Assessing the interaction between landscape characteristics and biodiversity

Charlotte Emily Carter

MEnvSci (Hons.)

Submitted for the degree of

Doctor of Philosophy

School of Life Sciences,

University of Warwick

August 2014

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would firstly like to thank my supervisor Mr Andrew Mead for allowing me the opportunity to do a PhD and for his support, guidance and patience throughout. You have been extremely helpful, and I really appreciate all the time you have made for me, especially this past year. Thank you also to Dr Dave Skirvin for your supervision at the start of my PhD and to the Warwick Chancellors Scholarship for funding my research.

I would also like to thank my external supervisor Dr Philip Fermor for his ecological expertise and for setting up the CASE partnership with Middlemarch, for which I am extremely grateful. Thank you to everyone at Middlemarch for making these past four years so enjoyable. I have learnt so much from you all, in particular from Dr Susanna Saari, you have been a wonderful line manager. Enormous thanks to Jarek Sobanski for all your help with python scripting, and for showing me lots of neat things you can do with ArcGIS.

Thank you to Chris Talbot for supplying the Warwickshire Habitat Biodiversity Audit and for sharing your knowledge of the data base. Thank you also to David Lowe and Louise Sherwell from Warwick County Council and Marc Botham from the UKBMS for the provision of butterfly records.

I would also like to thank Ellie, Lauren and Ellen for making my experience at Warwick so enjoyable and also to Badger, Fox and Stag for all the fun times we've had.

A massive thank you goes to Dr Greg 'WolfMAN' Amos for everything you have done to help me, from keeping me company in the field to reading every page at the end. I would not have been able to get through this without you.

Thank you to Mum, Dad, Rachel and Nan for all of your help, love and support. Thank you for bringing me food to the office and for helping me in the field.

Declaration

In accordance with the University of Warwick regulations for the degree of Doctor of Philosophy, I certify that this thesis has been written solely by me. The work contained in this thesis is my own unless otherwise stated. No aspect of this work has been submitted to any other institution for any other degree of award.

Summary

Severe declines in biodiversity have been attributed to anthropogenic changes in the composition and structure of our landscapes. Predicting the impact of landscape change on biodiversity is essential to halt further declines. In this thesis butterflies were used as indicators of biodiversity, and spatial assessments of butterflies were summarised at 1 km scale across Warwickshire to assess whether landscape characteristics can be used as surrogate measures of butterfly distribution and community measurements.

When determining the optimal scale (grain size) for capturing landscape patterns, a grain size of 25 m was found most appropriate for maximising landscape discrimination and detecting landscape patterns which occur within the perceptual range of butterfly species. Utilising a grain size of 25 m landscape metrics measuring the composition, connectivity and structure of the 1 km landscapes, were extracted from the Land Cover Map 2000 (LCM) and the Warwickshire Phase 1 Habitat map (PH1). Logistic regression analysis based on landscape metrics created predictive models of butterfly distribution for all species and species grouped by their ecological attributes (EAGs). Model performance was improved when the landscape metrics were considered in a combined landscape model, and different combinations of landscape parameters were important for the EAGs.

Models derived from the PH1 were most accurate in predicting observed presence-absence and were successfully transferred when tested using temporally independent data. The models were also successfully transferred to collected butterfly data which was spatially and temporally independent. This data was also collected alongside information on the local habitat such as vegetation composition. Probability of butterfly occurrence derived from the presence-absence models was successfully related to butterfly community characteristics and measures of local habitat quality.

To conclude developed models provide indications of habitat suitability, which together with successful transfer demonstrates their potential for identifying biodiversity hotspots and facilitating targeted conservation efforts.

Abbreviations

ANOSIM	Analysis of Similarities
ANOVA	Analysis of Variance
BARS	Biodiversity Action Reporting System
BC	Butterfly Conservation
BNM	Butterflies for the New Millennium
BRC	Biological Records Centre
CBD	Convention on Biological Diversity
CEH	Centre for Ecology and Hydrology
Defra	Department for Environment and Rural Affairs
DLC	Diversity of Land Covers
EAG	Ecological Attribute Group
EOF	Environmental Observation Framework
EU	European Union
GIS	Geographic Information Systems
Ha	Hectares
HBA	Habitat Biodiversity Audit
JNCC	Joint Nature Conservation Committee
LCM	Land Cover Map (1990, 2000, 2007)
MPS	Mean Patch Size
NBN	National Biodiversity Network
NCA	National Character Area
NGO	Non-Governmental Organisation
NLC	Number of Land Cover Classes
OS	Ordnance Survey
PCA	Principal Component Analysis
PH1	Phase 1 Habitat Map (2000, 2010)
SSSI	Site of Special Scientific Interest
TA	Total Area
UK	United Kingdom
UK BAP	UK Biodiversity Action Plan
UK BMS	UK Butterfly Monitoring Scheme
WBRC	Warwick Biological Records Centre
WCC	Warwick County Council

Compositional Metrics

LSIDI	Landscape Simpsons Diversity Index
NLAND	Number of Land Cover Classes

Connectivity Metrics

IIC	Integral Index of Connectivity
varIIC	Absolute variation in the connectivity component of the Integral Index of Connectivity after patch removal
varIICconn	Absolute variation in the connectivity component of the Integral Index of Connectivity after patch removal
varIICflux	Absolute variation in the flux component of the Integral Index of Connectivity after patch removal

Structural Metrics

AREA_AM	Area-Weighted Mean Patch Size
AREA_MN	Mean Patch Size
AREA_RA	Range in Patch Size
CIRCLE_AM	Area-weighted Mean Related Circumscribing Circle
CIRCLE_MN	Mean Related Circumscribing Circle
CIRCLE_RA	Range in Related Circumscribing Circle
COHESION	Patch Cohesion Index
CONNECT	Connectance Index
CONTAG	Contagion index
CONTIG_AM	Area-weighted Mean Contiguity Index
CONTIG_MN	Mean Contiguity Index
CONTIG_RA	Range in Contiguity Index
CONTIG_SD	Contiguity Index Standard Deviation
CWED	Contrast weighted edge density
ECON_AM	Area-weighted Mean Edge Contrast Index
ECON_CV	Edge Contrast Index coefficient of variation
ECON_MN	Mean Edge Contrast Index
ECON_RA	Range in Edge Contrast Index
ECON_SD	Edge Contrast Standard deviation
ENN_AM	Area-weighted Mean Euclidean Nearest-Neighbour Distance
ENN_CV	Euclidean Nearest-Neighbour Distance co-efficient of variation
ENN_MN	Mean Euclidean Nearest-Neighbour Distance
ENN_SD	Euclidean Nearest-Neighbour Distance standard deviation
FRAC_AM	Area-weighted Mean Fractal Dimension Index
FRAC_CV	Fractal Dimension Coefficient of Variation

GYRATE_AM	Area-weighted Mean Radius of Gyration
GYRATE_CV	Radius of Gyration Coefficient of Variation
GYRATE_MN	Mean Radius of Gyration
GYRATE_RA	Range in Radius of Gyration
IJI	Interspersion and juxtaposition index
LSI	Landscape Shape Index
MESH	Effective Mesh Size
PRD	Patch richness density
PROX_AM	Area-weighted Mean Proximity Index
PROX_CV	Proximity Index co-efficient of variation
PROX_MD	Median Proximity Index
PROX_MN	Mean Proximity Index
SHAPE_AM	Area-weighted Mean Shape Index
SHAPE_CV	Shape Index Coefficient of Variation
SHAPE_MN	Mean Shape Index
SHAPE_SD	Shape Index standard deviation
SIDI	Simpson’s Diversity Index
SIMI_AM	Area-weighted Mean Similarity Index
SIMI_CV	Similarity Index co-efficient of variation
SIMI_MD	Similarity Index co-efficient of variation
SIMI_MN	Mean Similarity Index
SIMI_RA	Range in Similarity Index

Chapter 1: Introduction

1.1 The biodiversity crisis

The world's population is growing at an exponential rate and with it the demand for resources and land (EEA-FOEN, 2011; Gil-Tena, *et al.*, 2012). The subsequent landscape change due to urban sprawl, intensification of agriculture, and the development of infrastructure (EEA-FOEN, 2011), is considered to be one of the main causes of habitat destruction and fragmentation worldwide and is happening at a faster rate than ever before (Bulte, *et al.*, 2005). The detrimental impact of increasing rates of landscape change and pollution on biodiversity are further exacerbated by the overexploitation of natural resources, introductions of invasive species and climate change (Watson and Albon, 2011).

