresholds of 1 cm, to include all recorded trees in the datasets, and a variable threshold to isolate the signal of adult trees were applied in separate analyses. This was done because the locations of adult trees are expected to be more informative about topographic niche preferences, being less subject to partially stochastic mortality effects than seedlings and saplings (Brokaw & Busing 2000; Hu et al. 2011). Following Bagchi et al. (2011), adult trees were defined as those with a dbh greater than the 99th percentile dbh for that species raised to a power of $\frac{2}{3}$. This is a biologically-motivated and conservative threshold, thought to be successful in excluding non-reproductive juveniles from the adult class. It is used in place of any other arbitrary threshold as it is variable, allowing for differences between species’ growth forms, and also because it is less likely to classify juveniles as adults than some commonly-used alternatives (e.g. Condit et al. 2000).

3.2.3 ENVIRONMENTAL METRICS AND REGRESSION

We summarised environmental variation in each plot in terms of topography, which is the only environmental data consistently available across all plots. It is also strongly correlated with variation in resources such as water availability and soil conditions, and a potentially important axis of niche differentiation in plant communities (Wright 2002; Silvertown 2004; Vormisto et al. 2004; Ediriweera et al. 2008). Both the total range and the spatial scale of topographic variation may affect patterns of overlap between species’ niches, and so we used five metrics designed to capture different aspects of these: the range and standard deviation of elevation; the surface to planimetric area ratio; the proportion of convex grid cells; and the average slope between grid cells. These metrics are summarised in Table 3.2. Because of the possibility that regular diurnal variation in cloudiness influences irradiance regimes as a function of aspect, we also summarised differences in aspect between plots in terms of sub-plot aspects (including the dominant aspect and variance in aspect). We used linear regression to examine the relationships between each measure of environmental heterogeneity and the standard deviations of the xPODs across plots.

We also checked for any relationships between biogeographical factors and xPOD standard deviations using linear regression. No clear predictions about these relationships exist, and because it is also unclear how the effects might combine or interact, we performed a single regression for each. By chance, the latitude of plots was found to correlate with elevation range (see Fig. A3.5), and so any relationship between spatial structure and elevation range is expected to produce a similar apparent relationship with latitude. No further relationship between spatial structure and latitude was expected. Finally, we checked for relationships between the topography metrics and the moments of each plot’s species abundance distribution (the mean, standard deviation and skewness), and the Shannon diversity index for each plot. This was to determine whether environmental heterogeneity has any systematic effect on relative abundances, as it is expected to on spatial structure. We predicted that no relationships of this kind would be found.
<table>
<thead>
<tr>
<th>Name</th>
<th>Identity</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation range</td>
<td>( \max(z) - \min(z) )</td>
<td>range</td>
</tr>
<tr>
<td>Elevation standard deviation</td>
<td>( sd(z) )</td>
<td>range; roughness</td>
</tr>
<tr>
<td>surface: planimetric area ratio</td>
<td>following method in Jenness (2004)</td>
<td>range; roughness</td>
</tr>
<tr>
<td>Proportion of convex cells</td>
<td>( n_{\text{convex}} = \frac{n_{\text{cells}} {z_i &gt; z_j : j = 1: n_n }}{n_{\text{cells}}} )</td>
<td>roughness</td>
</tr>
<tr>
<td>Mean slope</td>
<td>( \bar{s}<em>{\text{cell}} = \frac{\sum</em>{i=1}^{n_{n}} s_{\text{diff}}^{i}}{n_n}, \bar{s} = \frac{\sum_{i=1}^{n_{n}} \bar{s}_{\text{cell}}}{n_e} )</td>
<td>roughness</td>
</tr>
</tbody>
</table>

Table 3.2: The names, identities and characteristics described by the environmental metrics calculated for each plot.

- \( z \) = elevation
- \( z_i \) = elevation of cell \( i \)
- \( z_j \) = elevation of the \( j \)th neighbour cell of cell \( i \)
- \( n_e \) = total number of cells
- \( n_n \) = number of neighbours of each cell ( = 6)
- \( z_{\text{diff}} \) = difference in elevation between two (neighbouring) cells
- \( s_{\text{diff}} \) = horizontal surface distance between two cells

### 3.3 Results

Figure 3.2 shows the topography and xPOD for Pasoh and Gutianshan, the two plots at opposite ends of the range of physical heterogeneity (the corresponding results for the other plots are presented in Figure A3.1). As indicated by these plots, the standard deviations of the xPODs increase with increasing elevation range. This pattern is observed at all the abundance and dbh thresholds we applied, but is found to be strongest at an abundance threshold of 100 individuals and a dbh threshold of 1 cm (Fig. A3.2). Figure 3.3 shows the relationships of the xPOD standard deviation with elevation range, elevation standard deviation and average slope (all significant at 5% level by linear regression). Two remaining non-significant relationships are shown in Figure A3.3.
Figure 3.2: Topographies and cross-pair overlap distributions for Pasoh ((a), (c)) and Gutianshan ((b), (d)).

The y-scales on each plot are identical (0-270 m), but the colour scales only span the range of the data, 25 m in Pasoh and 270 m in Gutianshan
Figure 3.3: Significant relationships between cross-pair overlap distribution standard deviations and environmental metrics.

Fitted regression lines and their associated p-values are shown. Plots are identified as follows:

BCI = Barro-Colorado Island, Di = Dinghushan, Fu = Fushan, Gu = Gutianshan, Li = Lienhuachih, Lu = Luquillo, MS = Mo Singto
Mu = Mudumalai, Pa = Pasoh, Si = Sinharaja, Xi = Xishuangbanna, Ya = Yasuni

p = 0.0128
p = 0.0141
p = 0.0449
Both the standard deviations of the xPODs and the residuals of the significant models listed above were additionally plotted and regressed against biogeographical factors including mean elevation, dominant aspect, latitude, mean annual rainfall, maximum monthly average rainfall difference, maximum monthly temperature difference, species diversity and stem density (Fig. A3.4). The only significant relationship was between xPOD standard deviation and plot latitude (Fig. A3.4a, p-value = 0.0252). This relationship is presumably driven by the fact that the plots at lower latitudes are, by chance, more homogeneous than those at higher latitudes (Fig. A3.5), therefore being a consequence of the relationship between physical heterogeneity and xPOD standard deviation identified above. We also checked for plot-size effects by deriving environmental metrics and xPODs separately for each half of the three 50 ha plots (Pasoh, BCI and Mudumalai) and including these in the regressions as independent points. In every case they strengthened the existing relationships, but this was largely attributable to the increase in the number of data points. Regressions of the moments of each plot’s species abundance distribution and Shannon diversity index against environmental metrics produced no significant results (Table A3.1).

3.4 Discussion

This chapter was intended to test the hypothesis that the spatial overlap of species should vary with physical environmental heterogeneity, as suggested by niche theory. We have shown that the hypothesis is supported in a test involving 2,016,263 spatially mapped woody plants on twelve plots located on three continents. This result is incompatible with neutral theory, which predicts that no systematic relationship between species overlap and environmental heterogeneity exists.

Datasets used in this study span almost the entire geographic range of tropical forests, varying greatly in their location, elevation, rainfall, seasonality and disturbance regimes, each of which is known or expected to have profound consequences for forest ecology (Davidar et al. 2007; Wiens 2011). The clear relationship shown between spatial overlap of species and environmental heterogeneity, despite these biogeographical differences, suggests that niche differentiation linked to topography is a pervasive property of tree species in tropical rainforests.

The standard deviation of xPODs was found to be significantly positively related to the range and standard deviation of elevation, and to average slope. No relationship was found with the proportion of convex cells or the surface to planimetric area ratio. The fact that three out of five relationships were positive at the 5% significance level, while none were negative, strongly suggests that type I errors did not affect our findings. The relationship with elevation range is perhaps the most intuitive, relating directly to the geographical spread of species via their environmental niches. It is important to note that all of the plots included here contain reasonably regular gradients between extremes of elevation, with none containing, for example, two relatively homogenous areas separated by very steep ground (which might be expected to give a bimodal xPOD distribution). Given this, the relationship between spatial structure and elevation range is as expected under niche differentiation, with species separating out across a gradient according to their environmental preferences. It also conforms to previous findings of changes in tropical rainforest species composition along single altitudinal gradients at small (Lieberman et al. 1985; Sri-Ngernyung et al. 2003) and large (Lieberman et al. 1996; Vázquez & Givnish 1998) scales.

The effects of standard deviation of elevation and mean slope show that the ‘roughness’ of the environment is also, as expected, a strong determinant of spatial structure (Chapter 2). In this case, the scale of physical variation determines the spatial extent of niche, and hence species, overlap. The standard deviation of elevation is not a pure measure of roughness, but may be seen as an intermediate measure between that of overall range and that of mean slope. As such, these
relationships together provide a consistent picture of strong niche differentiation along elevation gradients.

We are not suggesting that elevation is directly responsible for the spatial overlaps of species. A number of environmental variables are likely to be correlated with elevation, such as soil nutrient and water availability, and light regimes (Becker et al. 1988, Ashton 1992, Daws et al. 2002, Baker et al. 2003, John et al. 2007). Each of these might vary closely with elevation and so any or all could be responsible for the observed effects. The lack of an effect of aspect, however, perhaps indicates that light regimes are either largely independent of aspect (being more related to disturbances and canopy gaps, for example) or that their influence is limited to certain sites. In any case, it appears that the total range of potential habitats, captured if not explained by simple measures of elevation, is a stronger determinant of spatial structure at the scale over which we measure it than finer environmental variations within that range.

We predicted that the signal of niche differentiation would be stronger among adult trees, because the inherent stochasticity in gap formation, seed arrival and germination mean that recruitment may be effectively neutral despite any niche differentiation at early life stages (Brokaw & Busing 2000). Although none of our analyses include seedlings (the minimum dbh threshold being 1 cm), we still expect the association of trees and environments that are favourable to them to increase as they grow (Condit et al. 2000; Bagchi et al. 2011). Instead, the relationships we find are strongest when all trees belonging to species with at least 100 individuals are included, and not when our analysis is restricted to adult trees or to more abundant species. This may be because environmental factors exert a similar influence on young trees as on adults, so that niche effects are established by the time a tree’s diameter reaches 1 cm (e.g. Comita & Engelbrecht 2009; Brenes-Arguedas et al. 2011), or because limited dispersal around adult trees simply leads to a reinforcement of the existing, environmentally-driven, spatial pattern. It is also possible that the loss of information when a higher dbh threshold is used leads to more noise in the signals, which would explain the fact that relationships are less clear under higher abundance thresholds. This would be exacerbated if rarer species were more strongly specialised to particular niches (Brown 1984).

We find no evidence that the effects of niche differentiation are discernible in species abundance patterns, with no significant relationships found between environmental metrics and the moments of each plot’s species abundance distribution (SAD). This is consistent with the findings of Chapter 2, and suggests that the inclusion of spatial information is of great value in studies of this kind. Neutral drift is an adequate explanation of first-order structure in these plots, but not of second-order structure.

Previously, mixed evidence has been found for niche differentiation in the plots that we include. Neutral speciation and dispersal processes are regularly found to account for at least some observed spatial structure (e.g. Condit et al. 2000; Seidler & Plotkin 2006; Horvat et al. 2010; Zhang et al. 2011). Niche differentiation is often found in a minority of species (Harms et al. 2001), or only at particularly large (Bin et al. 2010) or small spatial scales (Svenning et al. 2011). Spatial point pattern modelling suggests a similar mix of processes, with niche differentiation being identified most clearly in forests with high levels of physical heterogeneity (Potts et al. 2004; Shen et al. 2009; Lin et al. 2011). For example, Zhang et al. (2011) found that 80% of species in the highly heterogeneous Gutianshan plot were associated with particular soil properties, and Hu et al. (2011) that most species in Xishuangbanna had associations with several environmental factors which increased with plant age. Gunatileke et al. (2006) found that the majority of species at Sinharaja were habitat specialists, as did Condit et al. (2000), who nevertheless found little evidence of niche differentiation in Mudumalai, in agreement with the findings of John & Sukumar (2004). Kraft & Ackerly (2010), though, found widespread niche differentiation in the homogeneous Yasuni plot, and Debski et al. (2002) found evidence of it in all of the 12 species they considered in Pasoh.
It is difficult to draw broad conclusions from these studies because they use different methods, consider different variables, proportions of the populations and areas of each plot, and look for niche differentiation on different axes. Our approach, in contrast, produces a trend across all communities that integrates data-sets from plots under different biogeographical conditions. This allows us not only to test the compatibility of spatial patterns with the predictions of neutral theory at very large scales, but to assess how the strength of niche differentiation varies with environmental heterogeneity. More restricted comparisons of several plots have previously identified some signals of environmental structuring or deterministic coexistence mechanisms (Condit et al. 2000; Wills et al. 2006; John et al. 2007; Swenson et al. 2011) and our results are consistent with these, demonstrating that neutral theory alone cannot account for observed variation in spatial structure.

We also find, however, that the strength of topographical niche differentiation clearly increases with environmental heterogeneity, being strongest in plots such as Gutianshan and Dinghushan and weakest in Pasoh, Yasuni and BCI. Additional processes, whether neutral or related to other effects (such as the Janzen-Connell effect, which our findings do not speak to but which are investigated further in Chapter 5) may be occurring in the plots, as may niche differentiation in ways that we do not detect, for example temporally, at smaller scales, or at early life stages (e.g. Grubb 1977; Schulman et al. 2004; Comita & Engelbrecht 2009; Kanagaraj et al. 2011). Nevertheless, our observation of changes in spatial structure with environmental heterogeneity over a wide range of tropical biogeographical conditions is not consistent with a fully neutral explanation. A parsimonious interpretation is that the increasing range of spatial overlaps with increasing heterogeneity stems from niche differentiation of the species. We can conclude that environmental heterogeneity is an important driver of spatial differentiation and hence of species coexistence in tropical rainforests, especially in heterogeneous environments.
Appendix 3.1: Calculation of cross-pair overlap distribution (xPOD)

To avoid edge effects without the need to discard data or impose additional correction terms, we develop the cross-pair overlap distribution (xPOD) here to take account of the area over which its constituent data are gathered. The estimated cross-pair correlation function for species \( i \) and \( j \) at radius \( r \), \( \hat{g}_{ij}(r) \), is a ratio of the observed number of members of species \( j \) within annuli of areas \( a(r) \) around the members of species \( i \) to the number expected given the total area of these annuli and the average density \( \lambda_j \) of individuals of species \( j \) per unit area. Here we use only the fraction of each \( a(r) \) that falls within the plot in the calculations. This ensures that apparently empty areas beyond the plot boundaries do not influence the result, while all of the data within the plot is used and given equal weighting:

\[
\hat{g}_{ij}(r) = \frac{\sum_{k=1}^{s} n_{ja_k(r)}}{\lambda_j \sum_{k=1}^{s} a_k(r)}
\]

Where \( s \) is the number of individuals in species \( i \), and \( n_{ja_k(r)} \) is the number of individuals of species \( j \) found within the annulus of area \( a_k(r) \) at radius \( r \) around the \( k_{th} \) member of species \( i \).
3.6 Appendix 3.2: Additional plots and tables

Pasoh
25 m elevation range

Yasuni
32 m elevation range

BCI
40 m elevation range

Luquillo
90 m elevation range

Mo Singto
93 m elevation range

Fushan
130 m elevation range

Mudumalai
130 m elevation range

Sinhara
150 m elevation range

Xishuangbanna
150 m elevation range

Lienhuachih
174 m elevation range

Dinghushan
246 m elevation range

Gutianshan
269 m elevation range
Figure A3.1: Topographies (above) and cross-pair overlap distributions (below) for all plots.

Cross-pair overlap distributions are calculated with a dbh threshold of 1cm and an abundance threshold of 100 individuals.
Figure A3.2: Relationships between xPOD standard deviations and elevation range in each plot with dbh and abundance thresholds of (a) 1 and 100, (b) 1 and 500, (c) species-specific adult dbh thresholds (see methods) and abundance of 100.

Fitted regression lines and their associated p-values are shown.

BCI = Barro-COLORADO Island, Di = Dinghushan, Fu = Fushan, Gu = Gutianshan, Li = Lienhuachih, Lu = Luquillo, MS = Mo Singto, Mu = Mudumalai, Pa = Pasoh, Si = Sinharaja, Xi = Xishuangbanna, Ya = Yasuni.
Figure A3.3: Relationships, fitted regression lines and their associated p-values for each of the environmental metrics against standard deviation of cross-pair overlap distributions.

BCI = Barro-COLORADO Island, Di = Dinghushan, Fu = Fushan, Gu = Gutianshan, Li = Lienhuachih, Lu = Luquillo, MS = Mo Singto, Mu = Mudumalai, Pa = Pasoh, Si = Sinharaja, Xi = Xishuangbanna, Ya = Yasuni.
Figure A3.4: (previous page) Standard deviations of cross-pair overlap distributions against biogeographical variables.

(a) Latitude, (b) dominant aspect, (c) maximum monthly temperature difference, (d) average annual rainfall, (e) species density (ha$^{-1}$), (f) stem density (ha$^{-1}$), (g) mean elevation, (h) maximum monthly temperature difference. The only significant relationship is with latitude, and the regression line and associated p-value are shown. This relationship is assumed to be driven by the strong but arbitrary relationship between latitude and elevation range of the plots (Fig. A3.5).

BCI = Barro-Colorado Island, Di = Dinghushan, Fu = Fushan, Gu = Gutianshan, Li = Lienhuachih, Lu = Luquillo, MS = Mo Singto, Mu = Mudumalai, Pa = Pasoh, Si = Sinharaja, Xi = Xishuangbanna, Ya = Yasuni.

Figure A3.5: The chance relationship between latitude and elevation range of plots.