This ever-increasing destruction of the world's ecosystems is exceeding the tolerance limits of natural change (Christensen, *et al.*, 1996) and the subsequent loss of biodiversity is a major global concern. Nature has changed throughout history, and biodiversity is dynamic, but never before have ecosystems been under so much pressure (Chapin, *et al.*, 2000; Christensen, *et al.*, 1996). Pimm *et al.*, (2014) estimated that 21st Century species extinction rates are 1000 times higher than the background rate of extinction, the effects of which are detrimental to both ecosystem function and ecosystem services.

Biodiversity is essential for sustaining everyday life (Chapin, *et al.*, 2000; Ehrlich and Wilson, 1991), as it incorporates not only the diversity within and among species but also the diversity within and among ecosystems and their processes (Watson and Albon, 2011). More specifically Noss (1990) introduced a hierarchal concept for defining biodiversity which suggests that biodiversity is related to the composition, structure and function of an ecosystem across a multitude of scales, ranging from that of genetic structure to landscape patterns and processes (Failing and Gregory, 2003; Lindenmayer, *et al.*, 2000). It is this biodiversity across a multitude of scales which underpins the provision of ecosystems services which are pivotal for everyday life (Butchart, *et al.*, 2010).

1.2 Cause of crisis: Land cover change

During the 20th century the UK experienced severe declines and fragmentation of semi-natural habitats, including species-rich grassland and ancient woodland (Lawton, *et al.*, 2010). In particular, since the mid-1940s changes in agricultural practices have resulted in the decline of 73 % of 'priority habitats' designated under the UK Biodiversity Action Plan (UK BAP), including semi-natural heathland, chalk grasslands and lowland wet grasslands (Jackson, 2000; JNCC, 2010b; Lawton, *et al.*, 2010). The detrimental effects of habitat loss are also accompanied by a more recent deterioration in the quality of the remaining habitat patches due to changes in management practices such as the reduction of grazing and coppicing (Lawton, *et al.*, 2010).

1.2.1 Land cover change: Semi-natural grassland

Fuller (1987) estimated that 97 % of semi-natural grassland (acid, neutral and calcareous grasslands, purple moor grass and rush pasture) has been lost in England and Wales between 1930 and 1984. This has resulted in only 2 % of all grasslands being semi-natural in the UK (Bullock, *et al.*, 2011; Lawton, *et al.*, 2010). This extensive loss has been attributed to the conversion of semi-natural grassland into improved grassland or arable land due to agricultural intensification and mechanisation (Bullock, *et al.*, 2011). Further losses of 2-10 % per annum have been recorded for some parts of England during the 1980s and 1990s (Maddock, 2008). However, since the launch of the UK Biodiversity Action Plan (UK BAP) in 1995, over 3500 hectares (ha) of species–rich grassland (3 % of total coverage in England) has been created (Lawton, *et al.*, 2010), and 68 % of semi-natural grassland area is now protected within Sites of Special Scientific Interest (SSSI) (Bullock, *et al.*, 2011). Despite this, only 21 % of non-statutory semi-natural grasslands were recorded to be in a healthy state (favourable condition) in 2005 (Bullock, *et al.*, 2011). Degradation of non-statutory semi-natural grasslands has been linked, amongst other factors, to under grazing associated with inadequate management, resulting in the domination of rank vegetation and scrub (Bullock, *et al.*, 2011).

1.2.2 Land cover change: Broad-leaved woodland and hedgerows

In England broad-leaved woodland cover has increased by 23 % since 1945, and median woodland patch size has recently been estimated to be 3.9 ha (Lawton, *et al.*, 2010). However, this increase in cover has arisen from afforestation and secondary plantations and masks the destruction and fragmentation of ancient woodland which has occurred over several centuries, mostly due to the expansion of agriculture (Bailey, *et al.*, 2002; Quine and Watts, 2009; Rackham, 1986; Watts, 2006). In England loss of ancient woodland has been estimated at 7 % between 1930 and 1985 (Spencer and Kirby, 1992). Reductions in traditional management practices such as coppicing in the last 60 years has contributed towards the deterioration of remaining woodland fragments, as the resulting closure of the woodland canopy leads to a reduction in ground and field layer vegetation (Kirby, 2005; Lawton, *et al.*, 2010; Quine, *et al.*, 2011). Agricultural intensification since 1945 has also been accompanied with a severe reduction and deterioration of hedgerows (Croxton, *et al.*, 2005; Gillings and Fuller, 1998). Exact figures of loss are unknown; however net hedgerow loss between 1947 and 1985 has been estimated at 20 % (Evans, *et al.*, 2003). Barr *et al.*, (1993) estimated hedgerow length has reduced by an average of 7.8 % between 1985 –1990, due to both removal and neglect. Hedgerows have the potential to provide resources to support a variety of species predominately associated with woodlands and the woodland edge, in addition to facilitating daily dispersal events and long distance movement between connected fragments of semi-natural habitat (Hinsley and Bellamy, 2000).

1.3 Land cover determines landscape pattern and biodiversity dynamics

Landscape pattern (i.e. the configuration and composition of the landscape) is dynamic and driven by a multitude of abiotic factors including climate, geology, landform, and natural disturbance (Turner, 2005). A key determinant of landscape pattern is land-use change which, often driven by anthropogenic pressure, directly determines landscape composition (what is in the landscape and its quality) and its configuration (spatial arrangement and connectivity) (Bergerot, *et al.*, 2012; Turner, 2005).

The composition and configuration of habitats within a landscape determines landscape permeability (resistance of the landscape to species movement) for

specific species. Landscape composition in conjunction with abiotic factors directly influences the quality of habitat patches and the surrounding matrix (Dauber, *et al.*, 2003; Rossi and van Halder, 2010). The spatial arrangement of habitat patches within the landscape combined with the behavioural responses of individual species to landscape structure determines functional connectivity and, in turn, landscape connectivity (Baguette and Van Dyck, 2007; Ockinger and Van Dyck, 2012; Tischendorf and Fahrig, 2000). The impact of landscape structure on species movement depends upon the scale to which that species perceives its environment and the extent to which it can disperse (Ockinger and Van Dyck, 2012; Tews, *et al.*, 2004). Permeable landscapes therefore have high levels of functional connectivity, comprising well-connected habitats that have a low resistance to species movement (Goodwin and Fahrig, 2002; Kindlmann and Burel, 2008).

Within heterogeneous landscapes connectivity determines population dynamics and stability of meta-populations due to its influence on dispersal and migration (Goodwin, 2003; Kindlmann and Burel, 2008; Rossi and van Halder, 2010; Saura, *et al.*, 2008). Meta-populations, coined by Levins (1969) and later developed by Hanski (1994) exist as a number of connected sub-populations or colonies divided into 'sources' (donor patches) and 'sinks' (receiver patches). The persistence of the meta-population is dependent on a balance between colonisation and extinction between sources and sinks (Begon, *et al.*, 2006). Dynamics of individual colonies are determined by birth, death, immigration and emigration (Begon, *et al.*, 2006). Large-scale migration between colonies is essential for gene exchange and small-scale movement patterns determine local population survival (Ockinger and Van Dyck, 2012; Rossi and van Halder, 2010; Saura, *et al.*, 2008).

Landscape connectivity is heavily influenced by changes to landscape structure such as through the process of habitat fragmentation which involves not only net habitat loss but the segregation of a large continuous habitat into small isolated patches (Fahrig, 2003; Tscharntke, *et al.*, 2002). It is well established that species richness is a function of habitat area (Kennedy and Southwood, 1984) and as such reductions in habitat area, coupled with reduced resource availability are likely to be associated with reduced species richness. As an explanation for this species-area relationship, principles have been drawn from the Theory of Island Biogeography, proposed by MacArthur and Wilson (1967), considering habitat fragments analogous to 'islands'.

Within this theory, reductions in habitat area lead to an increase in species extinction rate due to reductions in available resources, and increased isolation of a habitat patch leads to a decrease in the immigration rate. Despite many critics of this analogy, this theory has been important in highlighting the effects of habitat size and isolation on species richness and population persistence (Turner, *et al.*, 2001).

Habitat fragmentation is considered to be one of the most important threats to biodiversity on a global scale, as it undermines habitat connectivity and disrupts the distribution and quality of habitats (Schindler, *et al.*, 2008). However, the magnitude of these effects is entirely species and scale specific, dictated by the degree of specialism and dispersal capabilities of a species and in turn the permeability of the surrounding landscape (Rossi and van Halder, 2010). In particular, the process of fragmentation is associated with increased exposure of the boundary of remaining habitat fragments (edge habitat), which has implications on both habitat quality and structure (Ewers and Didham, 2006; Haila, 2002). Edge habitats have been associated with unfavourable micro-climatic conditions for more specialist species, as well as increased predation and competition by more generalist species (Aurambout, *et al.*, 2005). Edge habitats can, however, provide favourable conditions, typically for more generalist species, through increased vegetation structural heterogeneity (Bergman, 2001; Dover and Settele, 2009). Furthermore, due to changes in the spatial pattern of habitat, habitat fragmentation can lead to an increase in landscape heterogeneity (Franklin, *et al.*, 2002). Heterogeneous landscapes characterised by diverse land cover types, have the capacity to support more species, and positive associations with biodiversity, particularly within agricultural landscapes, have been identified (Fischer and Lindenmayer, 2007; Schindler, *et al.*, 2008; Tscharntke, *et al.*, 2005). Habitat fragmentation can therefore be thought to have a mixture of both positive and negative effects.

1.4 A 'landscape scale' approach to conservation and research

Landscape composition and structure directly influences functional connectivity and permeability, and habitat fragmentation involves a multitude of inter-related processes that operate over a landscape scale (Bailey, 2007). As such, focusing conservation efforts on community assemblages across the landscape is considered vital for species preservation (Gil-Tena, *et al.*, 2012) due to it maintaining the complex interactions both between species and with the abiotic components of their surrounding environment (Noss, 1990).

Within the last decade, conservation efforts and scientific research has experienced a paradigm shift from focusing on the scale of the site or reserve to that of the landscape, considering the spatial configuration of semi-natural habitats (Dover and Settele, 2009). The theory of Island Biogeography (MacArthur and Wilson, 1967) and meta-population models (Hanski, 1994; Levins, 1969) has played a major part in driving conservation efforts globally, with the aim of addressing the negative impacts of habitat fragmentation at a landscape scale, particularly through the creation of wildlife corridors (Dover and Settele, 2009; Shreeve and Dennis, 2011). Many conservation strategies are formulated around the critical goal of enhancing landscape functional connectivity, through the provision of habitat networks to facilitate species dispersal (Gil-Tena, *et al.*, 2012; Vogt, *et al.*, 2009). Enhanced dispersal between sub-populations not only increases species density, but through gene flow increases both genetic and species diversity, allowing for better persistence of the meta-population (Ewers and Didham, 2006; Harrison and Bruna, 1999). This so-called 'rescue effect' can therefore limit local extinction of small isolated populations (Bailey, 2007).

As a signatory to the Convention on Biological Diversity (CBD), the UK government is committed to 'the Aichi targets' agreed in Nagoya in 2010 (Defra, 2011b; Defra, 2011a). An integrated landscape-scale approach is considered essential for achieving these targets and as such is central to recent environmental policy and conservation initiatives in the UK. Examples of such initiatives are:

- **Making Space for Nature: A review of England's Wildlife Sites and Ecological Network** – report led by Professor John Lawton highlighted that wildlife sites in England are fragmented and vulnerable. The report

recommended ‘reconnecting nature’ with ‘better, bigger, more and joined up’ priority habitats across the landscape (Lawton, *et al.*, 2010).

- **UK National Ecosystem Assessment (2011)** – assessment of the state of the UK ecosystem services revealed that many services are in decline and the underpinning habitats are fragmented. A move towards an integrated ecosystem approach to management is required (Watson and Albon, 2011).
- **The Natural Choice - Natural Environment White Paper (2011)** – the first natural environment white paper for 20 years outlined environmental policy for the next half century, shifting towards a landscape approach to the protection of the natural environment (Defra, 2011a).
- **ThinkBIG: How and why landscape-scale conservation benefits wildlife: people and the wider economy** – produced by Natural England on behalf of the England Biodiversity Group this report provided information for local authorities, communities and land managers to aid the implementation of a landscape approach to nature conservation (England Biodiversity Group, 2011).
- **Biodiversity 2020: A strategy for England’s wildlife and ecosystem services** – outlined the government’s strategy to implement international and EU commitments to halt the overall loss of biodiversity in England by 2020. The strategy builds upon environment policy in the Natural Environment White Paper, emphasising the need to establish a landscape-approach to improve ecological networks (Defra, 2011b).

1.4.1 Methods for measuring biodiversity at a landscape scale

Landscape scale assessment of biodiversity is essential for monitoring progress towards conservation and the achievement of the Aichi Targets. Since the Convention for Biodiversity at the Rio de Janeiro Earth Summit in 1992, the term biodiversity has rapidly evolved in use and meaning (Büchs, 2003). Biodiversity is a multifaceted concept, and it is this complexity which means that it would not only be too expensive but impossible to measure all entities of biodiversity (Lindenmayer and Likens, 2011; Vandewalle, *et al.*, 2010). Direct measurement can only be conducted on a small community subset (Bräuniger, *et al.*, 2010) and therefore the use of (biodiversity) indicators has been proposed as a method of representing biodiversity.

Indicator taxa or species are widely used to represent the compositional and functional aspects of biodiversity. Many indicator species are monitored at a national scale in the UK and across Europe including butterflies (*Lepidoptera*) and birds (*Aves*). Extensive biodiversity surveillance schemes are conducted across the UK by numerous programmes run by non-governmental organisations (NGOs), societies and research organisations and this data is collated by the UK Environmental Observation Framework (EOF) (UKEOF, 2014). Data inventories and national monitoring schemes are cost effective for evaluating species diversity over landscape scales, relying heavily on the contributing effort of a coordinated network of volunteers (citizen science) (Roy, 2012). The contribution of citizen science to recording schemes has been criticised due to bias in relation to spatial and temporal coverage (Roy, 2012). However, such schemes, particularly those included within the EOF, often involve standardised methods of data collection that can be repeated and transferred across the UK in order to facilitate the assessment of spatial and temporal trends in the distribution of indicator species in response to different arrangements of land cover (UKEOF, 2014).

The National Biodiversity Network (NBN) Gateway and the Biodiversity Action Reporting System (BARS) support the UKEOF by providing open access to biodiversity data (JNCC, 2014). The NBN integrates data from different sources including the Biological Records Centre (BRC) which collates data from national recording schemes and from individual local record centres, and provides a portal of access for use of the data by individuals and research organisations (NBN, 2011).

1.4.2 The UK Butterfly Monitoring Scheme

The UK Butterfly Monitoring Scheme (UK BMS) is one of the most comprehensive examples of national biodiversity monitoring conducted within the UK. The UK BMS launched in 2006, successfully combined the long established Butterfly Monitoring Scheme (BMS) (operating since 1976) and the more recent Butterfly Conservation (BC) transect project, into one single coherent scheme with over 1,000 sites surveyed across the UK per year (Fox, *et al.*, 2011). The UK BMS (and the two preceding schemes) use a standardised method of data collection based on that developed by Pollard (1977). This method involves conducting counts of butterfly numbers at sites along a fixed transect route under suitable weather conditions throughout the flight period (April to September) resulting in 26 transects per year.

Extensive monitoring of butterflies is conducted due to recognition of their intrinsic value and role as biodiversity indicators. As such, there is a wealth of data available on the distribution and abundance of butterfly species across the UK. Short term trends in butterfly populations can be assessed from these standardised annual counts.

1.5 Butterflies as indicators of biodiversity

One of the goals of the UK BMS is to produce official indicators of the broad state of biodiversity and the environment for use by the government (Fox, *et al.*, 2011). A range of multi-species butterfly indicators have been developed to assess the status of species of European importance and also to assess changes between habitat specialist species (confined to discrete localised habitat patches) and wider countryside species (generalists or habitats are widespread) in farmland and woodland habitats (Asher, *et al.*, 2001; Defra, 2012; Defra, 2013; Fox, *et al.*, 2011). These indicators are currently used to track progress towards the delivery of the Biodiversity 2020 strategy for England, which sets out commitments towards implementing international and EU targets (Defra, 2013).

Spatial and temporal trends in butterfly populations are considered to provide a good indication of change in other insects and in turn overall biodiversity as insects make up more than 50 % of terrestrial wildlife (Fox, *et al.*, 2011). Butterflies are recognised as strong biodiversity indicators both by the scientific and political communities (Asher, *et al.*, 2001; Dennis, 2009; Fox, *et al.*, 2011); several pioneering studies of meta-population ecology are based on the response patterns of butterflies (Dover and Settele, 2009; Kumar, *et al.*, 2009). The ecological attributes of butterfly species meet numerous criteria regarded as ‘essential’ for the successful application of biodiversity indicators; wide ecological breadth, responsiveness to environmental change, and well-studied life histories (Dennis, 2009; Heink and Kowarik, 2010; Hilty and Merenlender, 2000; Robinson, *et al.*, 2014; Rossi and van Halder, 2010).