BCI = Barro-Colorado Island, Di = Dinghushan, Fu = Fushan, Gu = Gutianshan, Li = Lienhuachih, Lu = Luquillo, MS = Mo Singto, Mu = Mudumalai, Pa = Pasoh, Si = Sinharaja, Xi = Xishuangbanna, Ya = Yasuni.
<table>
<thead>
<tr>
<th></th>
<th>SAD Mean</th>
<th>SAD standard deviation</th>
<th>SAD skewness</th>
<th>Shannon index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation range</td>
<td>0.491</td>
<td>0.461</td>
<td>0.211</td>
<td>0.085</td>
</tr>
<tr>
<td>Elevation standard deviation</td>
<td>0.509</td>
<td>0.452</td>
<td>0.261</td>
<td>0.117</td>
</tr>
<tr>
<td>Average slope</td>
<td>0.415</td>
<td>0.347</td>
<td>0.397</td>
<td>0.227</td>
</tr>
<tr>
<td>Area ratio</td>
<td>0.536</td>
<td>0.384</td>
<td>0.276</td>
<td>0.391</td>
</tr>
<tr>
<td>Proportion of convex cells</td>
<td>0.637</td>
<td>0.190</td>
<td>0.715</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Table A3.1: p-values for single linear regressions of environmental metrics against moments of species abundance distributions (SAD) and the Shannon diversity index for each plot.
CHAPTER 4 - Spatial statistics for the inference of underlying process in plant communities
4.1 Introduction

This chapter builds on the work presented in Chapter 2 to attempt to fully describe the spatial consequences of the ecological processes modelled there, and to determine which of a range of spatial statistics are most sensitive to those consequences.

4.1.1 SPATIAL STATISTICS

Statistics that summarise spatial relationships between individuals are of great potential value in ecology, where many processes influence, and are influenced by, spatial structure (Watt 1947; Greig-Smith 1961; Bolker & Pacala 1997; Law et al. 2009; Chapter 1). As discussed previously, these processes include niche differentiation (Wang et al. 2011), positive and negative interactions between species (Stoll & Prati 2001), negative density dependence effects within species (Clark & Clark 1984; Packer & Clay 2000), and dispersal limitation (Seidler & Plotkin 2006). All of these, alone or in combination, are thought to play a substantial role in the dynamics of plant communities, and the detection and quantification of such processes is a major aim of forest ecology in particular.

Of particular interest are processes that might enable very large numbers of species to coexist in tropical rainforests (Wright 2002; Chapter 1). The long-running debate over the identity of these processes has been given fresh impetus in recent years by findings that the observed first-order characteristics of rainforests may be matched by neutral models (Hubbell 1979; 2001; Bell 2001). Attempts to ascertain the role of neutral dynamics, as opposed to the more established niche differentiation (Grinnell 1917; Hutchinson 1958) or Janzen-Connell effects (Janzen 1970; Connell 1970), have been manifold (e.g. Hardy & Sonké 2004; Silvertown 2004; Chase 2005; Ricklefs 2006; Jabot et al. 2008; Zhou & Zhang 2008; Ruokolainen et al. 2009; Tang & Zhou 2011). While popular first-order summaries of species occurrence, diversity or abundances have failed to provide a resolution (e.g. Hubbell 2001; Chave 2004; McGill et al. 2007; Chapters 1 and 2), it is hoped that summaries of spatial structure may do so (Chapter 2; Holyoak & Loreau 2006).

The majority of spatial statistics used in ecology describe intra- or inter-specific clustering that has clear ecological relevance. The use of such statistics began with tests for non-random associations between or within species (Clapham 1936; Greig-Smith 1961; Taylor, 1971; Haase 1995), and continued through various quantifications of these (Clark & Evans 1954; Hurlbert 1971; Iwao 1977). Spatial point process theory, which provides statistical models and descriptions of patterns formed by interacting or non-interacting individuals, has become increasingly important as a basis for analyses of this kind (Matérn, 1960; Stoyan & Penttinen 2000). Ripley's K-function and its non-cumulative equivalent, the pair correlation function, have been particularly widely used (Salonen et al. 1992; Haase 1995; Hardy & Sonké 2004; Wiegand & Moloney 2004; Wiegand et al. 2007b; Illian et al. 2008; Chapter 2). Meanwhile, the spatial variations in diversity implied by various forms of clustering have been investigated through measures of β-diversity (Whittaker 1960; 1972), which summarise some aspect of the turnover in species composition with site (e.g. Jaccard 1912; Odum 1950; Whittaker 1952; Brown et al. 1995; Legendre & Legendre 1998; Nekola & White 1999; Koleff et al. 2003; Legendre et al. 2005).

Spatial statistics have often proved successful in identifying the spatial signals of particular ecological processes or environmental variation (Mladenoff et al. 1993; Bolker & Pacala, 1997; Law et al. 2009; Wang et al. 2010), or in studying associations and interactions between particular species (Wiegand et al. 2007b; Martinez et al. 2010; Zhang et al. 2010; Wang et al. 2011). The work presented in this thesis and other recent studies suggest that spatial structure may indeed be used to distinguish neutral and non-neutral processes (Tuomisto & Ruokolainen 2006; Münkemüller et al. 2011; Chapter 2), and dispersal limitation and niche differentiation have both been frequently investigated through their spatial signals (e.g. Condit et al. 2002; Dray & Legendre 2008; Paoli et al. 2006; Smith &
Nevertheless, patterns generated by environmental variation and limited dispersal remain largely indistinguishable (Pickett & Cadenasso 1995; Tuomisto et al 2003; Wiegand et al 2007a; Kraan et al. 2010), although advanced techniques for spatial point process modelling offer promise for separating the effects (e.g. Rue et al. 2009; Simpson et al. 2011; Illian et al., in press), as do increasingly sophisticated environmental summaries (e.g. Kharuk et al. 2010; Obertegger et al. 2010; Leng et al. 2010).

4.1.2 SPATIAL STRUCTURE

A focus on particular applications, however, has led to a proliferation of statistics that are based on similar or identical spatial information. Many of these statistics are intended to capture the spatial signals of mechanisms of coexistence, such as intraspecific patterns related to limited dispersal or Janzen-Connell effects, or interspecific patterns related to competition or niche differentiation. Despite this, reviews of the relative abilities of different spatial statistics to describe patterns of interest have been scarce (Dale et al. 2002; Koleff et al. 2003; Tuomisto 2010a; 2010b), and the relationships between them have not been formally assessed. A further consequence of this is that the spatial signals of important processes have not been fully investigated or characterised.

Here, we carry out a comprehensive assessment of the sensitivity of a number of different statistics to the five principal theorised mechanisms of species coexistence in plant communities: neutral dynamics, environmental niche differentiation, temporal niche differentiation, the Janzen-Connell effect, and heteromyopia. In order to do so, we isolate the signals of these processes by running stochastic simulations of forest communities under a model of each process. Our primary aim is to investigate the sensitivity of different statistics to the modelled processes. This is expected to aid in the development of empirically testable predictions that differ between theorised coexistence mechanisms. A secondary aim is to fully characterise the spatial consequences of the processes that we model. Any one statistic is unlikely to capture the full spatial signal of a particular process and, being a summary of spatial structure, may discard some relevant information. It is therefore necessary to use a number of different statistics to completely characterise spatial patterns as we wish to here.

Specifically, spatial statistics can measure structure formed by some form of attraction or repulsion among individuals, where random distributions indicate a lack of interaction (Wiegand & Moloney 2004). This attraction or repulsion can occur between or within species, and we divide spatial statistics according to which of these combinations they describe. Those that focus on spatial structure within species are expected to be particularly sensitive to intraspecific processes, while those that focus on spatial structure between species are expected to be similarly sensitive to interspecific processes.

These two classes of spatial structure are referred to, respectively, as ‘scattering’ and ‘exposure’ in the literature on spatial statistics (Rajala & Illian, in press). Patterns of scattering vary between clustering and regularity, while patterns of exposure vary between mingling and segregation (Fig. 4.2). Low levels of scattering indicate that conspecifics are non-randomly clumped (or clustered), while high levels indicate non-random dispersion (or regularity) of individuals. Similarly, low levels of exposure indicate non-random segregation of species and high levels indicate non-random mingling.

Scattering and exposure are not directly measurable properties of spatial patterns, and both may be described in a number of different ways. Because of this, it is difficult to gauge the success of a single spatial statistic in summarising the signal of any given process (e.g. Chapter 2). We therefore identify the individual pieces of information that comprise each statistic, and use these to develop a comprehensive summary of the spatial signals of modelled processes. This enables us to propose a
highly informative combination of existing spatial statistics which together capture all the principal
spatial consequences of each process, and may allow for more confident inference, or rejection, of
underlying process from resulting spatial patterns in plant communities.

4.2 Methods

4.2.1 SPATIAL STATISTICS

We do not attempt a comprehensive review of published spatial statistics, but instead concentrate
on those that have proved popular, successful in their original purpose, or are particularly
representative of a group of similar statistics. Altogether, we consider three statistics describing
scattering and five describing exposure (these are summarised in Table 4.2. Twenty measures of β-
diversity were also used to assess how species turnover relates to scattering and exposure; these
results are presented in Appendix 4.1). The statistics of scattering and exposure comprise
combinations of different individual pieces of information (Table 4.1), and together incorporate all of
the information used in other relevant statistics. While there are many of these statistics, they all
make use of relatively few such pieces of information, which are generally simple counts of or
distances to neighbours. We locate these individual pieces of information within a multi-
dimensional framework describing the sensitivity to species identity (conspecific or heterospecific),
scale (scale-dependent or independent) and organisational level (individual, species or community)
at which they operate (Fig. 4.1). This allows us to identify more precisely where and in what form
the spatial consequences of each process manifest themselves.

4.2.1.1 Scattering
We consider three measures describing scattering (within-species structure): the degree of
aggregation (Coomes et al. 1999); the measure of interspecific segregation (Dixon 1994, which,
despite its name, describes scattering rather than exposure); and the proportion of conspecific
neighbours (e.g. Yurkonis et al. 2012) (Table 4.2a). All operate at the species or community level,
although it is possible to calculate the proportion of conspecific neighbours at the individual level.
The measure of interspecific segregation is the only wholly scale-independent measure, making use
of nearest neighbour identities and species abundances (Fig. 4.1). The degree of aggregation, in
contrast, is scale-dependent but makes use only of information relating to conspecifics.

The proportion of conspecific neighbours and the measure of interspecific segregation are sensitive
to the abundances of heterospecific individuals due to the inclusion of terms dependent upon local
and total numbers of individuals respectively. This makes them indirectly sensitive to community
abundance and diversity, neither of which necessarily affect spatial structure. As a result, the degree
of aggregation is expected to provide the most intuitive and sensitive description of patterns
produced by different spatial processes, as it is not affected by community size or diversity and is
expressed relative to a random distribution of points.

4.2.1.2 Exposure
Five measures describe exposure of species: the individual species-area relationship (ISAR) (Wiegand
et al. 2007a); the mingling index (Lewandowski & Pommerening 1996; Graz 2004; Rajala & Illian in
press); the spatial Simpson index (Shimatani 2001; Rajala & Illian in press), the degree of association
(Coomes et al. 1999) and the cross-pair overlap distribution (xPOD) (Chapter 2). Of these, the
simplest is the ISAR. Originally defined via the sum of bivariate emptiness probabilities $P_{ij}(0,r)$ that a
circle of radius $r$ centred on an individual of species $j$ does not contain an individual of species $k$, it
can also be expressed as the count of species in circles around a member of species $i$. Its value can
therefore be found at the individual, species or community level. Intended to detect species which
‘attract’ and ‘repel’ diversity in their surroundings, it is also useful in characterising the variation in spatial diversity across a community.

The mingling index is a simple mean proportion of heterospecifics in circles of radius $r$, equivalent to one minus the mean proportion of conspecific neighbours (and so is also sensitive to scattering). Once again, it is possible to calculate this index at individual, species or community level. The spatial Simpson index, in contrast, is a measure of the diversity of an entire community and, because it includes the product of local and global proportions of individuals belonging to a particular species, does not exist at any other level.

The two remaining measures of inter-specific exposure, the xPOD and degree of association, are defined at the level of pairs of species. Both make use of similar information to form a scale-dependent ratio of observed and expected numbers of neighbours around members of species $j$ belonging to another species $k$. The only difference in their information content is that the neighbour counts used in the xPOD are theoretically made at radius $r$ (in practice within a range $r + dr$) and later integrated over a set radius $R$, while those used in the degree of association are within radius $r$, as is the case with all other measures used here.

The ability of the xPOD to distinguish the processes modelled here has previously been established (Chapter 2), and the degree of association is expected to be similarly sensitive due to the information it shares with the xPOD. The mingling index, spatial Simpson index and the ISAR are all sensitive to community size or diversity, which are not expected to vary consistently with process in real communities, and this may limit their ability to discriminate between the processes. Nevertheless, they allow for the comparison of signals in spatial patterns with those in diversity or abundances, which other measures do not.
### Table 4.1: Separate pieces of information used in spatial measures

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of aggregation</td>
<td>$D_{agr}(r) = \sum_{j=1}^{N_j} \frac{N_{jj}(r)}{\pi r^2 n_j (1 - n_j)}$</td>
<td>Coomes et al. (1999)</td>
</tr>
<tr>
<td>Measure of interspecific segregation</td>
<td>$S_j = \log \frac{n_c/n_h}{n_j - 1/n_j}$</td>
<td>Dixon (1994)</td>
</tr>
<tr>
<td>Proportion of conspecific neighbours</td>
<td>$P_c = \frac{N_{jj}(r)}{N(r)}$</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4.2a: Definitions of measures of scattering
<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISAR</td>
<td>ISAR(<em>j)(r) = \frac{N</em>{ij}(r)}{N(r)}</td>
<td>Wiegand et al. (2007a)</td>
</tr>
<tr>
<td>Mingling index</td>
<td>(M_j(r) = \frac{N_{ij}(r)}{N(r)})</td>
<td>Lewandowski &amp; Pommerening (1996); Graz (2004); Rajala &amp; Illian (in press)</td>
</tr>
<tr>
<td>Spatial Simpson index</td>
<td>(S(r) = 1 - \sum_{j=1}^{N_s} P_j \frac{\bar{N}_{ij}(r)}{N(r)})</td>
<td>Shimatani (2001); Rajala &amp; Illian (in press)</td>
</tr>
<tr>
<td>Degree of association</td>
<td>(D_{as}(r) = \sum_{1}^{N_j} N_{jk}(r) \frac{1}{n_j n_k A_c})</td>
<td>Coomes et al. (1999)</td>
</tr>
<tr>
<td>Cross-pair overlap</td>
<td>(xP[OD(R) = \left{ A_{jk} = \int_{0}^{R} \log \left( \frac{N_{jk}(r)}{n_j n_k A_c} \right) dr ; ; j, k \right} )</td>
<td>Brown et al. (2011)</td>
</tr>
</tbody>
</table>

**Table 4.2b: Definitions of measures of exposure**
Figure 4.1: The class and level to which each piece of information used in the construction of the measures considered here belongs.

Terms are divided into those that contain conspecific and heterospecific, scale-dependent and scale-independent information. They are further divided into the levels – individual, species, or community, at which they operate. The first (lowest-level) appearance of a term is denoted by bold type, dashed cones show where they may be averaged up to higher levels. Complete measures, which are made up of combinations of the information in the diagram, are shown at their appropriate level. They are in bold at the level at which they were originally defined, and also included at other levels where it is possible to construct them.
4.2.2 ECOLOGICAL SIMULATIONS

In order to test the sensitivity of the different spatial measures to particular ecological processes, we use data from the simulations presented in Chapter 2. These are stochastic individual-based models of a plant community in continuous space, in which multispecies spatial patterns are allowed to develop through several million birth and death events. These events occur according to neutral, niche, lottery, Janzen-Connell or heteromyopia assumptions; chosen as the principal theorised mechanisms of species coexistence in diverse plant communities. This approach allows any spatial signals of these processes to be isolated and assessed.

Under neutrality, the only processes occurring are density-dependent mortality, dispersal limitation (both of which occur identically in all species) and the immigration of new species. These processes also occur in the other models but are adjusted as follows: in the niche model, individuals have lower death rates when they are within their species’ preferred (geographically defined) environment; in the lottery model, species’ death rates vary at random time intervals; in the Janzen-Connell model local density-dependent mortality is stronger within species; and in the heteromyopia model density-dependent mortality occurs over larger areas within species than between them. Ten realisations of each model are generated so that the variability in resulting spatial pattern can be assessed, and the spatial measures calculated for each.

4.2.3 COMPARING SPATIAL STATISTICS

Statistics describing scattering and exposure are compared on the basis of their ability to distinguish the above models from their resulting spatial patterns. This ability is assessed in several ways. At each level, all appropriate measures (those that may be constructed at the particular level) are
calculated and plotted. Scale-independent measures give one value per individual, species or community; scale-dependent measures give several values when calculated at different scales (at or to different distances; see Table 4.2). In all cases, scale-dependent measures are calculated at or to a maximum radius of one quarter of minimum plot dimension, following the recommendation of Baddeley & Turner (2005). Depending on observed sensitivity to spatial scale, the number of increments of radius within this distance varies between 30 and 100 (increments of 0.0083 and 0.0025 respectively).

Once calculated, measures are compared separately at each level. Only four of the pieces of information used to construct the statistics are available at the individual level (i.e. describe some property of individual plants) (Fig. 4.1). When included in statistics that describe structure at species or community level, these are averaged as appropriate. None of the measures themselves are intended to operate at the individual level although three – the proportion of conspecific neighbours, the ISAR, and the Mingling index – are able to. However, all data at the individual level suffer from being large in number and variation; a trait that renders measures at this level difficult to calculate and interpret. These attributes make such data ideal for the study of differences between small numbers of individuals, but unwieldy and excessively noisy for studies at larger scales, especially with the additional dimension of scale. As a result, results from this level are not presented here.

At species level, the results of each measure are plotted in their entirety, with separate results for each species plotted together (at the community-level, in contrast, results are average across species rather than being retained at this level). The success of each measure in discriminating between the models is then visually assessed. For scale-dependent measures, information from the radius or radii at which differences between the models appear greatest is used to construct boxplots of results; scale-independent measures are summarised immediately as boxplots. While differences between these boxplots are not formally measured, their degree of overlap is again assessed visually, and distributions of values plotted for the radius at which this is minimised. These distributions are finally characterised by their first 3 moments – mean, standard deviation and skewness.

At the community level, the range of results observed across 10 realisations of each model is used to construct 95% prediction intervals for each measure. As before, the radius at which these intervals are most distinct from one another is chosen for boxplots of values taken by the measure. Once again, the boxplots produced by each measure (within each group) are visually compared, and those which detect the greatest differences between models identified. Formal significance tests of the differences between model results are not carried out (but see Chapter 2) because we aim to identify the statistics that are most sensitive to modelled processes, regardless of the significance level of this sensitivity. However, the 95% prediction intervals generated at community level provide a clear indication of the significance of any differences between distributions.

4.3 Results

4.3.1 SPECIES LEVEL

The majority of spatial statistics operate on the species level. The pieces of information here are more numerous and, where they have been averaged up from lower organisational levels, subject to far less variability – or random noise – than their lower-level counterparts. Nevertheless, they remain sensitive to informative low-level (interspecific) variation. The distinction with the individual-level is blurred, however, as many species are represented by very small numbers of individuals. In these cases, the information effectively remains on the individual level, retaining the
random noise and spurious spatial signals associated with this. Many measures of spatial structure therefore ignore information from rare species entirely, as do measures at community level, below.

4.3.1.1 Scattering
It is possible to define all three measures of scattering which we consider at the species level (Fig. 4.1), and all make use of scale-independent information.

4.3.1.1.1 Measure of interspecific segregation
The measure of interspecific segregation, a misleadingly named description of the deviation in the number of conspecific nearest neighbours from that expected under random distributions of individuals, is entirely scale-independent (Table 4.2, Fig. 4.1), and its shape varies considerably within models (Fig. 4.3). As above, variation is minimised by the imposition of an abundance threshold. We use thresholds of 1, 100 and 500 individuals per species here, and find that a threshold of 500 minimises uninformative variation while still including sufficient numbers of species to produce a clear signal. Nevertheless, this measure remains unsuccessful in distinguishing different models, with the signal for each having similar means and being dominated by variations between species.

![Boxplot of interspecific segregation results for each model](image)

**Figure 4.3: Measure of interspecific segregation results for each model.**
Three abundance thresholds are used: 1 (left), 100 (middle), and 500 (right). Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.

4.3.1.1.2 Degree of aggregation
The degree of aggregation is intended to operate at the community level but can also be expressed at the species level. In addition, it makes use of scale-dependent information and so can be calculated both at different abundance and radius thresholds (Figs. 4.4a & 4.4b). At a threshold of 500 individuals, multi-species curves appear coherent and clearly differ between models. Both aggregation at small radii and variation between species results at all radii are greatest in the lottery model, and lowest in the Janzen-Connell model. This suggests that the measure may distinguish the models most reliably at small radii, and boxplots of results at each of the first four radii for which the measure is calculated confirm this (Fig. 4.4b).

While results from each model partially overlap in each of these cases, they do differ substantially from one another both in their mean values and their variance, which remains consistently largest in
the lottery model and smallest in the Janzen-Connell model. These differences are clearest at the smallest radius considered, and these results are therefore chosen for expression as a histogram (Fig. 4.4c). Differences in variance, mean and skewness are clear (Table 4.3), and suggest that the first three moments of the distributions may provide bases for distinguishing the models, with the niche and lottery models having higher and more variable levels of aggregation.

Figure 4.4a: The degree of aggregation for species in each model.

Abundance thresholds of 1 (top left panel), 100 (top right panel) and 500 (bottom left panel). Radius increments in units of 0.0083 of minimum plot dimension.
Figure 4.4b: The degree of aggregation for species in each model.

Radii of 1 (top-left), 2 (top-right), 3 (bottom-left) and 4 (bottom-right) times 0.0083 of minimum plot dimension. A minimum abundance threshold of 500 individuals per species is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
Figure 4.4c: Histograms of values of the degree of aggregation for species in each model at radii of 4 times 0.0083 of minimum plot dimension, with a minimum abundance threshold of 500.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janzen-Connell</td>
<td>1.933</td>
<td>0.323</td>
<td>0.114</td>
</tr>
<tr>
<td>Neutral</td>
<td>2.355</td>
<td>0.700</td>
<td>1.110</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td>2.613</td>
<td>0.795</td>
<td>0.881</td>
</tr>
<tr>
<td>Niche</td>
<td>6.076</td>
<td>2.092</td>
<td>0.842</td>
</tr>
<tr>
<td>Lottery</td>
<td>5.253</td>
<td>4.270</td>
<td>1.574</td>
</tr>
</tbody>
</table>

Table 4.3: Means, standard deviations and skews of the degree of aggregation for species in each model at radii of 4 times 0.0083 of minimum plot dimension, with an abundance threshold of 500.
4.3.1.1.3 Proportion of conspecific neighbours

The proportion of conspecific neighbours, when expressed at the species level (Fig. 4.5a), shows very substantial differences between models. In the niche model, for example, almost all species show a higher proportion of conspecific neighbours at small radii than in the neutral, heteromyopia or Janzen-Connell models. In addition, variation in behaviour between species clearly conceals large differences in the mean of the index between models, suggesting that the measure may discriminate better between models when expressed at the community level. Nevertheless, the proportion may usefully be summarised as boxplots or histograms of values taken at a particular radius in each model. Because the greatest differences between the models are found at very small radii, the proportion is calculated and compared at finer increments of radius (of 0.0025) than above.

The resulting boxplots (Fig. 4.5b) show that the neutral, Janzen-Connell and heteromyopia model results are reliably distinct from those of the niche model, but that the lottery model is not reliably distinct from any other. However, all model results differ from one another to some extent in their means and variances. A radius of 0.00625 of minimum plot dimension is chosen for further analysis of these differences, and histograms of the values produced at this radius are shown in Figure 4.5c. When summarised by their first three moments (Table 4.4), these distributions are found to differ substantially between all of the models.

Figure 4.5a: The proportion of conspecific neighbours for species with abundances of 500 or more in each model.
Figure 4.5b: The proportion of conspecific neighbours for species in each model. Radii of 1 (top-left), 2 (top-right), 3 (bottom-left) and 4 (bottom-right) times 0.00125 of minimum plot dimension. A minimum abundance threshold of 500 individuals per species is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
Figure 4.5c: Histograms of values of the proportion of conspecific neighbours for species in each model at a radius of 0.00625 of minimum plot dimension.

A minimum abundance threshold of 500 individuals per species is used.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janzen-Connell</td>
<td>0.101</td>
<td>0.010</td>
<td>0.013</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.126</td>
<td>0.018</td>
<td>0.191</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td>0.141</td>
<td>0.023</td>
<td>0.460</td>
</tr>
<tr>
<td>Niche</td>
<td>0.366</td>
<td>0.087</td>
<td>1.171</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.291</td>
<td>0.160</td>
<td>0.982</td>
</tr>
</tbody>
</table>

Table 4.4: Means, standard deviations and skews of the proportion of conspecific neighbours for species in each model at radii of 0.00625 of minimum plot dimension, with an abundance threshold of 500 individuals per species.
4.3.1.2 Exposure
Most of the measures of exposure considered here operate at the species level: the ISAR, mingling index, degree of association and xPOD. Because the ability of the xPOD to distinguish these models has previously been assessed (Chapter 2), it is our starting point in the comparison of measures of exposure.

4.3.1.2.1 Cross-pair overlap distribution (xPOD)
In contrast to the other measures of exposure we use, the xPOD is only calculated at one radius and is cumulative to this point (set here as 0.25 of minimum plot dimension). When plotted, xPODs for the neutral and niche models show the greatest differences (Fig. 4.6), but almost all show some clear divergence, with the exception of the almost indistinguishable neutral and heteromyopia results. The first three moments of the distributions (Table 4.5) differ to varying extents between all models, however, including these.

![Figure 4.6: Cross-pair overlap distributions for species in each model.](image)

A minimum abundance threshold of 500 individuals per species is used.
### Table 4.5: Means, standard deviations and skews of xPODs for species in each model, with an abundance threshold of 500 individuals per species.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janzen-Connell</td>
<td>-0.050</td>
<td>0.026</td>
<td>-0.262</td>
</tr>
<tr>
<td>Neutral</td>
<td>-0.053</td>
<td>0.040</td>
<td>-0.330</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td>-0.066</td>
<td>0.066</td>
<td>-0.688</td>
</tr>
<tr>
<td>Niche</td>
<td>-0.267</td>
<td>0.319</td>
<td>-0.556</td>
</tr>
<tr>
<td>Lottery</td>
<td>-0.107</td>
<td>0.132</td>
<td>-0.968</td>
</tr>
</tbody>
</table>

#### 4.3.1.2.2 Individual species-area relationship (ISAR)

Of the many measures which use species-level scale-dependent information, the ISAR is the simplest, using only the average count of species found in circular neighbourhoods around members of each species. As initially defined it is calculated only for species with 20 or more individuals, but we restrict it to species with 500 or more, as above. It is also possible to express ISAR curves as proportions of the total number of species in a community, to remove differences due to modelled fecundities of species. In our results, the neutral, niche, and Janzen-Connell models are readily distinguishable in the un-normalised ISAR due to the differences in numbers of species in the communities (Fig. 4.7a). When this is corrected for, the niche and lottery models are distinguishable from the others only by their variability, while the Janzen-Connell model retains a slightly higher level of species mixing.

When these differences are examined in boxplots of ISAR values at different radii, a number of additional features become apparent (Fig. 4.7b). At the smallest radii, the niche model produces the highest ISAR values for individual species, while the neutral model produces the highest mean value. Contrary to expectations, the Janzen-Connell and heteromyopia models, when normalised by overall diversity, produce lower values than the neutral model. As radius increases, however, values produced by the neutral and heteromyopia models increase relatively slowly compared to those of the niche, lottery and Janzen-Connell models. As a result, the niche and Janzen-Connell models give the highest ISAR values after only a small increase in radius. The lottery model produces the lowest values at all radii.

From these, the results at the second radius ($r = 2 \times 0.0083$) are chosen as a basis for comparison of histograms of ISAR values (Fig. 4.7c). These show limited systematic variation among models, particularly the neutral, Janzen-Connell and heteromyopia models. The first three moments of the distributions (Table 4.6) confirm this impression, with differences in skewness being the most dramatic. The niche and lottery models also produce results that are chiefly distinguishable by their skewness.
Figure 4.7a: ISAR curves for species with abundances of 500 or more in each model expressed in terms of species numbers (top) and proportion of total community size (bottom).
Figure 4.7b: ISAR result boxplots for species in each model at radii of 1 to 9 times 0.0083 of minimum plot dimension.

A minimum abundance threshold of 500 individuals per species is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
Figure 4.7c: Histograms of ISAR values for species in each model at a radius of 0.0167 of minimum plot dimension.
A minimum abundance threshold of 500 individuals per species is used.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janzen-Connell</td>
<td>0.033</td>
<td>0.001</td>
<td>-0.375</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.040</td>
<td>0.002</td>
<td>0.093</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td>0.038</td>
<td>0.001</td>
<td>-0.264</td>
</tr>
<tr>
<td>Niche</td>
<td>0.035</td>
<td>0.006</td>
<td>-1.188</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.022</td>
<td>0.005</td>
<td>-0.977</td>
</tr>
</tbody>
</table>

Table 4.6: Means, standard deviations and skews of ISAR values for species in each model at radii of 0.0167 of minimum plot dimension, with an abundance threshold of 500 individuals per species.
4.3.1.2.3 Mingling index

The mingling index is also intended to operate at the species level, and, already being expressed as a proportion, is less sensitive to overall community size and diversity. Unlike the ISAR, it does not take a value of zero at small radii, and so is also sensitive to differences between models at small spatial scales. These are substantial, with the niche and lottery models producing lower values at small radii than the neutral model (Fig. 4.8a). Their results are also far more variable, with many species showing less mingling at all radii for which the index is calculated than in the neutral model. Once again the Janzen Connell model shows a more restricted range of high values. These differences are clear in boxplots of the index’s value at small radii (Fig. 4.8b), which nonetheless show limited differences between the means of the neutral, Janzen-Connell and heteromyopia models. A radius of 0.0167 times minimum plot dimension is chosen for further analysis as differences between models – the neutral and Janzen-Connell models in particular – are eroded on either side of this point. Once again the distributions at this radius are chiefly distinguishable by their skewness, which is minimised, along with the standard deviation, in the Janzen-Connell model (Fig. 4.8c, Table 4.7).
Figure 4.8a: Mingling index for species with abundances of 500 or more in each model, for radii of 1-30 (top) and 1-10 (bottom) times 0.0083 of minimum plot dimension.
Figure 4.8b: Mingling index boxplots for species in each model at radii of 1 to 4 times 0.0083 of minimum plot dimension.

A minimum abundance threshold of 500 individuals per species is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
Figure 4.8c: Histograms of mingling index values for species in each model at a radius of 0.0167 of minimum plot dimension.

A minimum abundance threshold of 500 individuals per species is used.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janzen-Connell</td>
<td>0.909</td>
<td>0.009</td>
<td>-0.012</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.886</td>
<td>0.017</td>
<td>-0.258</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td>0.873</td>
<td>0.021</td>
<td>-0.526</td>
</tr>
<tr>
<td>Niche</td>
<td>0.652</td>
<td>0.088</td>
<td>-1.282</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.718</td>
<td>0.161</td>
<td>-1.006</td>
</tr>
</tbody>
</table>

Table 4.7: Means, standard deviations and skews of mingling index results for species in each model, with an abundance threshold of 500 individuals per species.
4.3.1.2.4 Degree of association

The degree of association is intended to be used at the species level, but is first calculated at the level of species pairs. In this form, the obvious differences between models are again in terms of variability of the measure, with the niche and lottery models being the most variable and the Janzen-Connell the least, with the highest consistent values at intermediate radii (Fig. 4.9a). These differences do not appear to be dramatic when the measure is plotted in boxplots at different radii (Fig. 4.9b), but are clearer in histograms of values at a radius of 0.0167 of minimum plot dimension (Fig. 4.9c). When the first 3 moments of each distribution are compared (Table 4.8), only skewness appears to differ substantially between all models, while standard deviation differs between the niche, lottery and other models.

When the measure is averaged to the species level (by averaging across values for all species pairs in which a particular species occurs), differences in variance persist (Fig. 4.10a). Boxplots additionally reveal small apparent differences in mean values between models, although these are eroded by increasing radius (Fig. 4.10b). Once again, however, only the skewness of distributions of values at a fixed radius differ substantially between models (Fig. 4.10c, Table 4.9).

Figure 4.9a: The degree of association for species with abundances of 500 or more in each model, at the level of species pairs.
Figure 4.9b: Degree of association boxplots for species pairs in each model at radii of 1 to 4 times 0.0083 of minimum plot dimension.

A minimum abundance threshold of 500 individuals per species is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
Figure 4.9c: Histograms of degree of association values for species pairs in each model at a radius of 0.0167 of minimum plot dimension.

A minimum abundance threshold of 500 individuals per species is used.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janzen-Connell</td>
<td>1.247</td>
<td>0.346</td>
<td>0.494</td>
</tr>
<tr>
<td>Neutral</td>
<td>1.217</td>
<td>0.412</td>
<td>0.902</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td>1.198</td>
<td>0.450</td>
<td>1.081</td>
</tr>
<tr>
<td>Niche</td>
<td>0.913</td>
<td>1.561</td>
<td>2.827</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.992</td>
<td>0.795</td>
<td>4.004</td>
</tr>
</tbody>
</table>

Table 4.8: Means, standard deviations and skews of degree of association results for species pairs in each model, with an abundance threshold of 500 individuals per species.
Figure 4.10a: The degree of association for species with abundances of 500 or more in each model, at species level.
Figure 4.10b: Degree of association boxplots for species in each model at radii of 1 to 4 times 0.0083 of minimum plot dimension.

A minimum abundance threshold of 500 individuals per species is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
Figure 4.10c: Histograms of degree of association values for species in each model at a radius of 0.0167 of minimum plot dimension.

A minimum abundance threshold of 500 individuals per species is used.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janzen-Connell</td>
<td>1.246</td>
<td>0.124</td>
<td>-0.321</td>
</tr>
<tr>
<td>Neutral</td>
<td>1.224</td>
<td>0.173</td>
<td>0.920</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td>1.203</td>
<td>0.144</td>
<td>0.759</td>
</tr>
<tr>
<td>Niche</td>
<td>0.894</td>
<td>0.534</td>
<td>1.953</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.918</td>
<td>0.395</td>
<td>0.619</td>
</tr>
</tbody>
</table>

Table 4.9: Means, standard deviations and skews of degree of association results for species in each model, with an abundance threshold of 500 individuals per species.
4.3.2 COMMUNITY LEVEL

Many community-level measures are comprised of results averaged from the species level, and few operate exclusively at this level. Nevertheless, expression of results at the community level is expected to have advantages over species level summaries because between-species variation is averaged out. This may produce a clearer signal of the processes modelled here, which are not species-specific. We have ten community-level results for each model; one for each realisation of the simulation. Together, these allow us to generate a range of results for each model rather than the single result produced by any one community.

4.3.2.1 Scattering

4.3.2.1.1 Degree of aggregation

The only measure of scattering intended to operate at the community level is the degree of aggregation, which is summed over all species in a community. In this case, community-level results from each model are highly dissimilar but not distinct (Figs. 4.11a & 4.11b) with only the lottery and Janzen-Connell models producing ranges of values almost wholly different from the neutral, niche and heteromyopia models. In the summation of values, however, information on overall community size (in terms of species numbers) is included, and this is known to differ because average fecundities vary between models. Being at least partially an effect of model design rather than modelled process, this information is excluded through the calculation of mean values for the index (Figs. 4.12a & 4.12b). These show greater differences between models, with the niche model producing a distinct range of values, and others differing in terms of means and variances — especially in the case of the neutral model which is distinct from all but the lottery model.
Figure 4.11a: Prediction intervals (95%) for the degree of aggregation (summed) at community level, based on 10 realisations of each model.
Information from species with an abundance of at least 500 is used.