There are 59 resident butterfly species in Britain, with three regular European migrants (clouded yellow *Colias croceus*, painted lady *Vanessa cardui* and red admiral *Vanessa atalanta*) (Asher, *et al.*, 2001). The 59 resident species of Britain can be grouped into five families; Hesperiidae (the skippers), Papilionidae (the

swallowtails), Pieridae (the whites), Lycaenidae (the hairstreaks, coppers and blues) and Nymphalidae (the nymphalids, fritillaries, browns and milkweeds), with the Duke of Burgundy classified within the family Lycanidae in accordance with Emmet & Heath (1990). The majority of butterfly species form colonies that exist in meta-populations (e.g. marsh fritillary *Euphydryas aurinia*), however some species are more wide ranging characterised by open population structures (e.g. orange tip *Anthocharis cardamines*) (Asher, *et al.*, 2001; Dover and Settele, 2009). Much variation exists between and within families in dispersal capabilities, which range from migratory (e.g. painted lady *Vanessa cardui*) to restricted (e.g. gatekeeper *Pyronia tithonus*) (Dover and Settele, 2009).

Certain species have a broad geographic range, with 16 out of the 59 resident species occurring throughout most of Britain, and generalist species occurring and breeding successfully within most terrestrial habitats (Asher, *et al.*, 2001; Dennis, 2009) (Figure 1.1a). This wide ecological breadth is considered essential for an indicator, in order to represent ecosystem health and biodiversity across a range of habitats (Heink and Kowarik, 2010). However the remaining species have a geographic range boundary marking the north western limits of these species in Europe (Figure 1.1b) (Asher, *et al.*, 2001). The range of each butterfly species is determined by the availability of suitable habitats and food plants which is in turn governed by topography, climate, geology, and land use (Asher, *et al.*, 2001; Dennis, 2010; Schweiger, *et al.*, 2006). Habitat specialist species, particularly at the edge of their geographic range (e.g. grizzled skipper *Pyrgus malvae*), are responsive to local environmental change due to their limited dispersal capabilities and sensitivity to micro-climate and local habitat quality (Asher, *et al.*, 2001; Fleishman, *et al.*, 2003; Robinson, *et al.*, 2014).

Butterflies have a well-developed proboscis with which they obtain liquid food, primarily nectar from flowers, tree sap, and juice from fruit (Dennis, 2010; Erhardt and Mevi-schutz, 2009). Butterfly species require complementary resources during their life cycle, including host plants for larvae, nectar plants for adults and sheltered, warm sites for resting and over-wintering (Dover and Settele, 2009; Ouin, *et al.*, 2004). As such species occupy several different vegetation types, and community structure is therefore directly related to vegetation composition and landscape characteristics (Dennis, 2010; Rossi and van Halder, 2010). Considering both

specialist and generalist species is therefore essential for capturing the relationships with wider biodiversity.

Butterflies as a group respond rapidly to disturbance events due to their short generation time (Kumar, *et al.*, 2009; Thomas, 2005). The life cycle of butterflies involves the four stages of metamorphosis; egg, larvae, pupae and imago. Location and availability of host plants is important for providing oviposition sites and proving a source of food for larva once hatched (Garcia-Barros and Fartmann, 2009). Larva feed extensively on the host plant and often on surrounding plants for approximately two weeks prior to pupation. Sheltered locations are important for protecting pupa, which remain attached to the leaf of a host plant, and depending on the species this stage may range from a few days to a few years, after which the adult butterfly (imago) emerges and remains in situ whilst the wings harden (Hoffman and Marktanner, 1995). Butterfly species are characterised by volitinism, with some species having two broods per year. Many species will hibernate as adults or overwinter as eggs or chrysalises (Munguira, *et al.*, 2009).

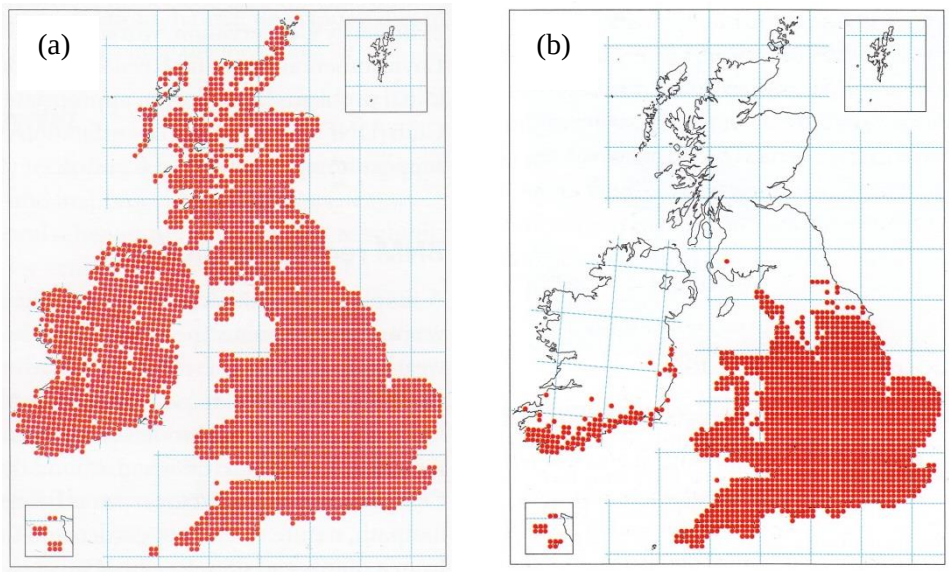


Figure 1.1: Contrasting distribution of two butterfly species across the UK (a) the widespread distribution of the green veined white (*Pieris napi*) and (b) north western limits in the distribution of the gatekeeper (*Pyronia tithonus*). Source: Asher *et al.*, (2001).

The standardised methodology of data collection primarily conducted by the UK BMS (see section 1.4.2) is transferable spatially and temporally and can be easily applied at the landscape scale. Furthermore, species can be easily identified within the field and their life histories are well studied (Asher, *et al.*, 2001; Fleishman, *et al.*, 2003; Robinson, *et al.*, 2014). All of these attributes make butterflies ideal species for monitoring temporal population trends in response to conservation action and policy initiatives (Fox, *et al.*, 2011).

1.5.1 Current threats to butterfly species

Butterflies are an important component of the ecosystem throughout their life cycle, providing prey for birds, bats and other insectivorous animals in addition to providing the ecosystem service of pollination during their adult stage (Erhardt and Mevi-schutz, 2009). Habitat specialist and wider country side species on both farmland and woodland have shown long term and short term decline, with a 49 % reduction in butterfly abundance on farmland and a 73 % reduction in butterfly abundance in woodland between 1990 and 2013 (Defra, 2013). More specifically, assessment of ten-year trends has revealed that abundance of habitat specialist species have declined by 18 % and wider countryside species by 24 % when comparing abundance between the 1995-99 and 2005-09 assessment periods (Fox, *et al.*, 2011). The year 2012 was recorded as a historical low for butterfly species due to extreme rainfall throughout the summer (Defra, 2013).

1.5.2 Important landscape elements for butterflies

Severe declines in Lepidoptera species over recent years, even for more widespread common species, have been attributed to changes in agricultural practices, increased urbanization and possible effects of global warming, which have together changed the structure of landscapes (Thomas *et al.*, 2001; Dover & Settele, 2009; Lutolf *et al.*, 2009; Shreeve & Dennis, 2011). Furthermore, changes in habitat management practices, such as reduction in grazing and coppicing (see section 1.1.1 and 1.2.1), within and surrounding remaining isolated habitats has undermined habitat quality (Ouin, *et al.*, 2004; Pywell, *et al.*, 2004). Butterfly population persistence is directly related to landscape configuration, connectivity, composition and habitat quality which are intrinsically linked (Bergerot, *et al.*, 2012; Schweiger, *et al.*, 2006).

Landscape configuration and connectivity

Many generalist species are considered to be mostly active within edge habitats (Rossi and van Halder, 2010), with behaviour along habitat edges dependent on the surrounding landscape type (Bergerot, *et al.*, 2012). Graded woodland edges, hedgerows with boundary strips, field margins and conservation headlands are also thought to improve connectivity between semi-natural habitat fragments, such as woodlands (Pywell, *et al.*, 2004).