Figure 4.11b: Boxplots of degree of aggregation (summed) values generated by 10 realisations of each model.
Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
Figure 4.12a: Prediction intervals (95%) for the mean degree of aggregation at community level, based on 10 realisations of each model. Information from species with an abundance of at least 500 is used.

Figure 4.12b: Boxplots of degree of aggregation mean values generated by 10 realisations of each model. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
4.3.2.1.2 Measure of interspecific segregation

The measure of interspecific segregation, when averaged to the community level, shows substantially greater variation between models (Fig. 4.13). Not only are differences in variance observed at the species level still apparent, but the range of values produced are entirely distinct for all models, with the only exception being a slight overlap between the lottery and heteromyopia models. The extent of segregation as measured by this index is minimised in the Janzen-Connell model and increases through the neutral, heteromyopia, lottery and finally niche model.

![Boxplot of mean values of the measure of interspecific segregation for each model, based on 10 realisations of each (500 abundance threshold).](image)

**Figure 4.13**: Boxplots of mean values of the measure of interspecific segregation for each model, based on 10 realisations of each (500 abundance threshold).

Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.

4.3.2.1.3 Proportion of conspecific neighbours

Similarly, the proportion of conspecific neighbours gives clearer results at the community level (Figs. 4.14a & 4.14b). Once again the niche and Janzen-Connell models are the most different, with the Janzen-Connell model giving the lowest proportion of conspecific neighbours and the niche model the highest. More significantly, the 95% prediction limits for every model are distinct, with the exception of some overlap between those of the niche and lottery models. Heteromyopia produces a higher proportion of conspecific neighbours than the neutral model at small spatial scales.
Figure 4.14a: Prediction intervals (95%) for the mean proportion of conspecific neighbours at community level, based on 10 realisations of each model. Information from species with an abundance of at least 500 is used.

Figure 4.14b: Boxplots of mean values of the proportion of conspecific neighbours for 10 realisations of each model. Information from species with an abundance of at least 500 is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
4.3.2.2 Exposure

4.3.2.2.1 Spatial Simpson index

One measure of exposure that operates only at the community level is the spatial Simpson index. As a measure of spatial diversity, it is intended to include all species, not just those above a certain abundance. In this form, the measure clearly detects lower means and greater variances of the niche and lottery model results relative to those of the other models, but far smaller differences between the remaining three models (Fig. 4.15a). The greatest differences are found at the smallest radius for which the measure is calculated, where the niche/lottery, neutral, heteromyopia and Janzen-Connell model results do not overlap (Fig. 4.15b). When the measure is calculated with an abundance threshold of 500, the variance of each of the models’ results increases, but differences between them remain similar (Figs. 4.16a & 4.16b). The niche and lottery models are still distinguishable only by their variances, while the results of the Janzen-Connell model become slightly more dissimilar to those of the neutral and heteromyopia models.

Figure 4.15a: Prediction intervals (95%) for the spatial Simpson index, based on 10 realisations of each model and including all species. Results are shown for radii of 1-10 times 0.0083 of minimum plot dimension.
Figure 4.15b: Boxplots of spatial Simpson indices for 10 realisations of each model, at a radius of 0.0083 minimum plot dimension, and including all species.

Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia. The plot on the right shows the neutral, Janzen-Connell and heteromyopia results on a smaller y range.

Figure 4.16a: Prediction intervals (95%) for the spatial Simpson index, based on 10 realisations of each model.

Results are shown for radii of 1-10 times 0.0083 of minimum plot dimension and an abundance threshold of 500 is used.
Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia. The plot on the right shows the neutral, Janzen-Connell and heteromyopia results on a smaller y range. An abundance threshold of 500 is used.

4.3.2.2 xPOD
Other measures of exposure are averaged from the species level, with the exception of the xPOD which, despite being a community-level measure, retains its constituent information at the level of species pairs. As a result, community-level distributions are the same as those at the species level, presented with variation between realisations captured in the form of prediction intervals, for example (Fig. 4.17a). For comparison with the other measures, however, the mean and standard deviation of the xPOD for each realisation is calculated and compared between models (Fig. 4.17b). Some of these results have previously been presented (Chapter 2) but are included below for convenience. These show substantial and robust differences between models, with each being wholly distinct apart from some overlap between lottery and heteromyopia results, in both summaries of the measure.
Figure 4.17a: Prediction intervals (95%) for the cross-pair overlap distribution, based on 10 realisations of each model.
4.3.2.2.3 ISAR

Figure 4.17b: Boxplots of means (left) and standard deviations (right) of the cross-pair overlap distribution for 10 realisations of each model.

Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.

The ISAR, in being averaged across species, becomes only slightly more informative about modelled process. When expressed as a mean proportion of the total number of species, it is able to distinguish only the lottery, Janzen-Connell and neutral/niche/heteromyopia model groups, with only limited differences in mean and variance occurring within these (Figs. 4.18a & 4.18b). These nevertheless indicate that a decreasing proportion of community diversity is found within local neighbourhoods of the neutral, heteromyopia, niche, Janzen-Connell and lottery models.
Figure 4.18a: Prediction intervals (95%) for the mean ISAR at community level, as a proportion of total community size, based on 10 realisations of each model.

Information from species with an abundance of at least 500 is used.
Figure 4.18b: Boxplots of mean values of the ISAR, as a proportion of community size, for 10 realisations of each model at radii of 1-4 times 0.0083 of minimum plot dimension.

Information from species with an abundance of at least 500 is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
4.3.2.2.4 Mingling index
The mingling index, in contrast, detects substantial differences between the models, especially at small radii, as before (Figs. 4.19a & 4.19b). In fact, only slight overlap occurs between the results of different models here. The niche model has the lowest levels of mingling as measured by this index, which increases through the lottery, heteromyopia, neutral and Janzen-Connell model results.

Figure 4.19a: Prediction intervals (95%) for the mean mingling index at community level based on 10 realisations of each model.
Information from species with an abundance of at least 500 is used.
4.3.2.2.5 Degree of association
The degree of association also detects the largest differences between the models at the smallest radii (Figs. 4.20a & 4.20b). In this case, however, only the neutral/Janzen-Connell/heteromyopia and niche/lottery groups are distinct; an improvement over the species-level measure but one which does not provide the discriminatory power of other measures. Despite this, mean values once again indicate that, in terms of species associations, the models are ordered Janzen-Connell > neutral > heteromyopia > niche > lottery.
Figure 4.20a: Prediction intervals (95%) for the mean degree of association at community level based on 10 realisations of each model.
Information from species with an abundance of at least 500 is used.

Figure 4.20b: Boxplots of mean values of the degree of association for 10 realisations of each model.
Information from species with an abundance of at least 500 is used. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia.
4.4 Discussion

We divide the statistics and information used here by the aspects of spatial behaviour that they describe and the levels at which they operate. The primary objective of this division is to identify which descriptions of intra-specific scattering and inter-specific exposure are the most sensitive to the processes we model, and at which level their sensitivity to these processes is maximised. The secondary objective is to fully characterise the spatial signals of the modelled processes, in terms of both scattering and exposure. It has previously been noted that the differing spatial implications of coexistence mechanisms are unlikely to be captured by any one spatial statistic, and that a comparison of several statistics is therefore necessary to describe them (e.g. Amarasekare 2003). However, a systematic comparison of this kind has not previously been carried out.

4.4.1 SCATTERING

The use of measures of scattering to quantify intraspecific aggregation is well established, although a substantial proportion of the measures used operate in discrete space, considering densities of individuals within set areas, for example (e.g. Lang et al. 1971; Sevenster 1996; Souza & Martins 2003; Veech 2005). Dixon (1994) proposed the measure of interspecific segregation simply as a robust test of spatial structure, and Coomes et al. (1999) used the degree of aggregation as part of a test for associations between species. Many other measures of scattering are similarly used to quantify aggregation or associations without suggesting further links to specific processes (e.g. Condit et al. 2000; Hao et al. 2007).

However, the measures of scattering considered here give a clear and consistent description of intra-specific spatial patterns under the different models. At species level, all three measures are dominated by inter-specific variation, which is greatest in the lottery model and decreases through the niche, heteromyopia, neutral and Janzen-Connell models. Nevertheless, some of the models are distinct in spite of this variation, particularly at small radii.

The measure of interspecific variation and degree of aggregation are unable to reliably distinguish any of the models at the species level, but do indicate that intra-specific clumping is greatest in the lottery and niche models, and smallest and more consistent across species in the Janzen-Connell model (although variation in the lottery model means that any one result may show clumping that resembles the results of other models). This is confirmed by the proportion of conspecific neighbours, the most successful of the measures of scattering at species level, which can reliably distinguish the niche model from the neutral, Janzen-Connell, or heteromyopia models. It also shows that the heteromyopia model produces slightly higher conspecific densities than the neutral model, especially at small radii.

The potential for linking species-level patterns of scattering to underlying processes is often thought to be limited by the strong effects of dispersal limitation (Pacala & Levin 1997; Plotkin et al. 2000), and it is important to note that we hold dispersal constant across species here. Nevertheless, scattering has previously been found to be informative about processes such as intraspecific and interspecific competition (Stoll & Prati 2001; Luo et al. 2012). In particular, high levels of intraspecific aggregation may be able to delay or prevent competitive exclusion in a wide range of ecological communities (Ives 1988; Tilman 1994; Shorrocks & Sevenster 1995; Monzeglio & Stoll 2005), and the degree of aggregation is among the measures suggested for use in studies of this (Rácz & Karsai 2006). Heteromyopia is able to support the coexistence of species for this reason (Vogt et al. 2010), and measures of aggregation have been used to detect Janzen-Connell or other mortality effects (e.g. Vacek & Lepš 2009; Bagchi et al. 2011). Less specifically, changes in patterns of scattering with time have been linked to changes in underlying ecological processes (e.g. He et al.
These observations apply at the species level and several of them imply the spatial signals that we identify above.

When measures are expressed at the community level, the differences between models become far clearer and provide a fuller picture of the spatial effects of the models. All three measures distinguish some of the models from one another, and the measure of interspecific segregation and proportion of conspecific neighbours can each distinguish all but one model pair (niche/heteromyopia and niche/lottery respectively), and even these overlap only slightly. This difference between the two measures is presumably due to the information which they make use of: the measure of interspecific segregation uses data about the first nearest neighbour and global proportions; the proportion of conspecific neighbours only data from a local neighbourhood.

Our findings therefore imply that the niche and heteromyopia models produce similar patterns at the immediate (first neighbour) spatial scale, but become increasingly dissimilar as spatial scale increases – an inference that is supported by boxplots of the proportion of conspecific neighbours at the species level (Fig. 4.5b), in which the niche and heteromyopia results overlap at the smallest scale but become discrete with increasing radius. It is possible to conclude that heteromyopia manifests itself, in terms of scattering, through elevated intra-specific clumping at the smallest scales, and compensatory dispersion at intermediate scale, producing a result at larger scales similar to that of the neutral model (which also agrees with the similarities in exposure found between the two models by the xPOD). This is consistent with the recognised decrease in intensity of intraspecific interactions at small radii and increase at large radii relative to the neutral case that results from heteromyopia (Murrell & Law 2003; Vogt et al. 2010).

More broadly, these results demonstrate that each of the models has a specific implication for patterns of scattering (Table 4.10). The niche model produces the least scattered species, because of their association with particular geographically-defined habitats; the lottery model the next-least, but more variably, scattered species, because of the domination of a few particularly widespread species at any given time; the heteromyopia model produces low levels of scattering at small scales and increased scattering at intermediate scale, because of species’ tolerance of high immediate densities of conspecifics but not of high densities over larger distances; the neutral model more scattered species which are separated only through limited dispersal and density-dependent mortality; and the Janzen-Connell model the most, and least variable, scattering, because of species’ uniformly lower tolerances of high conspecific density.

Detailed community-level findings of this kind have not previously been made with measures of scattering. It has been recognised that scattering has consequences for both α- and β-diversity that make it a suitable focus for studies of underlying process (Veech 2005), especially when measured at the community level (Luo et al. 2012). However, appropriate measures have not been identified for this purpose and, where used, measures of scattering have generally been applied at the species level to investigate the role of aggregation in maintaining coexistence, as above (e.g. Damgaard 2010).

4.4.2 EXPOSURE

We find that patterns of exposure are as consistent and informative as patterns of scattering. At the species level, all measures can distinguish the niche and lottery models from the others because of the higher variance in the results they produce. Despite being of similar magnitude, these variances have different sources. In the lottery model it is due to the previously identified variability in behaviour between species, with some being particularly dominant and widespread and others being marginalised as a result. In the niche model, however, behaviour is more consistent across species, and differences in exposure are attributable instead to the extent of similarities in species’
environmental preferences: species with similar preferences will show higher levels of exposure than those with dissimilar preferences (as captured most clearly by the degree of association, Fig. 4.9b).

Species-level differences are not limited to this, however. The ISAR can distinguish the lottery model from all others except the niche model, due to its substantially lower levels of local species diversity. The mingling index, which does not take account of heterospecific identity, can distinguish the niche model from all others except the lottery model due to its lower levels of mingling. Although difficult to interpret, this difference suggests that while greater proportions of total species numbers occur in local neighbourhoods of the niche model, the proportion of heterospecifics of any type (whether from a single species or several) is greater in the lottery model. High levels of exposure in the niche model therefore occur between a greater number of species than in the lottery model, and this is related to the lower levels of scattering under niche differentiation.

In addition, the ISAR suggests that the Janzen-Connell model has low proportional diversity at the smallest scales but that the rate of its increase with scale is greatest in this and the niche models, and that proportional diversity in the heteromyopia model increases slightly faster with scale than in the neutral model. The low proportional diversity of immediate neighbourhoods in the Janzen-Connell model is due to high overall diversity, while a rapid rate of increase with radius is related to high levels of scattering which ensure that many species are encountered at each radius increment. In the niche model, similar increases are due to species turnover with environmental variation, as implied by the xPOD. Under heteromyopia, higher rates of increase are found than under neutrality because of the low levels of scattering at very small scales, which produce immediate neighbourhoods dominated by conspecifics and intermediate neighbourhoods dominated by heterospecifics.

As with measures of scattering, measures of exposure have primarily been applied at the species level, and in order to describe patterns or to identify associations between species without suggesting an underlying cause. Numerous measures have been used in this way, including the mingling index (Pommerening 2002; Aguirre et al. 2003) and ISAR, which was developed as a description of spatial diversity (Wiegand et al. 2007a). Early measures were defined in discrete space (e.g. McIntosh 1957; Hurlbert 1969), but it was also recognised that distances between neighbours of different species could be used to distinguish associations between species (Clark & Evans 1954; Goodall 1965). This is reflected in the popularity of measures derived from spatial point process methodology (e.g. Szwagrzyk 1992; Manabe et al. 2000; Li et al. 2008). Nevertheless, the explanation of observed associations has long been regarded as beyond the scope of these measures (e.g. Cole 1957). Recently, patterns of exposure have been used to distinguish intraspecific and interspecific associations (Roxburgh & Chesson 1998), interspecific associations and environmental signals (Wiegand et al. 2007b) and to investigate the role of spatial structure in maintaining species coexistence (Luo et al. 2012).

At the community level, all measures of exposure clearly distinguish most or all models from one another. This includes the spatial Simpson index, which operates only at this level. Most are able to distinguish all but the niche and lottery models (the lottery model is distinguishable by the variability of its resulting spatial patterns only when several are available), while the xPOD and ISAR can distinguish these but not some of the others. The mingling index appears able to distinguish all of the models from one another.

From this, the following picture of neighbourhood diversity, or exposure, emerges. The niche and lottery models have the lowest levels of exposure, which are linked to their low levels of scattering and high densities of conspecifics. Exposure increases more rapidly in the niche model because of geographical environmental variation, allowing new species to dominate at intermediate scales. The heteromyopia model has the next lowest exposure levels, but these also increase relatively quickly
with scale as high intermediate-scale sensitivity to conspecifics leads to greater mixing. The neutral model has relatively high levels of exposure driven by dispersal limitation and produces results which are similar to those of the heteromyopia model, particularly at large scales. Finally the Janzen-Connell model shows high exposure levels which increase steadily with scale, because high levels of scattering ensure that species are uniformly mixed.

Previously, measures of exposure have been used at the community level to distinguish disturbed and undisturbed forests (Mladenoff et al. 1993) and to distinguish species associations and environmental signals (Wiegand et al. 2007b). Primarily, though, their use has focused on the description of forest structure, often to inform forest management policies. This is true of the spatial Simpson index (Shimatani 2001; Shimatani & Kubota 2004) and a version of the mingling index that is limited to data from four nearest neighbours (Gadow 1993; Motz et al. 2010). Our finding that all of our modelled processes can be distinguished through their implications for exposure has therefore not been anticipated.

<table>
<thead>
<tr>
<th>Janzen-Connell</th>
<th>Neutral</th>
<th>Heteromyopia</th>
<th>Niche</th>
<th>Lottery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scattering</td>
<td>High; highly consistent across species</td>
<td>Low at very small scales but higher at intermediate scales</td>
<td>Very low; consistent across species</td>
<td>Low but highly variable between species</td>
</tr>
<tr>
<td>Exposure</td>
<td>High; highly consistent across species</td>
<td>Low at very small scales but similar to neutral at intermediate-large scales</td>
<td>Low at small scales but high at intermediate-large scales</td>
<td>Low but highly variable between species</td>
</tr>
<tr>
<td>Beta-diversity (turnover)</td>
<td>Rapid rate of turnover with distance</td>
<td>Rapid at small scales, neutral at larger scales</td>
<td>Rate of turnover similar to neutral</td>
<td>Highly variable</td>
</tr>
</tbody>
</table>

Table 4.10: Principal characteristics of spatial patterns generated under each model in terms of scattering, exposure, and implications for turnover of species with distance (Beta-diversity).