Linear features such as hedgerows and green lanes (tracks between fields) have been found to act as valuable corridors within an agricultural matrix, linking isolated patches of semi-natural habitat, and as such providing conduits for movement as well as additional habitat area (Croxton, *et al.*, 2005; Dover, *et al.*, 2000; Ouin, *et al.*, 2004). In particular, Croxton et al., (2005) found double hedged green lanes to support significantly higher floral and butterfly diversity and abundance in comparison to tracks with a single hedge or grass tracks without a hedge in an agricultural environment (bounded by fields).

Landscape composition and habitat quality

The quality of isolated semi-natural habitat fragments, such as woodlands, and grasslands, in an agricultural or urban matrix is pivotal in determining presence and absence of species as well as abundance, particularly of those species which persist within meta-populations (Dauber, *et al.*, 2003; Rossi and van Halder, 2010). Habitats with high quality are likely to support higher diversity and abundance of species (Dennis, 2010; Flick, *et al.*, 2012; Ouin, *et al.*, 2004). Habitat quality within woodland patches is determined by a diverse structure of dense undergrowth, primarily associated with rides and open glades and canopy clearings achieved through shifting management, such as long rotation coppicing with standards (see section 1.2.1). Reduction in coppicing in recent years has been linked to the decline of several woodland fritillary species dependent upon the early stages of the coppice rotation for the growth of their primary food plant *Viola spp.* (Clarke, *et al.*, 2011).

Habitat quality of grassland fragments in agriculturally dominated landscapes is determined by the provision of shelter and conservation headlands, width of non-crop habitat (field margin), floral diversity, and abundance of nectar and larval food

plants (Ouin, *et al.*, 2004; Pywell, *et al.*, 2004). Green lanes with uncropped verges, whether bounded by hedgerows or stone walls are suggested to provide numerous benefits including enhanced shelter, stable environmental conditions (micro-climates), buffers against chemicals such as herbicides and pesticides, and greater vegetation structural diversity which in turn provides more habitat suitable for breeding, ovipostion and larval development (Dover, *et al.*, 2000; Dover and Sparks, 2001). The enhanced habitat quality provided by green lanes has led to the suggestion that they may be superior to hedgerows in aiding dispersal (Dover, *et al.*, 2000). However, grassy verges sheltered by bordering hedgerows or woodlands, still facilitate butterfly activity as exposed areas become unsuitable, for example through increased wind speed, which is particularly important for immobile species (Pywell, *et al.*, 2004). Furthermore, grassy strips around woodlands not only provide additional habitat with diverse vegetation structure but also 'buffer' the impact from surrounding land use. In particular, agricultural intensification has been coupled with increased use of insecticides, herbicides, and fertilisers which can lead to soil improvement, and in turn a reduction in the diversity of larval and adult food plants (Clarke, *et al.*, 2011).

Spatial scale

Dover and Settele (2009) suggest that all features within a landscape should be considered to be landscape elements, but with different attributes. As such, landscape elements can contain resources for different species during different aspects of their life cycle (see section 1.5), with some elements proving to be more important than others, depending on the species (Dennis and Hardy, 2007; Ouin, *et al.*, 2004). For example, the gatekeeper (*Pyronia tithonus*) utilises tall grassy areas near to shrubs as sites for oviposition, hibernation, and as a food source for larvae (Asher, *et al.*, 2001). As an adult, the gatekeeper will utilise shrubs as sites for mate location, basking and as a source of nectar (Dennis, 2004). The use of different habitat patch types within a landscape for obtaining vital resources is known as landscape complementation (Dunning, *et al.*, 1992), and is a process commonly associated with butterfly species. In contrast some species may substitute their resource use by utilising similar patches which are nearby (Ouin, *et al.*, 2004). Consideration of the process of landscape complementation is essential as the majority of empirical work seeking to understand the role of the landscape has focused only on (1) semi-natural

habitats or the single habitat patch of the study species, assuming the surrounding matrix to be inhospitable or (2) specific species with very limited dispersal capabilities (Ockinger and Van Dyck, 2012).

The dispersal capability of the species in question determines the behavioural response of that species to the landscape and in turn functional connectivity (Billeter, *et al.*, 2008; Shreeve and Dennis, 2011). For example, highly mobile or migratory species are less likely to be affected by landscapes characterised by compositional or structural aspects that are considered to be impermeable for species with low dispersal abilities (Cozzi, *et al.*, 2008). Even species characterised with a relatively average dispersal ability (such as the speckled wood butterfly) has been shown to have reduced mobility in highly fragmented landscapes, with distances of daily movement decreasing with increasing isolation between woodland patches (Bergerot, *et al.*, 2012). Consequently, much of the empirical work focused on specific species reveals varying responses to spatial and temporal changes in landscape structure, depending on the species in question and whether it has a high or low dispersal capability (Rossi and van Halder, 2010; Shreeve and Dennis, 2011). It is therefore important to consider multiple species taking into account the different scales at which they respond to their environments, in order to effectively study landscape complementation without bias (Billeter, *et al.*, 2008).

1.5.3 Capturing multiple species response to landscape pattern

Landscape pattern affects butterfly species in multiple ways and at multiple spatial scales, with different aspects of the landscape affecting different stages of a species life cycle (Robinson, *et al.*, 2014; Shreeve and Dennis, 2011). Therefore considering multiple species is essential for capturing these conflicting relationships, and measuring community assemblages, species diversity and richness is important for representing wider biodiversity. However, the application of species richness and diversity indices are limited as these measures do not reflect differences in species composition between different habitats (Aavik and Liira, 2009). Communities comprise an assemblage of various species with differing ecological and biological attributes and in turn different response patterns to the same ecological variable; measures of richness and diversity alone will not measure the ecological and biological attributes therefore are somewhat limited in their application (Rossi and van Halder, 2010; Vandewalle, *et al.*, 2010). Therefore, species richness measures

obscure the functional relationship between species and environmental variables, which may lead to misleading conclusions (Azeria, *et al.*, 2009; Lindenmayer and Likens, 2011). However, this limitation may be overcome by the grouping of species according to their habitat specialism and in turn ecological similarities and these groups will potentially have similar response patterns to environmental change (Azeria, *et al.*, 2009; Blamires, *et al.*, 2008; Mac Nally, *et al.*, 2008). Such an approach is known as 'deconstruction of richness patterns' and is essential for overcoming the limitations associated with species richness measures, allowing the detection of multiple response patterns (Blamires, *et al.*, 2008; Marquet, *et al.*, 2004).

Ecological Attribute Groups (EAGs)

Shreeve *et al.*, (2001) propose capturing the behavioural response of butterfly species movement in response to resource distribution (functional connectivity) by grouping species with similar resource requirements and ecological attributes, and in turn similar responses to landscape change. These groups consider the resource requirements and traits associated with the life cycle of individual species, which is reflected in the plant species present at different times as well as the voltinism of each species. As such this approach aggregates species associated with different aspects of the landscape for each stage within their life cycle.

This behavioural approach to grouping species is considered more robust for understanding the response of species groups to environmental change than those based on biotope associations or mobility, which are widely used (Shreeve *et al.*, 2001; Shreeve & Dennis, 2011). Incorporating measures of species traits within biodiversity indicators has become increasingly popular within the scientific community (Firbank, *et al.*, 2008; van Strien, *et al.*, 2009) and is considered to control for ecological differences between species, providing a more accurate response to environmental change, and enabling the comparison of biodiversity over differing bio-geographic regions (Gregory, *et al.*, 2005; Vandewalle, *et al.*, 2010). Therefore a trait-based approach facilitates prediction of community dynamics (Vandewalle, *et al.*, 2010).

1.6 Predicting biodiversity in response to landscape pattern

It is widely accepted that landscape composition and configuration influences meta-population dynamics, and so landscape models, which capture the interaction between landscape structure and functional connectivity, can be used as a surrogate measure for biodiversity (Kadoya and Washitani, 2011; Rossi and van Halder, 2010), and are often used to predict the biodiversity response to different arrangements of land-use (Cowley, *et al.*, 2000). This is particularly important considering current and future rates and patterns of land-use change as well as the potential effects of climate change (Lawler, *et al.*, 2011). Predictive models, providing insight into species distribution within un-surveyed areas, are vital considering the time and monetary constraints on surveying species at a landscape scale, and would overcome the lack of species distribution data as robust comprehensive data sets are only available for some species groups (plants, birds and butterflies) (Cowley, *et al.*, 2000; Kumar, *et al.*, 2009; Lawler, *et al.*, 2011). Species distribution maps show the probability of species occurrence (or habitat suitability) derived from predictive models and are widely used to target effective conservation, and inform landscape planning (Cowley, *et al.*, 2000; Vaughan and Ormerod, 2005). Furthermore, the suitability of sites derived from presence-absence models can also be used to predict the likely abundance within occupied sites (Cowley, *et al.*, 2000).