Results are expressed relative to the neutral model, which has intermediate levels of scattering, exposure and turnover due to density-dependent mortality and dispersal limitation.
4.4.3 LEVEL OF INFORMATION

Many of the measures of scattering and exposure considered here are able to distinguish some of the models we consider from their resulting spatial patterns. Without exception, their ability to do so is maximised when they are expressed at the community level. Few exist at the individual level, and none of these are found to be useful for this purpose, being dominated by inter-individual variation. At the species level, many of the measures become highly informative about differences in spatial structure between the models, and are particularly valuable for assessing whether these differences affect all species equally. It is at the community level, however, that the signals of the models emerge most clearly, presumably because they do not suffer from the confounding variation that occurs even between abundant species in stochastic models of the kind used here.

Not all measures are equally informative at community level, however. Those that best describe differences in scattering are the measure of interspecific segregation and the proportion of conspecific neighbours. The latter makes use only of individual-level neighbour counts; the former of scale-independent species-level neighbour counts and global proportions. Both therefore depend on some of the lowest level information used by any of the measures we consider: counts of conspecific and heterospecific neighbours.

The most successful measures of exposure are the mingling index, xPOD and the spatial Simpson index. Of these, only the mingling index is able to distinguish all models, while the xPOD and spatial Simpson index distinguish all but two. Once again, these measures are comprised of counts of neighbours made at the individual level and averaged up to species level, before construction of the measures themselves. The xPOD and spatial Simpson index further make use of some global information – species sizes and proportions, respectively. The comparative (and unexpected) failure of the degree of association is presumably due to the lack of information it conveys about inter-specific variation, in which respect it differs from the xPOD.

It is already known that the identity of nearest neighbours is an important indicator of ecological process (e.g. Clark & Evans 1954; Pielou 1961; Kint et al. 2003; Murrell & Law 2003; Wiegand et al. 2007b; Vogt et al. 2010). Indeed, far more complex data may be little more informative about spatial structure or its causes (Moeur 1993), and techniques for reconstructing full spatial point patterns from nearest neighbour data have been developed (e.g. Pommerening & Stoyan 2008). Despite this, links to particular processes have not previously been investigated, with the exception of studies that indirectly use neighbour identities to search for density-dependence, Janzen-Connell or competition effects (e.g. Duncan 1991; Hubbell et al. 2001; Peters 2003; Uriarte et al. 2004; Chapter 5). Apart from a general recognition that community-level summaries are best suited to studying processes that operate at the community level (e.g. Luo et al. 2012), the use of neighbour data averaged across individuals and species has seldom been considered.

4.4.4 GENERAL

In Chapter 2, the cross-pair overlap distribution was found to successfully discriminate between neutral/heteromyopia, niche, lottery and Janzen-Connell processes in a modelled forest community. At a general level, our results show that the spatial signals detected by the xPOD are strong and consistent consequences of the processes we modelled. Some aspect of the same spatial behaviour is detected by all the statistics that we consider, at every level and in terms of both scattering and exposure.

Specifically, the niche and (to a lesser extent) lottery models produce low levels of scattering and exposure, with highly clumped and poorly mingled species. The lottery model, though, produces more variable results across species and communities. This means that it is particularly difficult to
establish whether any single pattern was produced by this or another model. The neutral and heteromyopia models have similar but more tightly constrained spatial properties in which both scattering and exposure are intermediate between the niche and Janzen-Connell models. In the latter, species are widely distributed and uniformly mingled, giving high levels of scattering and exposure. These behaviours are at their most variable in the niche and lottery models, and at their least variable in the Janzen-Connell model (Table 4.10).

It is notable that these findings are consistent with levels of diversity and evenness in the modelled communities (see Chapter 2; Table 2.2), both of which broadly – if not reliably – increase through the lottery, niche, heteromyopia, neutral and Janzen-Connell models. As previously discussed, this is partly due to differences in the modelled fecundities of species, and measures of spatial structure are normalised to this in some way (or can be, as in the case of the ISAR). These first-order (non-spatial) properties do, nevertheless, have consequences for spatial structure that could be used to anticipate some of the findings of this chapter, particularly those relating to exposure, in which the level and consistency across species of exposure may be related to diversity and evenness, respectively.

These consequences are not fixed, however, and first-order measures do not provide a reliable indication of spatial structure, as demonstrated by the failure of the species abundance distribution to distinguish the models used here (Chapter 2). Furthermore, they offer no information on changes in spatial structure with scale, the properties of particular species, or levels of scattering. The most important example of this is the inability of first-order measures to distinguish the neutral and heteromyopia models, in contrast to the spatial statistics used here. Spatial statistics that offer the greatest advantage over diversity and evenness measures are therefore those that can operate at particularly small scales, where abundances can vary widely from their mean, or those that describe scattering and are fully normalised to overall abundances. These are capable of isolating spatial signals not detectable by any first-order methods.

The full description of these spatial signals allows contrasting predictions about spatial structure to be derived from the different theories, and for the empirical testing of these predictions. This has been identified as an important objective by studies of coexistence mechanisms (e.g. Amarasekare 2003) but has not previously been attempted, with the majority of comparative studies considering only the relative response of different measures to defined spatial structure (e.g. Kint 2005). Here, in contrast, we show not only that different measures are sensitive to different aspects of spatial structure, but that this may be used to develop robust methods of inferring underlying process from resulting spatial pattern. These may be best suited to comparisons in which the similarity of multiple spatial patterns can be assessed rather than the attribution of single patterns to any one process, but do nevertheless allow for inference of this specific kind.

In fact, we suggest that the likelihood of a given spatial pattern having been produced by any particular process may be estimated, especially where measures of scattering and exposure are combined. We therefore propose the combination of three metrics for the identification of underlying process in plant communities: the measure of interspecific segregation, the mingling index, and the standard deviation of the xPOD. Expressed at the community level and with an abundance threshold of 500 individuals per species, these describe all the aspects of spatial behaviour which we have found to vary with modelled process. The measure of interspecific segregation is scale-invariant and so does not require arbitrary selection of a scale at which to apply it; the mingling index uses slightly larger-scale information but works best in immediate neighbourhoods of tens of individuals; and the xPOD works over larger scales still, is relatively robust to variation in this scale, and, in being summarised through its standard deviation, provides information about inter-specific variation in patterns which the others do not.
When the results generated in this study are combined in this way and plotted (Fig. 4.21), results from the five models do indeed separate out in parameter space. Variation within models reduces this separation—particularly in the case of the lottery model—but does not entirely confound it. Together, these three statistics appear able to successfully and reliably distinguish all modelled processes, and their application to empirical data as a three-dimensional discriminator of underlying process is expected to provide valuable insights into the dynamics of plant communities.

Figure 4.21: A 3-D plot of results from each model based on mean (community-level) values of the measure of interspecific segregation, mingling index, and standard deviation of the xPOD.

Results are coded as follows: JC = Janzen-Connell, Nt = neutral, H = heteromyopia, L = lottery, Ni = niche.

To achieve this, further work is required in developing models of processes of interest and in generating sufficient realisations of these models so that multivariate distributions of the values taken by the three statistics in each case can be described. This is similar to previous and ongoing work intended to capture the range of spatial patterns produced by particular processes (e.g.
Chapter 2; Wiegand et al. 1998; 2003; DeAngelis & Mooij 2005; Grimm et al. 2005). The combined measures can then be applied to empirical data from diverse plant communities (e.g. Losos & Leigh 2004; Chapter 3) and the similarity of the spatial patterns observed to those generated by the modelled processes can be found.

This would not, of course, allow the certain identification of any one underlying process, especially given the potential strength of intra- and inter-specific variation in environmental responses and other relevant traits (Clark 2010; Clark et al. 2010). It may be necessary to investigate further the effects of such variation on the spatial signals of the processes modelled here, but preliminary findings suggest that the signals may be robust to realistic levels of variation (Appendix 2.2). The methods presented here are in any case expected to be beneficial in eliminating possible causes of spatial structure, and hence species coexistence, that first-order or single second-order statistics are unable to (Chapter 2; Chapter 3).

4.5 Conclusion

In recent years, increasing attention has been paid to links between coexistence mechanisms and spatial patterns (Pacala & Deutschman 1995; Tilman & Kareiva 1997; Murrell et al. 2001; Stoll & Prati 2001). However, the full implications of each on the other remain uncertain (e.g. Rejmánek 2002; Chesson & Neuhauser 2004). This work represents the first attempt to assess the full consequences of all the principal theories of coexistence on spatial structure.

We find that the spatial signals of the processes modelled here are strong and detectable with a range of spatial statistics. These signals have different, though related, implications for the scattering and exposure of species, and description of both of these aspects allows the models to be reliably distinguished from one another. The best performing measures of scattering are the measure of interspecific segregation and the proportion of conspecific neighbours; of exposure, the mingling index, xPOD and spatial Simpson index. All of these are most informative when expressed at the community level, but all share a reliance on counts of conspecific and heterospecific neighbours taken in local neighbourhoods at the individual level. This low-level information, when combined and averaged, therefore appears to be the most powerful discriminator between the effects of different processes. This further implies that the inference of underlying process from resulting spatial pattern in plant communities is a robust and reliable technique, possible through a number of existing spatial statistics and, especially, combinations of these.
Appendix 4.1: Measures of β-diversity

4.6.1 INTRODUCTION

The measures of β-diversity considered here all utilise counts of species represented in areas or quadrats (Table A4.1). As a result they are scale-independent assuming a set scale for quadrat size, and occur only at the community level. Every measure from the review paper of Koleff et al. (2003) was used, but only those which proved particularly informative or representative are presented here. In order to calculate their values, the area in which each simulated pattern occurs is divided into 400 cells of equal area, which are then treated as quadrats. This number of quadrats was chosen so that approximately 10 to 50 individuals were contained in each, as might be the case in real-world rainforest quadrats on a scale of several metres. To test the behaviour of the measures with changing spatial scale they are calculated for four different separations of quadrats, beginning with those that neighbour one another (Fig. A4.1). In operating at the community level, measures of β-diversity are insensitive to species-specific behaviour, and so are expected to discriminate best between processes which produce different levels of diversity at the community scale.

4.6.2 RESULTS

Measures of β-diversity describe some aspects of both scattering and clustering, concerned as they are with the rate of species turnover with distance. The measures are divisible into three main groups by their response to the models presented here. These groups are typified by the plots in Figure A4.2: those that increase with radius; decrease with radius; and show little change with radius. This last group appears to be the most sensitive to our models, showing both the largest differences between them and the smallest variation within them. The niche and lottery models are found to have the lowest levels of β-diversity and the Janzen-Connell model the highest. However, this group consists of measures which exclusively or primarily use information on the numbers of unique species in each quadrat, making them sensitive to overall community size. It is this signal that generates the differences between models found by this group of measures.

This is confirmed by fits of power laws to the community-level results of each simulation (Fig. A4.3). When divided by underlying model, estimated values of the power law parameter α are found to differ most greatly when using the measure of Magurran (1988). Here, the Janzen-Connell, neutral/heteromyopia and niche/lottery model groups are entirely distinct, with substantial differences in means and variances even within these. Where measures include information on the numbers of shared species between quadrats, however, so preventing differences in modelled fecundities of species from dominating the resulting signal, these differences are eroded. At best, the neutral, Janzen-Connell and neutral/lottery/heteromyopia groups are distinct.

4.6.3 DISCUSSION

The measures of β-diversity which perform best in distinguishing underlying model are those which emphasise unique counts of species. However, this is largely due to differences in modelled fecundities of species, and not to any inherent characteristics of the modelled processes themselves. Measures which do not suffer from this bias include information about species which are shared between sites, and are therefore normalised to some extent by community size. These, however, are less successful in distinguishing underlying model. Despite this, they do suggest that the rate of species turnover with distance is greatest in the niche and, to a lesser extent, Janzen-Connell models (and most variable in the lottery model). Once again, this can be traced to geographical environmental variation in the niche model and uniform mixing of species in the Janzen-Connell model.
<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
</table>
| 1. Whittaker’s measure      | \[
\frac{N_{ab} + N_a + N_b}{(2N_{ab} + N_a + N_b)/2}
\]                      | Whittaker (1960)  |
| 2.                          | \[
\frac{N_a + N_b}{2}
\]                                                  | Cody (1975)       |
| 3.                          | \[
N_a + N_b
\]                                                      | Weiher & Boylen (1994) |
| 4.                          | \[
\frac{(N_{ab} + N_a + N_b)^2}{(N_{ab} + N_a + N_b)^2 - 2N_aN_b}
\]            | Routledge (1977)  |
| 5.                          | \[
\frac{\log(2N_{ab} + N_a + N_b) - \frac{2N_{ab}\log2}{2N_{ab} + N_a + N_b}}{2N_{ab} + N_a + N_b}
\] | Routledge (1977)  |
| 6.                          | \[
\frac{\exp\left(\log(2N_{ab} + N_a + N_b) - \frac{2N_{ab}\log2}{2N_{ab} + N_a + N_b}\right)}{2N_{ab} + N_a + N_b}
\] | Routledge (1977)  |
| 7.                          | \[
\frac{N_a + N_b}{2N_{ab} + N_a + N_b}
\]                                      | Wilson & Shmida (1984) |
| 8. Jaccard’s index          | \[
\frac{N_{ab}}{N_{ab} + N_a + N_b}
\]                                      | Jaccard (1912)    |
| 9.                          | \[
\frac{2N_{ab}}{2N_{ab} + N_a + N_b}
\]                                      | Sorensen (1948)   |
| 10.                         | \[
\frac{(2N_{ab} + N_a + N_b){\left(1 - \frac{N_{ab}}{N_{ab} + N_a + N_b}\right)}}{N_{ab} + N_a + N_b}
\] | Magurran (1988)   |
| 11.                         | \[
\frac{\min(N_{ab}, N_b)}{\max(N_{ab}, N_b) + N_{ab}}
\]                      | Harrison et al. (1992) |
| 12.                         | \[
1 - \frac{N_{ab}(2N_{ab} + N_a + N_b)}{2(N_{ab} + N_a)(N_{ab} + N_b)}
\]           | Cody (1995)       |
| 13.                         | \[
\frac{N_a + N_b}{N_{ab} + N_a + N_b}
\]                                      | Colwell & Coddington (1994) |
| 14.                         | \[
\frac{\min(N_{ab}, N_b)}{N_{ab} + N_a + N_b}
\]                                      | Williams (1996a)  |
| 15.                         | \[
\frac{N_aN_b + 1}{((N_{ab} + N_a + N_b)^2 - (N_{ab} + N_a + N_b))/2}
\] | Williams (1996a) |
| 16.                         | \[
1 - \frac{2N_{ab}}{2N_{ab} + N_a + N_b}
\]                                      | Harte & Kinzig (1997) |
| 17.                         | \[
\frac{N_{ab}}{N_{ab} + N_b}
\]                                      | Ruggiero et al.   |
Table A4.1: Definitions of measures of $\beta$-diversity (individual terms are given in Table 4.1)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>(1998)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.</td>
<td>(\frac{\min(N_a, N_b)}{\min(N_a, N_b) + N_{ab}})</td>
<td>Lennon et al. (2001)</td>
</tr>
<tr>
<td>19.</td>
<td>(\frac{2</td>
<td>N_a - N_b</td>
</tr>
<tr>
<td>20.</td>
<td>(1 - \log\left(\frac{2N_{ab} + N_a + N_b}{N_{ab} + N_a + N_b}\right) / \log 2)</td>
<td>Lennon et al. (2001)</td>
</tr>
</tbody>
</table>

Figure A4.1: The labelling of cells or quadrats used for the calculation of measures of beta-diversity.

Each measure is calculated for five different quadrat separations as shown here.
Figure A4.2: Example results for measures of beta-diversity, typical of the three types of curve produced by the 21 measures considered.

Shown are means across realisations (solid lines) with 95% limits (dashed lines). Measures are calculated for pairs of quadrats separated by 0, 1, 2, 3, and 4 uncounted quadrats.
Figure A4.3: Boxplots of estimated values of parameter alpha in the power law which best fits results for 3 measures of beta diversity in each model.

Results based on 10 realisations of each model. Models are identified as follows: 1 – Neutral; 2 – Niche; 3 – Lottery; 4 – Janzen-Connell; 5 – Heteromyopia. Numbering of measures refers to numbers in Table A4.1.
CHAPTER 5 - Neighbourhood effects on mortality and diversity in a tropical rainforest
5.1 Introduction

In this chapter we investigate the occurrence of small-scale Janzen-Connell effects of various forms in a single tropical rainforest plot using Bayesian modelling. We do so in order to assess the likely strength of spatial signals of the kind identified in previous chapters generated by Janzen-Connell effects. Here we focus on tree mortality, which can be caused by numerous factors in tropical rainforests that are very hard to disentangle and identify. The effort is justified, however, by the search for processes that maintain the exceptional levels of diversity found in the tropics. This is especially true if density-dependent mortality is as strong and ubiquitous as has been suggested (Lambers et al. 2002; Peters 2003). If the intensity of density-dependent mortality does not vary substantially between species then its role in supporting diversity must be limited, but if it does, and particularly if it makes abundant species disproportionately likely to die, then it may play a major role (Connell et al. 1984).

There are convincing reasons to expect just such an effect. Common species may support large populations of specialised pests and pathogens that can inflict substantial damage and mortality, but that do not affect rare species to the same extent (Ridley 1930; Leigh et al. 2004). On the local scale, this can lead to 'overcompensating' mortality, under which young trees cannot survive in the immediate vicinity of their parents or other conspecifics (Janzen 1970; Connell 1971). Known as the Janzen-Connell effect, this would promote the spatial mixing of species on small spatial scales, and coexistence at larger scales (Wright 2002; Petermann et al. 2008; Chapter 1).