Empirical models are the most widely applied model type for predicting species distribution, and these are based on a correlative relationship between species occurrence and landscape components (Guisan and Zimmermann, 2000; Lawler, *et al.*, 2011). Using the configuration and composition of the landscape in a predictive model therefore requires both the identification and quantification of relevant aspects of landscape and biodiversity components. Predictive models of butterfly distribution have been developed based on landscape components most notably by Cowley *et al.*, (2000), Rossi & van Halder (2010) and Fjellstad *et al.*, (2001). However, Cowley *et al.*, (2000) use a habitat-association method, so that only the compositional aspect of the landscape has been considered. Structural aspects of the landscape were considered in the empirical models developed by Rossi & van Halder (2010), across a multitude of scales. However, landscape models were developed with butterfly species richness. As discussed in section 1.5.3 measures of species richness encompass a range of species with often conflicting responses to landscape

components, and as such the variation in species response to landscape components has not been considered. Transferability of this model is also questionable, due to the delineation of habitats based on aerial photos and a non-standardised methodology for defining habitat patches. Landscape analysis and the transferability of landscape models is strongly dependent of the definition of land cover types (Rossi and van Halder, 2010; Turner, *et al.*, 2001). Rossi & van Halder (2010) conclude that the feasibility of predicting biodiversity from landscape metrics remains an open question. In particular, Dover and Settele (2009) state that, 'an adequate description of how landscapes influence butterflies is wanting.' The effects of landscape pattern on biodiversity remain unclear and further validation of the use of landscape pattern as an indicator of biodiversity is required (Feld, *et al.*, 2009; Haines-Young, 2009; Saura, *et al.*, 2008).

1.6.1 Methods for the measurement of landscape composition

Defining boundaries by landscape character

Assessment of environmental change at the landscape scale, in line with delivering the objectives of the Natural Environment White Paper, the Biodiversity 2020 strategy and the European Landscape Convention, is conducted considering landscape boundaries defined by National Character Areas (NCAs). Natural England has developed 159 National Character Areas (NCA) across England and these are natural areas with similar intrinsic 'character', considering characteristics that make that area unique such as biodiversity, geo-diversity and cultural activity. As such, boundaries between NCAs reflect natural change in landscape character providing a framework for assessing the state of ecosystem services and a context for landscape planning (Natural England, 2012a).

Land cover data: Land Cover Map

Advancement in computer power, development of Geographic Information Systems (GIS) and the availability of satellite imagery has facilitated the display and analysis of spatial patterns across the landscape, aided landscape planning, and ultimately played a critical role in the advancement of landscape ecology (Cowley, *et al.*, 2000; Lawler, *et al.*, 2011; Turner, *et al.*, 2001). In particular, the availability of satellite imagery has been pivotal for the analysis of land cover at the landscape and regional scale (Turner, *et al.*, 2001). For example, in the UK the Land Cover Map (LCM),

derived from satellite imagery obtained from the Landsat sensor, has been developed by the Centre for Ecology and Hydrology (CEH) for the assessment of Broad Habitats outlined by the Joint Nature Conservation Committee (JNCC) for the delivery of the UK BAP (Jackson, 2000). The LCM 2000 identifies 26 sub classes corresponding to 20 terrestrial broad habitats and the LCM 2007 identified 23 classes corresponding to 17 terrestrial broad habitats. LCM is available for the years 1990, 2000 and 2007, however, direct comparison between the different time periods is not advisable due to the differing classification techniques employed and improved spatial accuracy of the LCM 2007.

Land cover data: Phase 1 Habitat Map

The Phase 1 Habitat Classification (PH1) and field survey was originally developed in the 1970s and has since been widely used across the UK, particularly within the planning industry for the development of ecological baselines and environmental impact assessments (Cherrill and McClean, 1999; JNCC, 2010a). PH1 is a standardised method for recording semi-natural vegetation and habitats over large areas, with classification of habitat type dependent on the defining features within the habitat (Stevens, *et al.*, 2004). Each habitat has a unique mapping symbology and this has facilitated the application of this mapping technique within GIS. Ten broad habitats (not directly corresponding to UK BAP habitats) are identified with a further distinction of 155 specific habitat types, providing a realistic representation of semi-natural habitats reflecting habitat quality, particularly for grassland habitats with the distinction between different levels of improvement (JNCC, 2010a).

The PH1 mapping technique has facilitated the country-wide survey of Wales as well as being applied by local counties for the assessment of semi-natural habitats (Cherrill and McClean, 1999; Lucas, *et al.*, 2011; Stevens, *et al.*, 2004). For example, the Habitat Biodiversity Audit (HBA) managed by Warwickshire Wildlife Trust provides PH1 habitat information for Warwickshire, Coventry and Solihull (WCC, 2014). The HBA was developed as a joint project between the local authorities of Warwickshire County, Solihull and Coventry as well as Natural England and the Environment Agency in 1995 (WCC, 2014). The HBA is continually updated on an annual basis concentrating on different sections of Warwickshire each year with each field re-surveyed at least every 5 years (WCC, 2014). As such at different time

periods the PH1 map reflects changes that have occurred continuously over time and not just the habitats present in that year. Due to the time and resource constraints for developing PH1 maps at a county scale, digitised PH1 maps are only available for a limited number of counties (Cherrill, *et al.*, 1995), including Kent (ARCH, 2011).

1.6.2 Methods for the measurement of landscape pattern

Approaches for measuring landscape pattern are centred on the use of landscape pattern metrics, which quantify different aspects of the landscape, including landscape composition and configuration. The increased availability of landscape pattern calculation software has resulted in the development of numerous landscape pattern metrics (Li, *et al.*, 2005; Turner, 2005), from just three initially proposed by O'Neill et al., (1988) including fractal dimension, dominance and contagion, to more than one hundred in the most widely used programme FRAGSTATS (HainesYoung and Chopping, 1996; McGarigal and Marks, 1995).

Due to the vast number of landscape metrics available HainesYoung and Chopping (1996) proposed the division of metrics into groups according to the landscape aspect being measured. Metrics have been divided into the following five categories according to the landscape aspects in the most recent version of FRAGSTATS (version 4); area and edge; shape; contrast; aggregation; and diversity (McGarigal, *et al.*, 2012).

Metrics can be applied to measure aspects for each scale of heterogeneity within the landscape including the level of the cell, patch, habitat class or across the landscape as a whole. Categorical maps are based on patches, which are typically defined by composition (e.g. LCM or PH1 habitats) and landscape pattern metrics can capture the characteristics of each patch (patch level) and the spatial distribution of patches for each class type (class level), or across the landscape as a whole (landscape level) (McGarigal and Marks, 1995). As such patch attributes form the basis for the computation of several metrics which occur at the class and landscape level. Additionally, several metrics at the class and landscape level are derived from summarising patch level metrics providing a range of summary statistics which represent patch distribution (e.g. average, area-weighted average, standard deviation etc) (McGarigal and Marks, 1995). Metrics can also be computed to capture the local neighbourhood around each cell in the landscape, either considering (1) the

configuration of cells which make up a patch (cell level), or (2) the configuration of patches at the class or landscape level (Moving Window Analysis - MWA). Broad descriptions of the structural metrics associated with each landscape aspect, adapted from McGarigal *et al.*, (1995; 2014), are provided here and are discussed in more detail in Chapter 3.

Area and edge metrics

These metrics represent patch size and extent in addition to the amount of edge created by patches within the landscape and as such capture both the compositional and configuration aspect of the landscape (McGarigal, 2014). Measuring patch size is important when considering the species area relationship (Kennedy and Southwood, 1984) and the potential impacts of fragmentation and isolation on species dispersal and population persistence (Fahrig, 2003). Edge metrics also provide information on the fragmentation of the landscape, and, together with measures of the variability in patch size, provide information regarding spatial heterogeneity (McGarigal, 2014).

Shape metrics

The complexity of a patch shape directly influences the amount of edge habitat, and as such patch shape is considered to effect dispersal activities between patches such as migration and foraging (McGarigal, 2014; Saura, *et al.*, 2008). There are several methods for computing the complexity of patch shape, with the simplest method in FRAGSTATS comparing the perimeter of a patch with that of a standard shape of the same size to provide a measure of complexity (McGarigal, 2014).