As a plausible and potentially fundamental process in tropical forests, the Janzen-Connell effect has generated intense interest. Evidence for it has been limited and inconsistent, however. Interspecific effects of density on growth rates and asymmetric size competition have been recognised for some time (e.g. Bella 1971; Diggle 1976), and are regularly modelled (Kohyama 1992; Canham et al. 2006; Freckleton & Lewis 2006). Some intraspecific interactions have been shown to be detrimental, but the majority appear to be beneficial or neutral (Connell et al. 1984; Condit et al. 1992; Hyatt et al. 2003; Bagchi et al. 2011). Even where negative effects of conspecific neighbours exist, they may be so weak as to be unimportant (Paine et al. 2008; Svenning et al. 2008). Some evidence even suggests that rare species suffer stronger conspecific density dependent mortality than common species (Comita et al. 2010).

However, it is possible that Janzen-Connell effects are often hidden by other processes. In particular, environmentally-driven mortality and recruitment can both produce patterns that resemble or confound those of density-dependence (e.g. Menge & Sutherland 1987; Condit 1995; Leigh et al. 2004; Kraft et al. 2008). The Janzen-Connell effect may also take several different forms. Originally thought to operate only within species, it might also spread beyond taxonomic boundaries to affect several species that are similar in some way. This similarity may be in terms of environmental niche preference, meaning that species tend to co-occur geographically and represent a single broad target for agents of mortality (e.g. Cavender-Bares & Wilczek 2003; Canham et al. 2006). It may also be phylogenetic, so that related species are all suitable for specialised pests, pathogens or predators (Wills et al. 1997; Carson et al. 2008). Considerable evidence supports the second of these hypotheses, from findings that disease risk is inversely proportional to species diversity because of pathogen dilution (Mayer & Pimm 1997; Haas et al. 2011), to findings that explicitly link phylogenetic relatedness to increased mortality (Boyden et al. 2008; Paine et al. 2011; Yguel et al. 2011). Direct studies of pest species also identify specialisation to particular families or genera of prey (e.g. Barone 1998; Novotny et al. 2002a). Agents of mortality may nevertheless be influenced by unexpected or unobservable factors, so affecting groups of species that have no obvious relationship to one another (e.g. Coley & Barone 1996; Van Bael et al. 2003; Richards & Coley 2007).
A wide range of techniques have been used to investigate Janzen-Connell effects. Recently, spatial statistics have shown some promise for this (Chapter 2) and may be able to distinguish patterns produced by Janzen-Connell and environmental effects in empirical data (Bagchi et al. 2011). Several studies have attempted to do so by analysing small-scale variation in functional traits, but find evidence for both effects (Kraft et al. 2008; Cavender-Bares et al. 2009; Paine et al. 2011). More widely used are the small-scale experimental manipulation of densities (e.g. Connell 1970; Augspurger 1983; Schupp 1992; Petermann et al. 2008; Svenning et al. 2008; Bagchi et al. 2010) or statistical model fitting to look for evidence of effects in growth or mortality rates (e.g. Uriarte et al. 2004a; 2004b; Ruger et al. 2011; Lin et al. in press). Both approaches have limitations: the former is time-consuming, costly and only possible for a very small number of species at one time, while the latter requires large amounts of data and is computationally expensive, again restricting the number of species which can be considered in practice. Nevertheless, the availability of detailed long-term datasets from large rainforest plots (e.g. Condit 1998; Losos & Leigh 2004) makes the second approach increasingly attractive. Some plots, such as that on Barro Colorado Island in Panama (Hubbell et al. 2005), have been regularly censused over a period of 30 years, producing data ideally suited to studies of growth or mortality. Meanwhile, maximum-likelihood approaches enable the robust comparison of many competing models (Johnson & Omland, 2004), which can even produce meaningful results for very rare, data-poor species when hierarchical Bayesian modelling is used (Clark, 2005).

These methods have been widely applied in studies of patterns of recruitment, growth and mortality in forests (e.g. Monserud, 1976; Ribbens et al. 1994; Wyckoff & Clark 2000; Uriarte et al. 2004; 2005; 2010). However, while strong age- or diameter-related effects have been identified, small-scale density-dependence and Janzen-Connell effects have rarely been investigated because they require many more data and more complex models to describe each individual’s neighbourhood. Only a small number of species and forms of the Janzen-Connell effect have been considered as a result. In addition, studies have tended to focus on recruits and saplings as the main subjects of density-dependent mortality, overlooking the effects of pests and pathogens that are not restricted by host age (Carson et al. 2008).

Here, we use hierarchical Bayesian maximum-likelihood modelling to investigate these neglected aspects of Janzen-Connell effects. We test for the occurrence and form of Janzen-Connell effects in all species within a tropical rainforest plot, considering the impact of small-scale neighbourhood spatial structure on individual mortality. We model four kinds of density-dependence within neighbourhoods: one that does not vary with species identity, one that varies in strength between conspecific and heterospecifics (the standard Janzen-Connell model), one that varies in strength according to the relatedness of species (a phylogenetic Janzen-Connell model), and one that varies in strength according to the distance between species on an axis of similarity used to express undefined factors that determine the suitability of different species to agents of mortality (a Janzen-Connell axis model). The identity of this axis is not specified so that species can be located along it to maximise model fit, but it may, for example, describe the likelihood of co-occurrence or the similarity of particular traits between species. In comparing these models we expect to discover whether observed mortality is affected by Janzen-Connell dynamics and, if so, what form they take and whether they affect all species equally.

### 5.2 Methods

#### 5.2.1 STUDY AREA/DATA

We used data from the 50ha plot on Barro Colorado Island (BCI) in Panama (9°9'N, 79°51'W), which is a 1,567 ha island in the Panama Canal forested with tropical lowland moist forest (Hubbell et al. 2005).
The location, status (alive or dead), diameter at breast height (dbh, with breast height defined as 130 cm from the ground) and taxonomic identity of all free-standing woody plants with dbh ≥ 1 cm were recorded first in 1982 and then every 5 years from 1985 onwards. We use the 1990 census for size, identity and neighbourhood information, derive growth rates (change in dbh in mm between censuses) from the 1985 census (plants not alive in 1985 are omitted from the analysis) and mortality from the 1995 census. Altogether, 15% of the individuals that were alive in 1990 died during this period, and 88% of species suffered at least one mortality. Approximately 6.5% of the growth rates were negative, presumably because of recording errors in one or both of the 1985 and 1990 censuses. We assume a similar number of undetectable errors in the positive growth rates and so do not arbitrarily alter negative rates as previous studies have done (e.g. Uriarte et al. 2004; Ruger et al. 2011). We also assume that deaths, being easier to classify, are accurately recorded.

The climate and environment on BCI are relatively uniform with the exception of rainfall, which varies between averages of 23 mm in March and 364 mm in October. Mean monthly minimum temperatures vary between 22.8°C (January) and 23.7°C (May), while elevation varies between 120 and 160 m above sea level. The plot is mainly comprised of old-growth forest except for 2 ha of 100 year-old secondary forest. Human disturbances are substantial, including historical harvesting of mahogany, some settlement activity over 1000 years ago and much more recent isolation (BCI only became an island upon construction of the Panama Canal), defaunation, species introductions and intense research activity (Sheil & Burlsem 2003). Natural disturbances, though, are minimal, and include occasional wind-storms which contribute to an estimated 126-year recurrence of treefall at any particular point in the forest. Severe droughts associated with El Niño cause the most substantial natural disturbances, dramatically increasing recruitment and mortality roughly every 10-20 years (Leigh et al. 2004b). One such event occurred in 1983. Altogether, 244,016 stems of 304 species were included in our analysis. The dbh, taxonomic information and growth rate of each of these individuals was recorded, along with the number, dbhs, distances and taxonomic information of all neighbours with dbh ≥ 1 cm within 5 m. Neighbourhood size was limited by computational power, but 5 m has previously been identified as the approximate range of autocorrelation in density effects (Uriarte et al. 2004) and Janzen-Connell effects are expected to be detectable at this scale (Condit et al. 1992). The mean number of neighbours was 41.

We used phylogenetic data for angiosperm families from Davies et al. (2004) to construct a species phylogeny for 268 of the 304 species on BCI, using the Phylomatic software (Webb & Donoghue 2005) and R package ape (Paradis et al. 2004). Inter-family distances not included in these data were set to the average value. As in previous uses of this data (e.g. Kembel & Hubbell 2006, Hardy 2008), genera and species within families were represented as polytomies with constant node ages, but in contrast to previous studies we did not impose values on these, instead selecting them through maximum likelihood modelling.

5.2.2 MODEL DESIGN AND SELECTION

We use maximum likelihood estimation to fit hierarchical and non-hierarchical Bayesian models of mortality for the BCI plot data (Hillborn & Mangel 1997; Johnson & Omland 2004). Hierarchical models included nested terms to ensure that parameter values with large uncertainties conformed to distributions defined by well-supported values, with variable mean and variance (see below). All models were of logistic form,

\[ \mu = \frac{1}{1 + e^{-\kappa}} \]  

**eqn 5.1**
where \( \mu \) is the probability of death and \( K \) is a linear combination of mortality terms as described below. For each combination of terms, parameters were set to the values that maximised the likelihood of predicted mortality matching observed mortality. The full list of terms to be included in the models (Table 5.1) was determined prior to the analysis, although the form of some was varied in order to maximise their fit to the data. They include an intrinsic mortality term, a \( \text{dbh} \)-dependent term, a shading term, and neighbourhood terms. Different forms of the \( \text{dbh} \) term \( k_1 \) were tested, and the form

\[
k_1 = a \cdot \log(\text{dbh}) + b \cdot \text{dbh} \quad \text{eqn 5.2}
\]

was found to best describe the data (Fig. A5.2). The shading term used the Crown Area Index for each individual, an estimate of the number of canopies extending above the individual calculated by the relationship

\[
d_c = 0.1 \times \text{dbh} \quad \text{eqn 5.3}
\]

where \( d_c \) is crown diameter in metres and \( \text{dbh} \) is in cm. Where this diameter exceeds the distance between the focal individual and the neighbour, a value of one is added to the Crown Area Index.

Neighbourhood terms included a simple count of neighbours of the form \( j_1. n \), a neighbour total basal area term \( j_2. n_{\text{area}} \), a Gaussian neighbour distance term \( j_3. e^{-\left(\frac{d_{\text{dist}}}{\sigma_{\text{dist}}}\right)^2} \), a neighbour-\( \text{dbh} \) term \( j_4. n_{\text{dbh}}^\beta \), and three different Janzen-Connell effect terms to describe the alternative models detailed above.

The basic Janzen-Connell effect was modelled by a term that varied depending on whether or not neighbours were conspecifics of the focal individual, \( j_5(\text{con}) \) or \( j_6(\text{hetero}) \). The second form, in which interaction strength varied according to the distance between species along an undefined axis, was modelled by allowing each species to take a value chosen by the maximum-likelihood model in a \([0,1]\) interval, and then by calculating the distances between these values when species shared a neighbourhood. Values on this axis could represent traits such as heights or leaf characteristics, environmental niche preferences, or any other factor of potential relevance to mortality agents. The form of interaction between species was Gaussian in shape, and its strength depended upon the distance \( j_{\text{dist}} \) of species from one another on the axis and a standard deviation parameter \( \sigma_{jC} \). Because the values assigned to each species were parameters in the model, these distances changed during model fitting.

\[
j_7. e^{-\left(\frac{j_{\text{dist}}}{\sigma_{jC}}\right)^2} \quad \text{eqn 5.4}
\]

The third, phylogenetic Janzen-Connell effect was modelled as a Gaussian response to phylogenetic distance, using true inter-family distances and variable parameters for family-genus and genus-species distances, with a standard deviation parameter as above.

\[
j_8. e^{-\left(\frac{\text{phylo.dist}}{\sigma_{\text{phylo}}}\right)^2} \quad \text{eqn 5.5}
\]

The neighbourhood model was therefore of the form
\[ k_n = j_1(n) \times j_2(n_{area}) \times j_3 \cdot e^{\left(\frac{n_{dist}}{\sigma_{dist}}\right)^2} \times j_4 \cdot n_{dbh}^\beta \times \sum_{i=1}^n \left\{ \begin{array}{ll} j_5(\text{con}) & \text{or} \ j_6(\text{hetero}) \\ j_7 \cdot e^{\left(\frac{\text{phylo}_{dist}}{\sigma_{phylo}}\right)^2} & \end{array} \right. \]  

_eqn 5.6_

and the full mortality model was

\[ K = k_0 + k_1(ddbh) + k_2(CAI) + k_3(growth) + \sum_{nghbrs} k_n \]  

_eqn 5.7_

where \( K \) is a measure of mortality of one individual as in equation 5.1 and \( n \) is the number of neighbours of that individual.

Every term in the model was able to be global or species-specific and hierarchical or non-hierarchical. Where hierarchies were used we assumed normal hyperdistributions, having found that values for common species with tightly constrained credible intervals were approximately normally distributed (as did Ruger et al. 2011). Models were constructed term by term, with the term and form (global or species-specific, hierarchical or non-hierarchical) that produced the greatest improvement in model fit at each stage being retained. No pre-selection of terms was carried out. Altogether 92 different models were run, each taking between 0.5 and 36 hours. All the combinations of terms that could have improved model fit were considered, except for meaningless combinations of terms (such as the different Janzen-Connell models) that would have violated the implicit assumption in the model selection process that all models under comparison could be accurate (Johnson & Omland 2004).

<table>
<thead>
<tr>
<th>parameter</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic mortality term ( k_0 )</td>
<td>-3.5 - -1.0</td>
</tr>
<tr>
<td>log(ddbh) coefficient ( a )</td>
<td>-1.0 – 0.0</td>
</tr>
<tr>
<td>dbh coefficient ( b )</td>
<td>0.0 – 0.05</td>
</tr>
<tr>
<td>Crown Area Index coefficient ( k_2 )</td>
<td>-0.1 – 0.1</td>
</tr>
<tr>
<td>growth coefficient ( k_3 )</td>
<td>-0.1 – 0.05</td>
</tr>
<tr>
<td>number of neighbours coefficient ( j_1 )</td>
<td>-0.1 – 0.1</td>
</tr>
<tr>
<td>neighbour basal area coefficient ( j_2 )</td>
<td>-0.5 – 0.5</td>
</tr>
<tr>
<td>neighbour distance coefficient ( j_3 )</td>
<td>0.0 – 0.2</td>
</tr>
<tr>
<td>neighbour distance sensitivity ( \sigma_{dist} )</td>
<td>0.0 – 0.05</td>
</tr>
<tr>
<td>neighbour dbh coefficient ( j_4 )</td>
<td>-0.01 – 0.01</td>
</tr>
<tr>
<td>neighbour dbh power ( \beta )</td>
<td>0.0 – 1.1</td>
</tr>
<tr>
<td>conspecific neighbour coefficient ( j_5 )</td>
<td>-0.05 – 0.05</td>
</tr>
<tr>
<td>heterospecific neighbour coefficient ( j_6 )</td>
<td>-0.05 – 0.05</td>
</tr>
<tr>
<td>Janzen-Connell axis value ( JC_x )</td>
<td>0.0 – 1.0</td>
</tr>
<tr>
<td>Janzen-Connell axis distance coefficient ( j_7 )</td>
<td>-0.2 – 0.2</td>
</tr>
<tr>
<td>Janzen-Connell axis sensitivity ( \sigma_{JC} )</td>
<td>0.0 – 0.3</td>
</tr>
<tr>
<td>Phylogenetic distance coefficient ( j_8 )</td>
<td>-0.2 – 0.2</td>
</tr>
<tr>
<td>Phylogenetic distance sensitivity ( \sigma_{phylo} )</td>
<td>0.0 – 0.3</td>
</tr>
<tr>
<td>species-genera phylogenetic distance ( d_{sp-gen} )</td>
<td>0.0 – 1.0</td>
</tr>
<tr>
<td>genera-family phylogenetic distance ( d_{gen-fam} )</td>
<td>0.0 – 1.0</td>
</tr>
</tbody>
</table>
Table 5.1: The identities and allowed ranges of all model parameters.

Ranges were constrained only to ensure that non-infinite values were generated for the combinations of terms considered.

Models were coded in C++ and fitted using the Filzbach software package (Microsoft Corporation 2012), which provides versatile C++ libraries for rapid maximum-likelihood estimation (see e.g. Dalchau et al. 2011; Joppa et al. 2011). Filzbach employs an adaptive Markov Chain Monte Carlo (MCMC) analysis with phases for burn-in, sampling, and for maximum likelihood analysis using a likelihood profile method (Microsoft Corporation 2011). The adaptive MCMC uses a Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970) to explore the posterior distribution of the parameters, with samples from the burn-in discarded. The number of iterations following burn-in is specified (below), and each of these generates a sample of a vector of parameters from the joint posterior distribution. These can be regarded as dependent samples from the posterior distribution. Because of memory limitations we thin these samples, recording every 100th vector along with its corresponding log-likelihood. The maximum likelihood phases generate further samples from the joint posterior distribution using a targeted likelihood profile method. We ran the MCMC algorithm for 200,000 iterations and the maximum likelihood phase for a further 100,000 iterations. Together, these form a single analysis chain, and we checked for convergence within each chain and between two independent chains for each model using the Brooks-Gelman-Rubin convergence diagnostic (Gelman & Rubin 1992; Brooks & Gelman 1998).

Prior knowledge about possible parameter values is limited or absent, with only the range of values that produce tractable likelihoods known. Non-informative priors are therefore placed on the parameters. These are Gaussian in form and in practice have far greater variance than the possible ranges of the parameters, which are defined separately (a full list of parameters and their ranges is given in Table 5.1). A Bayesian analysis is nonetheless used because of the ease and speed with which it can be carried out, and the posterior means and 95% credible intervals estimated by the Metropolis Hastings phases are used below. Only half of the data (every other record, the ‘training’ data) was used for model fitting, with the other reserved for model selection.
Several model selection criteria were recorded, including the AIC, BIC, DIC and maximum likelihood of the model. However, we carried out final selection using the maximum-likelihood fit of the models to the reserved half of the data (the ‘test’ data) which the models had not previously been confronted with. The ability of the models to predict mortality in these data depended upon accurate description of mortality in the training data but without overfitting to its idiosyncrasies. This technique is used to prevent under- or over-fitting of models to one particular dataset and to avoid the consequent need for a term to penalise model complexity, the value of which is a largely subjective choice (Atkinson 1980; Guyan & Yao 1999; Pitt & Myung 2002).