Contrast metrics

The contrast between neighbouring patch types directly influences landscape permeability, and so the capacity for a species to disperse, and in turn the functional connectivity of the landscape (Baguette and Van Dyck, 2007; Ockinger and Van Dyck, 2012; Tischendorf and Fahrig, 2000). As such edge contrast directly influences ‘edge effects’ and the degree to which species will utilise the surrounding matrix (landscape supplementation or complementation) (Dunning, *et al.*, 1992; McGarigal, 2014).

Aggregation metrics

Aggregation refers to dispersion (spatial distribution of a class type), interspersion (intermixing between classes), sub-division (fragmentation of patches) and isolation (distance between patches). In particular, isolation measures the distance from a patch to the nearest neighbouring patch of the same type, based on edge to edge distance (McGarigal, 2014). Knowledge of the aggregation of the landscape is important when researching habitat fragmentation, island biogeography (MacArthur and Wilson, 1967) and meta-population theory (Hanski, 1994; Levins, 1969).

Diversity

There are numerous metrics for calculating the richness and diversity of class types at the landscape level, including patch richness, Shannon's diversity index, and the Simpson's diversity index (Maurer and McGill, 2011). These indices capture the composition of the landscape but do not reflect the spatial distribution of patches, or the ecological importance of individual class types (McGarigal, 2014). However, the diversity of land cover types in the landscape can be important for supporting species diversity through a greater provision of niches (Fischer and Lindenmayer, 2007; Schindler, *et al.*, 2008; Tscharntke, *et al.*, 2005).

The aim of this thesis was to enhance our understanding of the application of metrics to measure landscape characteristics and to develop landscape-based models to predict the distribution of butterfly species and community characteristics.

1.7 Hypotheses

1. Landscape structure metrics are influenced by spatial scale; differences in landscape patterns are consistent across scales (Chapter 3).

2. Composition, connectivity and structure of landscapes can be used to predict the presence-absence of butterfly species and butterfly Ecological Attribute Groups (EAGs) (Chapter 4); a combination of these measures will produce the best predictive model.

3. Landscape models developed in (2) can be used to predict the presence-absence and community assemblage of butterfly species from a temporally independent landscape data set (Chapter 5).

Chapter 2: Materials and methods

2.1 Assessment of spatial scale (Chapter 3)

2.1.1 Landscape extent

Two landscape extents were considered; firstly landscape extents were selected from within the UK, which have been defined, in relation to the National Character Areas (NCAs) described in section 1.6.1 (Section 3.2) (Natural England, 2012a). Secondly, the county of Warwickshire was considered and landscape extent was defined by imposing a 1 km x 1 km grid across Warwickshire, resulting in a total of 2467 individual landscapes, referred to hereafter as grid squares (Section 3.3) (Figure 2.1).

2.1.2 Landscape data

Two data sources were considered; The Land Cover Map 2000 (LCM 2000) provided by the Centre for Ecology and Hydrology (Fuller, *et al.*, 2002) and the Warwickshire Habitat Biodiversity Audit for the year 2000 (referred to as Phase 1 Habitat Map (PH1 2000) hereafter), provided by Warwickshire County Council (see section 1.6.1). The LCM 2000 raster version identifies 26 level 2 land cover classes corresponding to 20 terrestrial broad habitats classified at a 25 m x 25 m (25 m hereon) grain size (Fuller, *et al.*, 2002). The PH1 2000 was similarly rasterised to enable application of landscape structure metrics and extraction of landscape composition information using the software FRAGSTATS (McGarigal, *et al.*, 2012), and this was conducted at 25 m grain size to enable comparison between the two data sources. The Phase 1 Habitat mapping technique has the potential to identify up to 155 specific habitat types (JNCC, 2010a), however, the Warwickshire PH1 2000 classifies only semi-natural habitats. Therefore, prior to rasterisation unclassified polygons were re-classified as either 'built up' or 'infrastructure' according to the classification of corresponding polygons in the OS open data 'Vector Map Local' (Ordnance Survey data © Crown copyright and database right 2011). Only major roads (motorways, A roads and B roads) and railways were classified as 'infrastructure' and minor roads were classified as 'hardstanding'. The PH1 2000 vector map was rasterised using the maximum combined area method which combines all features with the same value within a cell and then assigns a value to the cell according to the feature with the largest combined area (ArcGIS, 2011). Therefore all features are considered within a cell during the rasterisation process not just the feature within the centre of the cell or a single feature with the largest area.

Once rasterised, unclassified pixels were reclassified corresponding to the most frequent pixel value within its neighbourhood (ArcGIS, 2011).

Land cover was derived from the LCM 2000 for landscapes defined by NCA boundaries. However, land cover was derived from both the LCM 2000 and PH1 2000 for landscapes defined by the Warwickshire grid squares. The coverage of the PH1 2000 across Warwickshire is smaller than that achieved by LCM 2000 (*n* = 2427) and as such there were fewer grids squares for the PH1 2000 (*n* = 2080) (Figure 2.1).

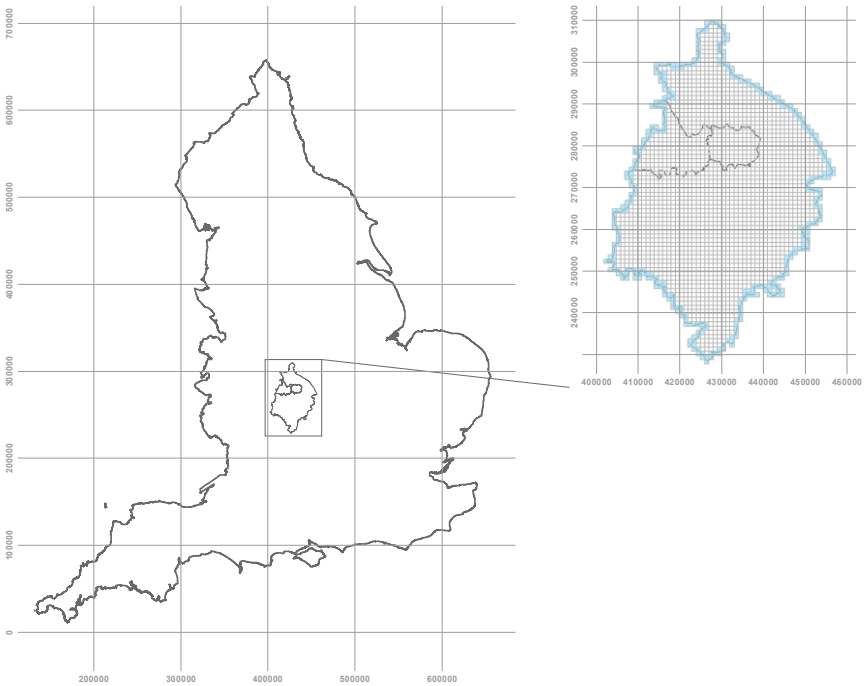


Figure 2.1: Location of Warwickshire within England, including the metropolitan districts of Coventry and Solihull, and the application of 1 km grid squares across the district. The blue squares indicate grid squares on the periphery of Warwickshire, which were not covered by the PH1 2000 data set.

Classification of landscape characteristics

For the NCA landscapes (n = 159) and grid square landscapes (n =2467; n = 2080) the Mean Patch Size (MPS), Diversity of Land Cover (DLC) and Number of Land Cover (NLC) classes were computed. Additionally the Total Area (TA) was computed for each of the NCA landscapes. NCA landscapes were grouped according to the quartile ranges of TA and MPS (Table 2.1a), and additionally by contrasting NLC (low 3 – 16 and high 17 - 23). Grid square landscapes were grouped by MPS and NLC (Table 2.1b).

Diversity of land cover groups were identified for each landscape by means of a hierarchal cluster analysis (HCA). HCA was used to group landscapes depending on their similarity in landscape composition, considering the distribution of area between each land cover type. The similarity coefficient between landscapes was calculated using Euclidean Distances, and landscapes were then clustered using the Complete Link method, which considers the minimum similarity between two samples in a cluster to define the similarity between two clusters. The HCA for the NCA landscapes (based on the LCM 2000) was performed by Mead (Warwick) and nine groups were identified at a cut off threshold of 95 % (Table 2.2a). For the grid square landscapes cut-off values of 85 % were chosen to define compositional groups and 14 groups were identified based on the LCM 2000 (Table 2.2b) and 19 groups identified based on the PH1 2000 (Table 2.2c). Clusters were identified at similarity levels which maximises homogeneity of clusters (average similarity of the members) and separation of the clusters (average dissimilarity of each cluster to its nearest neighbour) (Jongman, *et al.*, 1995). The Simpsons Diversity Index was calculated considering the land cover composition for each group identified from the HCA, and the groups are described hereon as Diversity of Land Cover (DLC) groups.