5.3 Results

5.3.1 MODEL SELECTION CRITERIA

Different model selection criteria gave highly divergent results (Fig. 5.1). The maximum likelihood fit to the training data increased steadily with model complexity as expected, and was maximised by a model with the second-largest number of parameters of those we considered. Because of the large sample size, the AIC and DIC scores followed the fit to training data closely, and also selected this model. The BIC, though, selected one of the simplest models, and its value increased steadily with model complexity after this. The maximum likelihood fit to the test data, in contrast to all other selection criteria, improved rapidly with model complexity as the first terms were added to the models, more slowly after the two most successful terms were included, and then gradually decreased after being maximised by a model of intermediate complexity.

5.3.2 HIERARCHIES

The presence or absence of hierarchical terms had little effect on model selection (Fig. 5.2). Hierarchies improved the maximum likelihood fit to the training data because of the addition of likelihoods relating to the hierarchies used. The fit to test data did not take account of this, however, and maximum likelihood values remained almost identical. The reason for this is apparent from the posterior values of species-specific terms in hierarchical and non-hierarchical models, which show that hierarchies constrain parameter values, and hence improve model fit, only for the rarest species for which data is particularly scarce, as expected (Fig. 5.3). Because the overall likelihood is summed over individuals, the contribution of rare species is very small. The imposition of a hierarchy does, however, ensure that reasonable parameter values are selected for rare species while having little or no effect on the well-constrained parameter values of common species. It also provides a strongly justified global parameter in the hierarchical mean. For these reasons we chose hierarchical models over their non-hierarchical equivalents below.
Figure 5.1: Model selection criteria against model complexity.
(a): Maximum likelihood ranges for training data (black) and test data (red). Training data likelihood is maximised by a model with 1,836 parameters. (b): AIC, BIC and DIC rank, with the model identified as best by each criteria ranked 1 and the worst ranked 79.

The training data likelihood, AIC and DIC all selected an overfitted model with hierarchical and species-specific terms for intrinsic, dbh- and growth-dependent mortality terms, and for neighbourhood effects related to neighbour dbh and conspecific or heterospecific neighbour identity:

$$k_0 + k_1(dbh) + k_3(growth) + \sum_{nghb} j_1s \cdot n_{dbh} \beta_s \times \begin{cases} j_5_s(con) \\ j_6_s(hetero) \end{cases} \quad \text{eqn 5.8}$$

The BIC selected an underfitted model with global terms for intrinsic and dbh-dependent mortality, and a hierarchical species-specific term for the number of neighbours:

$$k_0 + k_1(dbh) + \sum_{nghb} j_1s \quad \text{eqn 5.9}$$

None of these selection criteria successfully identify the model which maximised fit to the test data.
Figure 5.2: Training and test data likelihoods for the five models with the highest test data maximum likelihoods, with (a) and without (b) hierarchies.

Figure 5.3: Posterior 95% credible intervals for the species-specific intrinsic mortality parameter $k_0$ with and without a hierarchy.

The imposition of a hierarchy substantially constrains credible intervals only for species with abundances less than 100.
The model that maximised the test data likelihood contained hierarchical and species-specific intrinsic, dbh- and growth-dependent terms, along with neighbourhood terms relating to neighbour dbh or to the phylogenetic Janzen-Connell effect:

\[ k_{0s} + k_{1s}(dbh) + k_{3s}(growth) + \sum_{\text{neighbrs}} \left\{ J_{4s} \cdot \left( \frac{\text{dbh}}{\text{adj}} \right)^{\beta_s} \right\} \]

The range of likelihoods produced by models with each of the neighbourhood terms overlapped considerably, while a model containing both terms gave slightly lower likelihoods that were nevertheless not significantly different from these (Fig. A5.1).

The highest maximum-likelihood values were produced by the model containing a neighbour dbh term \( J_{4s} \cdot \text{dbh}^{\beta_s} \), with the posterior distribution of \( \beta \) tightly constrained around a value of 0.6, but with highly variable values for the species-specific coefficient \( J_{4s} \). Crucially, these coefficients take both positive and negative values, implying that the presence of large neighbours increases mortality in some species and decreases it in others. Of the 304 species, coefficients for 49 were significantly non-zero, with 95% credible intervals that do not include zero; 30 of these are negative, suggesting a beneficial effect of large neighbours, and 19 positive (Fig. 5.4). The hierarchical mean was also significantly negative, providing strong evidence that the average effect of large neighbours was beneficial.

The model with a phylogenetic Janzen-Connell effect contained similarly variable parameter values. The family-genera and genera-species ages \( d_{\text{gen-fam}} \) and \( d_{\text{sp-gen}} \) were tightly constrained at values of 0.028 and 0.744 respectively, relative to the mean of the true inter-family values at 0.7836. The standard deviation term \( \sigma_{\text{phylo}} \), which controls the sensitivity of related species to one another, had a mean value close to 1.93, suggesting that species interact with conspecifics, congenerics and, to a certain extent, confamilials, but are unlikely to interact with heterofamilials. The species-specific coefficients \( J_{5s} \), however, took a range of positive and negative values like the coefficients in the neighbour dbh model above (Fig. 5.5). In this case, 61 of the 304 species’ parameters have non-zero values, 44 of which are negative, meaning that the presence of conspecifics or related species in a neighbourhood decreases mortality. Once again the hierarchical mean was significantly negative as well. Of the 61 species with non-zero coefficients in this model, 33 also had non-zero coefficients in the neighbour dbh model, and all of these were of the same sign in both cases.
Figure 5.4: Significant (non-zero) posterior 95% credible intervals for the neighbour diameter parameter $k_4$.

Species sizes (abundances) are on a log scale. Negative values imply a decrease in mortality.

Figure 5.5 Significant (non-zero) posterior 95% credible intervals for the neighbour phylogenetic relatedness parameter $k_{phylo}$.

Species sizes (abundances) are on a log scale. Negative values imply a decrease in mortality.
Discussion

Our results strongly support some general conclusions. Firstly, they demonstrate that the maximum likelihood of models against previously unencountered (test) data provides a robust criterion for model selection. The training data likelihood was maximised by one of the most complex models we considered, and the AIC and DIC also both selected this model. The test data likelihood, however, indicated that this model and others of similar complexity were strongly overfitted to the training data, with parameters describing uninformative noise rather than general processes. The BIC, in contrast, chose a strongly underfitted model with fewer terms and global, rather than species-specific parameters. This confirms the interpretation of the BIC as being suited to descriptive rather than predictive purposes, in that it rejects all but the terms that contribute most to model fit rather than accepting additional higher-dimensional approximations (Smith & Spiegelhalter 1980; Buckland et al. 1997). All of the model selection criteria that we considered chose a different model to that identified as the best by the test data likelihood, and all did so quite conclusively. Reliance on any of these would therefore have resulted in the selection of an inferior model. This reinforces existing doubts about the use of these criteria alone for model selection, especially given the implicit assumption (in the BIC at least) that a ‘true’ model of finite dimension exists and is contained in the suite of models being compared (Buckland et al. 1997; Johnson & Omland 2004).

We are also able to conclude that hierarchies usefully constrained parameters for rare species but made little difference to model selection in this case. None of the selection criteria we used showed substantial differences between equivalent hierarchical and non-hierarchical models. We nevertheless use hierarchical models rather than their non-hierarchical equivalents in order to generate robust and informative values for the hierarchical mean and variance and for parameters of rare species. It may be particularly important to derive accurate parameter estimates for rare species where Janzen-Connell effects are included, as rare species that are shade-tolerant and produce large seeds might be the most susceptible to such effects (Carson et al. 2008; Comita et al. 2010) – although this may be concealed to some extent by the imposition of a hierarchical parameter range.

Due to our use of test data likelihoods for model selection, we are unable to conclude that any one model offers the best description of mortality. This method of model selection is not in any case infallible, and its failure to identify a single superior model emphasises the general need for subsequent model interpretation (Pitt & Myung 2002). Here, two models with contrasting treatments of neighbourhood effects give the best, but similar, test data likelihoods (Fig. A5.1). Both contain the same individual-level effects, however. The first term to be included in both, providing the greatest initial increase in test data likelihood, is the species-specific intrinsic mortality term $k_{a}$. This clearly does not describe any particular process, but provides a background mortality rate which differs significantly between species. The second term, a species-specific dbh term, is more informative. Its form is shown in Figure A5.2, with high mortality at small dbhs (young ages) which declines steeply but subsequently increases again with increasing dbh. This agrees with frequent observations of elevated mortality levels in young and old trees when compared to mortality rates of trees of intermediate age. These, of course, are also the result of other processes, including competition at young ages and senescence and sensitivity to natural disturbances at old ages (Milton et al. 1994; Coomes & Grubb 2003; Moles & Westoby 2004). It is beyond the scope of this study to further distinguish these effects, however.

The final individual-level term is dependent upon growth over the previous census interval (1985-1990), and is also species-specific. As mentioned above, a large number of negative growth rates are recorded, and a correspondingly large number of errors in the positive rates are assumed. As a result, the values of parameters in this term must be interpreted with care, but their individual and hierarchical forms clearly show that mortality decreases with increasing growth. It is possible to
interpret this finding as relating to light regimes, with higher growth rates expected in more exposed areas. It is known that the light environment has an effect on mortality, with lower rates often found at higher irradiance (Denslow 1987; Wright 2002), although this certainly varies between species and many show the opposite effect (Kobe 1999; Ruger et al. 2011). In this context, the significant variation between species is particularly notable. The selection of the growth term rather than the shading term based on the Crown Area Index, however, suggests that the growth effect is more general, and not just a surrogate for light levels. Instead, it may be related to biotic stress factors such as competition, disease or herbivory, with light regimes only playing an incidental role. In any case, the presence and form of each of these individual-level terms in the most successful models agree with the findings of previous studies of mortality in BCI, including those that use the same census interval (Ruger et al. 2011; Uriarte et al. 2004).

The interpretation of the neighbourhood terms is more difficult, especially given the presence of contrasting terms in the two best models. The forms of relationships identified in each are convincing, with neighbour dbh effects scaling by a well-constrained parameter value of 0.6 and phylogenetic interactions occurring only at family level and below. However, the effects of these relationships on mortality are variable and unexpected, with a majority of significant effects acting to reduce mortality (Fig. 5.4). The effect of neighbouring trees is expected and known to vary with their size, with large trees having a greater effect on the growth and mortality rates of their neighbours than smaller ones (Kohyama 1992). Competition that is asymmetric with respect to size is also well documented (e.g. Bauer et al. 2004; Damgaard et al. 2002), although the form of our model does not allow for such an effect.

Previous evidence for phylogenetic Janzen-Connell effects is more limited, especially in studies of this size. Queenborough et al. (2009) point out that this is partly due to poor phylogenetic resolution below genus level in many cases, but identify conspecific and congeneric density-dependent effects within one rainforest tree family (Myristicaceae). Boyden et al. (2008) find that neighbour relatedness, together with individual size and the spatial structure of neighbourhoods, determines mortality in two related Eucalyptus populations, with competition being most intense between the most closely related individuals. Canham et al. (2006) find some evidence of Janzen-Connell effects which extend across similar species, but measure similarity in terms of environmental requirements rather than phylogenetic relatedness.

Considerable evidence exists of phylogenetic specialisation by pest species, however, to the level of particular families or, more often, genera, and rarely to either single species or more than one family (Barone 1998; Novotny et al. 2002a). These findings are consistent with the phylogenetic extent of the effect we identify here, which is strongest within species and genera, weaker within families, and does not occur between families. Most other studies making use of the phylogenetic data we use arbitrarily set the family-genera and genera-species distances to two- and one-thirds of the mean family-family distance, and therefore have not demonstrated this (e.g. Hardy & Senterre 2007; Hardy 2008). Other research, however, has suggested that a sufficiently large proportion of generalist herbivorous pests are found in rainforests to make any contribution of these effects to coexistence negligible (Novotny et al. 2002b).

Two difficulties remain: that neither model of neighbourhood interactions clearly fits the data better, and that the majority of effects identified reduce mortality rates. The first may be explained by a link between the effects, suggested by the fact that many of the same species show significant effects of the same sign in each case. This link would occur if species’ size (dbh) ranges differ, so that conspecific neighbours also happen to be large, for example. However, when plotted (Fig. 5.6), no obvious relationships exist between the dbh ranges of species and their phylogenetic or neighbour-diameter effects. This suggests that the effects are both genuine and distinct, or that they describe some other, unidentified process. It is certainly true that the implication that many species benefit
from high densities of related species in their neighbourhoods directly contradicts the Janzen-Connell hypothesis, and would actively inhibit the coexistence of unrelated species. This is not in itself evidence against beneficial interactions, however, and it has previously been suggested that some species in BCI do indeed benefit from having conspecific neighbours (Leigh et al. 2004). Mechanisms supporting this could include soil-microbe facilitation (Reynolds et al. 2003; Silvertown 2004), although these are also likely to increase competition in some cases (Connell & Slatyer 1977).

![Figure 5.6: Diameter (dbh) ranges for species with significant phylogenetic and neighbour-diameter effects. Positive effects (that increase mortality) are shown in black and negative effects in red. 95% dbh ranges are shown.](image)

Apparently beneficial relationships between related or large neighbours and mortality may also be generated by strong environmental preferences in some species. We do not include environmental terms in our models because of the difficulties of determining which environmental factors and data to include in models, but it has been found by others that a substantial number of species in BCI show evidence of environmental niche differentiation (Condit et al. 2000; Harms et al. 2001; John et al. 2007). It has also been noted that this can give spurious positive relationships between survival and the density of conspecific, or related, individuals, particularly where it is the dominant process (Leigh et al. 2004; Swenson et al. 2011). When we compared the spatial distribution of species with positive and negative phylogenetic and neighbour-diameter effects, we find little evidence of any systematic differences, with some species in each case appearing highly aggregated and others highly dispersed. We therefore summarised the aggregation of each species using the Clark and Evans aggregation index (Clark & Evans 1954), and the resulting ranges of values for species with positive and negative effects are shown in Figure 5.7. Species with negative coefficients that benefit from the presence of related or large neighbours show a wider range of aggregations but are on
average more aggregated than those with positive coefficients. This may indicate a slightly greater degree of environmental specialisation in species with negative coefficients, but the difference between the groups is small and unconvincing. However, it would be necessary to control for aggregation produced by different dispersal mechanisms before drawing firm conclusions from this finding, and so we cannot yet attribute the beneficial effects we find to environmental niche differentiation.

![Graph showing the range of Clark & Evans aggregation index values produced by species with significant phylogenetic and neighbour-diameter effects. Separate ranges are given for species with positive and negative coefficients.](image)

**Figure 5.7: The range of Clark & Evans aggregation index values produced by species with significant phylogenetic and neighbour-diameter effects.**

Separate ranges are given for species with positive and negative coefficients.

Although our findings do not suggest widespread Janzen-Connell effects, it is not necessarily the case that none exist. Instead, it is possible that they take a different form or operate on different spatial scales than those we consider. Depending on the principal agents of mortality, the effect may be highly localised, being driven by, for example, soil pathogens or sedentary larvae (Packer & Clay 2000; Novotny et al. 2002a), or may not be distance-dependent at all, as a meta-analysis has suggested (Hyatt et al. 2003). It certainly is not constrained to act in circular neighbourhoods of the kind we consider, with fixed (small) size and no height structuring. Pests and herbivores may operate at particular level (e.g. canopy), and on large, even landscape, scales. This would make the species composition across communities more important than small-scale spatial structure, and the Janzen-Connell effect more relevant to overall $\alpha$-diversity (Condit et al. 1992; Carson et al. 2008). Evidence is sparse: Uriarte et al. (2004) found that the average effective neighbourhood size in Luquillo (for survival data) was 11.85m, ranging between species from 4.2m to 14m, but Condit et al. (1992) found negative conspecific density effects, where present, almost always occurred within 5m.
Furthermore, the effect could be driven by periodic outbreaks of pests that may or may not occur during any particular census interval, making it temporally rather than spatially variable (Chesson & Warner 1981; Carson et al. 2008). This would also result from effects driven by disturbance events such as drought or windstorms (Uriarte et al. 2004). It has previously been noted that strong artefacts of this kind can affect mortality rates from one census interval (Sheil & May 1996). We find some evidence that species that benefit from related neighbours are those which suffered highest mortality during the previous (1982-1985) census interval, when an unusually strong drought occurred in BCI (Leigh et al. 2004). It has previously been noted that species which had suffered high mortality in the drought had unusually low mortality during the next census interval, while those that were less affected by the drought suffered far higher mortality after it (Leigh et al. 2004). Finally, it is possible that Janzen-Connell effects are strongest during recruitment, as was concluded by, for example, Wills et al. (1997), who found that the great majority of species suffered from it at this stage.

It may of course be the case that we do not detect Janzen-Connell effects because they are minor or absent. It has often been found that they occur only in a minority of cases, perhaps only in rare species for which data is scarce, and are not strong or widespread enough to explain tropical diversity (Lambers et al. 2002; Hyatt et al. 2003; Comita et al. 2010). This may be partly attributable to an analogous but opposing process occurring in other trophic levels, with pests of common species becoming so numerous that they themselves are heavily predated upon and diminished in number (Visser et al. 2011). We can, however, conclude that Janzen-Connell effects, where present, are likely to depend upon phylogenetic relatedness rather than simple species identity, and that they are not strong and ubiquitous in BCI in the forms that we model here.
5.5 Appendix 5.1: Supporting plots

Figure A5.1: Test data likelihood values for the three most successful models:

1. With a neighbour diameter at breast height term; 2. With a neighbour phylogenetic relatedness term; and 3. With both these terms. All other models gave substantially lower likelihood values.

Figure A5.2: The form of the relationship between tree diameter at breast height (dbh) and mortality, for a dbh range of 1 – 50 cm.

The relationship is expressed by:

\[ k_1 = a \cdot \log(dbh) + b \cdot dbh \]

With posterior hierarchical mean values for the parameters \(a\) and \(b\) used here.
CHAPTER 6 - Discussion
The aims of this thesis are to contribute to the available methodology for inferring underlying processes from the observed structure of ecological communities, and to extend empirical understanding of the dynamics of plant communities by testing hypotheses derived from theory in both simulated and natural communities. These aims are important because they may help to resolve the long-running debate over the mechanisms responsible for species coexistence in diverse communities such as tropical rainforests. As the most diverse ecosystems on Earth (Whitmore 1998; Primack & Corlett 2005), the conservation of tropical rainforests is a global priority that depends upon accurate knowledge of underlying ecology. The premise of this research is that analysis of spatial structure can substantially extend this knowledge.

As discussed earlier, theories about coexistence mechanisms have tended to be tested and compared on the basis of non-spatial first-order community characteristics such as species abundance distributions (SADs). In Chapter 2, we performed the first systematic comparison of first- and second-order signals of the five main theorised mechanisms of species coexistence. We demonstrated that SADs are unable to distinguish these processes from one another, with stochastic variation within models confounding any systematic variation between them. These findings are consistent with many earlier studies that found that the SAD was unable to distinguish particular processes and with its failure to resolve the debate over mechanisms of coexistence (Chave et al. 2002; Purves & Pacala 2006; Etienne & Alonso 2007; McGill et al. 2007; Alonso et al. 2008). Our results also cast further doubt on the assumption that general patterns of this kind are the result of fundamental ecological processes (Andrewartha 1958; Hajnal 1958; Prigogine 1980; Cramer 1993; Warren et al. 2011). Although we suggest that the SAD may contain some information about underlying processes, and others have found evidence that external factors such as disturbance may be discerned in it (Mouillot & Lepretre 2000; Hurlbert 2004; Thibault et al. 2004), it appears difficult if not impossible to reliably extract this information.

Spatial structure, in contrast, is found to be highly informative in Chapter 2. We developed a community-scale summary statistic of the spatial overlaps of all pairs of abundant species in a community, the cross-pair overlap distribution (xPOD), and show that it can reliably distinguish all but two of the modelled processes (neutrality and heteromyopia). Its ability to do so is unaffected by random variation within models, and is particularly great between neutral and niche models. The consequences of each of the modelled processes for spatial structure have not been widely considered previously, but it is recognised that the large-scale spatial predictions of neutral and niche theories differ, and this has formed the basis for empirical tests of neutral theory using measures of β-diversity (e.g. Terborgh et al. 1996; Chave & Leigh 2002; Condit et al. 2002; Dornelas et al. 2006). Although these tests operate at greater (cross-site) scales than our study, the differences they detect may be seen as extensions of the small-scale signals we find, in which niche-structured communities show a greater and systematic turnover of species with environmental variation. At larger scales, as more environmental variation is included, differences in community composition between neutral and niche models are therefore predictable from our results. The small-scale intraspecific aggregation produced by niche differentiation has also been previously noted (e.g. Hamill & Wright 1986; Plotkin et al. 2000; Zillio & Condit 2007; Murrell 2010).

The spatial signal we detect of the Janzen-Connell effect, in which species are found to be uniformly dispersed and mingled, matches both the original observations that led to the development of the theory (Janzen 1970; Connell 1970), and its implication that a greater proportion of heterospecific neighbours should be observed where the effect operates. This finding has therefore been anticipated in a broad sense (Clark & Clark 1984; Schupp 1992; Batista & Douglas 1998). The lack of any clear difference between the neutral and heteromyopia results in Chapter 2 was not predicted, and may be attributable to a decreased sensitivity to conspecific neighbours at small scales under heteromyopia that is balanced by an increased sensitivity at larger scales.
This highlights one limitation of the xPOD: that it averages across behaviour occurring within the scale over which it is defined. A solution to this is to form a ‘radius-weighted xPOD’ by weighting the constituent scale-dependent values that make up the xPOD; these are the areas under the cross-pair correlation functions at defined radius ranges, and are summed unweighted to form the xPOD used here (see Chapter 2). We have begun work on this modification (Studeny et al. 2011), and initial results suggest that the heteromyopia model does indeed result in opposing signals at small and medium spatial scales, and that it is consequently distinguishable from neutrality using the weighted xPOD. Similarly, the use of spatial statistics that operate at different scales would be expected to distinguish these processes, and we confirm this in Chapter 4.

The results of Chapter 2 are nevertheless limited in scope as their applicability depends upon the design of the models used. These models are intended only to describe the basic form of each process acting in isolation, and so we do not expect the same signals, or even differences between signals, to be present in empirical data. Specifically, variation in both the strength and form of the processes might be expected, perhaps even between species in a single community (Condit et al. 2000; Harms et al. 2001; Leigh et al. 2004). This is especially true of the niche, Janzen-Connell and heteromyopia effects, for which too little empirical evidence is available to constrain model parameters. The spatial signals of the niche model that we identify, for example, are dependent upon the range and scale of environmental variation, and also upon species’ sensitivity to these. We are unable to explore the effects of variations in these factors due to time constraints, but this would certainly justify further research, especially where this focuses on interspecific variation in the effects or on combinations of different effects.

Nevertheless, the spatial signals that we identify in Chapter 2 can confidently be regarded as characteristic of the modelled processes. Niche differentiation must result in the separation of species along environmental gradients, Janzen-Connell effects must result in increased spatial mixing of species, and lottery effects must result in considerable variation and spread between species. Furthermore, we have shown in the second Appendix to Chapter 2 that the large differences between neutral and niche models are robust to variation in dispersal kernels, and so far more likely to be detectable in real-world communities.

The neutral and niche models are particularly interesting because they generate contrasting and testable predictions for the effects of environmental heterogeneity on spatial structure. In a niche-structured community, increasing environmental heterogeneity is expected to increase the spread of spatial overlaps between pairs of species, while this is not expected in a neutral community. We tested these predictions in Chapter 3, and found that spatial patterns across 12 rainforest plots from around the world were not consistent with a fully neutral explanation of community structure, and instead matched the above prediction of niche theory. We concluded that the importance of niche differentiation to community structure was greater in more heterogeneous environments.

There have been few previous comparative studies of spatial structure in rainforest plots (Condit et al. 2000; Wills et al. 2006) and even fewer that have addressed the predictions of neutral and niche theories (John et al. 2007; Swenson et al. 2011). Such studies have generally found some evidence for niche differentiation, but its importance relative to neutral processes has not been confirmed. The same is true of the more numerous investigations of spatial structure in individual plots (Harms et al. 2001; Debski et al. 2002; Potts et al. 2004; Seidler & Plotkin 2006; John et al. 2007; Shen et al. 2009; Horvat et al. 2010; Kraft & Ackerley 2010; Hu et al. 2011; Lin et al. 2011; Zhang et al. 2011). Our findings are largely consistent with these, but extend previous studies in providing systematic and large-scale evidence of niche differentiation that increases in strength with environmental heterogeneity.

The work detailed in Chapter 3 was not intended to investigate the role of niche or neutral processes in supporting the coexistence of species, and we do not draw any direct conclusions about this.
While it is notable that we find no evidence that niche differentiation promotes increased diversity, with heterogeneous plots being no more diverse than homogeneous plots, substantial biogeographical differences between the plots mean that this cannot be interpreted as evidence against coexistence through niche differentiation. Nevertheless, coexistence may also be the result of niche differentiation on other axes or scales (Grubb 1977; Schulman et al. 2004; Comita & Engelbrecht 2009; Kanagaraj et al. 2011), neutrality within broad environmental niches, Janzen-Connell, heteromypenia or other effects, or some combination of these. We do not confirm or refute any of these, but the great majority of species pairs in every plot do have near-random overlaps indicated by values on the xPOD near to zero. To this extent our findings are consistent with studies that suggest that a range of neutral and niche processes support coexistence (Burslem et al. 2001; Wright 2002; Tuomisto et al. 2003; Hardy & Sonké 2004; Tilman 2004; Hubbell 2005; Hubbell 2006; Adler et al. 2007). In particular, Chave (2004) suggests that there is scope for neutrality within habitats, and that this would explain the difficulty of identifying habitat specialisation from SADs. Cadotte (2007) similarly argues that even strongly niche-differentiated communities may have multiple equivalent species within niches. This is reminiscent of Hubbell’s (2001; 2003) suggestion that, while differences between species do occur, they may not be the principal mechanism responsible for observed distributions or abundances.

We can conclude from Chapter 3 that spatial statistics such as the xPOD are able to distinguish the signals of ecological processes operating in real communities. Furthermore, spatial signals of niche differentiation are identified in Chapter 3 despite great and ecologically relevant differences between the plots in terms of latitude, temperature, rainfall, disturbance regimes and diversity. This is an important confirmation of the findings of Chapter 2, and justifies further investigation of the ability of spatial statistics to reliably distinguish mechanisms of coexistence.

In Chapter 4, we compared the ability of a range of spatial summary statistics to distinguish the processes modelled in Chapter 2. We found that all of the statistics we considered were able to distinguish some of the processes but that some were more successful than others. Measures expressed at the community level were most useful because stochastic variation meant that individual and species-level results were too variable to provide a clear signal. More interestingly, the most informative measures used information gathered at the individual level and local spatial scale, with the identities of nearest neighbours appearing especially informative. This confirms previous findings that distances to or identities of nearest neighbours are indicative of ecological processes (Pielou 1962; Hubbell & Foster 1984; Getzin et al. 2006; Lieberman & Lieberman 2007).

This comparison also enabled us to build a complete picture of the spatial signals of each of the modelled processes. These signals were strong and robust to random variation, and the potential for their empirical detection appears good. This suggests that testable predictions about spatial structure may be a valuable tool for investigations of coexistence mechanisms (Amarasekare 2003). The Janzen-Connell model produced the most spatially uniform communities, with high levels of interspecific scattering and intraspecific exposure. At the other extreme, the niche model produced communities with low levels of scattering and exposure and high rates of species turnover with increasing spatial scale. Systematic investigations of the spatial signals of different processes have been rare, and so other relevant findings tend to come from studies of one particular process. As noted above, patterns of β-diversity have been shown to differ between neutral and niche-structured communities, and this has been used in empirical tests of neutral theory (Legendre et al. 2005). Spatial structure has also been used to try and identify Janzen-Connell effects on several occasions (e.g. Sterner et al. 1986; Nathan & Casagrandi 2004; Wiegand et al. 2007a; Bagchi et al. 2011).

Perhaps the most useful function of the comparison in Chapter 4 is that it allows us to propose a highly informative combination of statistics for the empirical inference of underlying process. Our
results suggest that this combination (of the community-average values of the mingling index, proportion of conspecific neighbours and the standard deviation of the xPOD) can robustly separate all modelled processes by their spatial signals. However, we do not have enough simulated data to describe a complete multivariate distribution of these statistics under each model, and neither do we apply this combined measure to empirical data. These are obvious next steps (on which work has begun), and might eventually allow for probability-based inference of underlying processes from spatial patterns. Identification of methods of this kind was one important objective of this thesis.

A notable limitation of the work presented here is that combinations of different processes are not considered. Evidence from studies of single or small numbers of species suggests that such combinations are in fact very likely to occur. Specifically, both niche differentiation and Janzen-Connell effects have been found in many communities, and appear to vary in strength between species (see above and e.g. He et al. 1997; Wiegand et al. 2007a). A simulation-based exploration of the results of such combinations is likely to be prohibitively large and complex, and so opportunities for empirically constraining parameters are valuable. This is perhaps most difficult for niche differentiation, where the variables on which differentiation occurs may themselves be unknown. The Janzen-Connell effect, in contrast, can take one of a limited number of forms and is dependent upon measurable densities of particular species. The spatial scale of the effect, though, is uncertain (Schupp 1992; Carson et al. 2008). Knowledge of this is crucial not only for defining the spatial signals that the effect generates, but for assessing its potential role in maintaining species richness. If it operates at very small scales as originally proposed (Janzen 1970; Connell 1970) spatial signals will be strong and of the form identified in Chapter 2. If it operates at large scales, however, then it may be more accurately described by the heteromyopia model, in which conspecifics influence one another over larger distances (Amarasekare 2003).

In Chapter 5 we investigated the Janzen-Connell effect using empirical data from a single rainforest plot. We used hierarchical Bayesian modelling to try and identify evidence for the existence and form of the effect at small spatial (and temporal) scales. We find some evidence for interactions between species that vary in strength with phylogenetic relatedness, but both the hierarchical mean and majority of specific effects are directly opposed to that expected, with high densities of related neighbours appearing to reduce mortality rates. In addition, the fit of the model to data is virtually indistinguishable from a model in which mortality varies with the size of neighbouring individuals. This effect is also positive in the majority of cases, so that mortality rates decline as neighbour size increases.

We considered several explanations for the unexpected form of these effects, and for the fact that neither clearly described the data better. These included environmental specialisation, relationships between species size and relatedness effects, and genuinely beneficial effects of high related- or large-neighbour density. We found no evidence for these, however, and so these suggestions remain speculative and are an obvious subject for further research. Nevertheless, we were able to conclude that true, small-scale Janzen-Connell effects were not widespread among plants with a diameter greater than 1cm, and were likely to span a number of related species (up to the level of confamilials) where they did occur. These tentative conclusions are supported by several studies and meta-studies that find that Janzen-Connell effects are rare or weak for established saplings ≥ 1 cm dbh (Connell et al. 1984; Condit et al. 1992; Hyatt et al. 2003; Paine et al. 2008; Svenning et al. 2008; Bagchi et al. 2011). They are also consistent with the known tendency of pests and pathogens to specialise at the level of genus or family (e.g. Barone 1998; Novotny et al. 2002a) and for observed Janzen-Connell effects, where found, to vary in intensity with phylogenetic relatedness (e.g. Boyden et al. 2008).

The lack of Janzen-Connell effects is also relevant to our earlier findings because it suggests that their spatial signals are unlikely to be strong in real-world communities and therefore are unlikely to
confound the opposing signals of niche differentiation. They further imply that the effects of specialised pests and pathogens, being largely undetectable at small spatial scales, may instead operate over larger scales, so being more accurately described by our heteromyopia model (Murrell & Law 2003; Barot 2004). In Chapter 4 we developed a robust method for distinguishing the spatial signals of heteromyopia and neutrality, and this is therefore another obvious subject for further research. In particular, a method of separating the combined effects of niche differentiation and heteromyopia from their resulting spatial patterns would be valuable. A careful consideration of changes in structure with spatial scale, perhaps using the radius-weighted xPOD mentioned above, might enable this.

This thesis has demonstrated that spatial statistics have great and previously unrecognised potential for the inference of underlying process in diverse forest communities from resulting spatial pattern. We have shown that ecological processes thought to be responsible for maintenance of the high diversity of tropical tree communities have inherent and strong consequences for spatial structure, and we have described these consequences in full on the basis of simulated data. We have also shown that a range of spatial statistics are sensitive to such signals and developed a combination of new and existing summaries of spatial structure that provide a basis for inference about underlying process using derivable likelihoods. In comparing spatial structure in 12 different large-scale rainforest plots from around the world, we demonstrated that we were able to distinguish the signal of niche differentiation in empirical data despite substantial ecologically-relevant differences and the action of many other undefined processes. We further investigated the occurrence of density-dependent mortality of stems ≥ 1 cm dbh in one large rainforest plot and found little evidence for it on the spatial and temporal scales that we consider, concluding that its spatial signature is likely to be weak. These findings have led us to suggest a targeted programme of further research to extend the application of spatial statistics in ecology and to further illuminate the processes responsible for species coexistence.

The ultimate aim of this work is to contribute to the effective management and conservation of tropical rainforests, some of the most valuable environmental resources on Earth. Discerning which of several possible processes is dominant in structuring these communities is essential to this aim, and the use of spatial statistics clearly holds great promise for this. The further development of statistical methodology together with the ongoing collection of spatially-explicit data is therefore an important supplement to the practical protection of highly diverse forest communities.
CHAPTER 7 - References


Ayala, F.J. 1972, "Competition between species: the diversity of environments in which most organisms live permits the coexistence of many species, even when they compete for the same resources", American Scientist, vol. 60, no. 3, pp. 348-357.


Chang, H.Y. 2010, *Woody Species's Composition, Structure, Distribution and Characteristics at 4 ha Plot of Low Altitude Evergreen Broad-Leaved Forest in Lienhuachih Area, Central Taiwan*, Providence University, Taiwan.

Chase, J.M. & Myers, J.A. 2011, "Disentangling the importance of ecological niches from stochastic processes across scales", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1576, pp. 2351-2363.


Clark, P.J. & Evans, F.C. 1954, "Distance to nearest neighbor as a measure of spatial relationships in populations", *Ecology*, vol. 35, no. 4, pp. 445-453.


Hutchinson, G.E. 1959, "Homage to Santa Rosalia or why are there so many kinds of animals?", The American Naturalist, vol. 93, no. 870, pp. 145-159.


Jaccard, P. 1912, The distribution of the flora in the alpine zone 1, Blackwell Publishing Ltd., Oxford.


Murrell, D. 2010, "When does local spatial structure hinder competitive coexistence and reverse competitive hierarchies?", *Ecology*, vol. 91, no. 6, pp. 1605-1616.


Prigogine, I. 1980, From being to becoming: time and complexity in the physical sciences, 1st edn, WH Freeman and Company, San Francisco.


Rajala, T. & Illian, J.B. in press, "Graph-based description of mingling and segregation in multitype spatial point patterns", Ecological and Environmental Statistics.


Whittaker, R.H. 1962, "Classification of natural communities", *Botanical Review*, vol. 28, no. 1, Classification of Natural Communities, pp. 1-239.