From a total of 159 NCAs, 32 were selected for further analysis, based on their landscape characteristics (TA, MPS, DLC and NLC) (Figure 2.2). Two NCAs were selected from each of the 16 quartile combinations of TA and MPS (Table 2.1), which also contrasted in the number of land classes (NLC) they contained. In addition to these criteria, NCAs were selected from within each of the nine DLC groups.

(a)

NCA Group	Total Area (TA) (ha)	Mean Patch Size (MPS) (ha)
small	1124 – 34645	4.0 – 7.9
medium	36487 – 63726	8.0 – 9.9
large	64071 – 98594	10.0 – 11.9
extra large	101141– 382606	12.0 – 23.15

(b)

Grid square Group	LCM 2000 Mean Patch Size (MPS) (ha)	PH1 2000 Mean Patch Size (MPS) (ha)
small	2.1 - 3.4	0.7 – 2.7
medium	3.5 - 4.2	2.8 – 4.0
large	4.3 – 5.3	4.1 – 5.9
extra large	5.4 – 20.0	6.0 – 33.3

Grid square Group	LCM 2000 Number of Land Classes (NLC)	PH1 2000 Number of Land Classes (NLC)
1	3 - 7	1-7
2	8	8 - 10
3	9	11-12
4	10 - 11	13 - 20

Table 2.1: The grouping of landscape extents by landscape characteristics; (a) the groupings of NCAs by total area (TA), and mean patch size (MPS) and (b) the groupings of grid squares derived from LCM 2000 and PH1 2000 by mean patch size (MPS) and number of land classes (NLC). Groups are determined by quartiles, and the range for the corresponding landscapes within each quartile group is provided.

(a)	NCA(n = 32)		
DLC group	LCM 2000 LSIDI		
	Min	Max	Average
1	0.787	0.843	0.822
2	0.860	0.872	0.866
3	0.778	0.837	0.814
4	0.762	0.842	0.802
5	0.750	0.865	0.821
6	0.835	0.859	0.848
7	0.716	0.772	0.744
8	0.615	0.719	0.684
9	0.683	0.762	0.716

(b)	1 km grid squares (n = 2427)		
DLC group	LCM 2000 LSIDI		
	Min	Max	Average
1	0.734	0.853	0.788
2	0.633	0.865	0.775
3	0.547	0.866	0.769
4	0.677	0.794	0.765
5	0.587	0.854	0.764
6	0.458	0.881	0.758
7	0.625	0.819	0.746
8	0.657	0.827	0.743
9	0.342	0.829	0.723
10	0.281	0.833	0.684
11	0.245	0.854	0.670
12	0.100	0.819	0.554
13	0.254	0.745	0.550
14	0.061	0.627	0.467

(c)	1 km grid squares (n = 2080)		
DLC group	PH1 2000 LSIDI		
	Min	Max	Average
1	0.793	0.870	0.832
2	0.695	0.843	0.789
3	0.761	0.822	0.787
4	0.679	0.879	0.781
5	0.675	0.907	0.778
6	0.556	0.885	0.755
7	0.617	0.862	0.753
8	0.271	0.860	0.723
9	0.637	0.792	0.714
10	0.588	0.807	0.711
11	0.574	0.851	0.705
12	0.292	0.880	0.687
13	0.524	0.771	0.681
14	0.516	0.773	0.669
15	0.446	0.843	0.637
16	0.496	0.732	0.614
17	0.118	0.908	0.594
18	0.000	0.851	0.513
19	0.172	0.758	0.452

Table 2.2: The grouping of landscape extents defined by the landscape characteristic Diversity of Land Cover (DLC). Groups were identified by a Hierarchical Cluster Analysis based on the composition of land covers/ habitats within each landscape extent. The Simpsons Diversity Index for each landscape (LSIDI) is provided for (a) the nine DLC groups identified for the NCA landscape using the LCM 2000 data set, (b) the 14 DLC groups identified for the 1 km grid squares based on the LCM 2000 and (c) the 19 DLC groups identified for the 1 km grid squares using the PH1 2000.

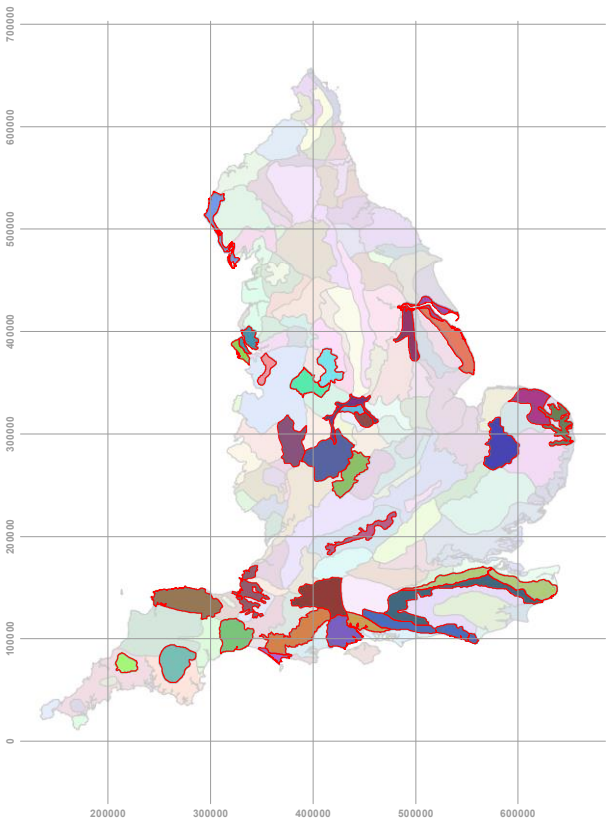


Figure 2.2: Location of 32 selected National Character Areas (NCAs) (bold colours) across England which range in the total area, mean patch size, diversity of land cover and number of land cover classes.

2.1.3 Measuring Landscape Structure

Using the software FRAGSTATS (version 4) (McGarigal, *et al.*, 2012) landscape structure metrics were calculated for each of the 32 NCAs using data derived from the LCM 2000, and for each Warwickshire grid square using the LCM 2000 and PH1 2000. A total of 69 metrics were initially considered which measured the landscape aspects of area/ edge, shape, aggregation, and diversity (see section 1.6.2). For the computation of metrics considering patch type adjacency, including edge contrast, contagion and interspersion a border was applied to each landscape (1 km border for NCAs and 500 m border for grid squares). Addition of a border enables the computation of metrics considering characteristics of patches which continue beyond the boundary of the landscape. Edge contrast and similarity weights for each land class adjacency for the LCM 2000 data classes was based on those developed by Skirvin and Mead (*pers.comm.*) considering similarities in plant attributes data set (Appendix A1). Edge contrast and similarity weights for each habitat adjacency for

the PH1 2000 was developed based on similarities in plant community composition derived from the literature (Appendix A2).

Co-linearity between landscape metrics is common (Schindler, *et al.*, 2008) and the number of landscape metrics was greater than the number of NCA landscapes. Pearson's Product Moment Correlations were therefore calculated on the output of the 69 landscape metrics applied to the 32 NCA landscapes, in order to identify pairs of highly correlated metrics at the 1% significance level, and therefore where metrics were potentially redundant (Appendix A3). Landscape structure metrics which did not meet the assumptions of normality and homoscedasticity, were $\log_{10}$ transformed (25 in total), and landscape structure metrics which were proportions or percentages (six in total) were arcsine transformed. Representatives from significant pairwise correlations were chosen on the basis of their ecological significance, resulting in a final selection of 32 metrics for further analysis (Table 2.3; Appendix A4). Selected metrics exhibited each type of scaling relation proposed by Wu (2004) (see section 3.1). For the calculation of the metric proximity index and similarity index a search radius of 2 km was considered.

2.1.4 Discriminating between landscapes: NCAs and grid squares

A One-way Analysis of Variance (ANOVA) was used to examine the impacts of the landscape characteristics (TA, MPS, DLC and NLC) on each of the 32 landscape structure metrics and determine the difference between different levels of each landscape component on the 32 landscape structure metrics. For three metrics (CIRCLE_RA, CONTIG_RA, and ENN_MN) derived from the PH1 2000 data set for the 1 km grid square landscapes, the non-parametric Kruskal-Wallis Test was performed due to non-normality of these three metrics.

Key landscape structure metrics were identified from a Principal Component Analysis (PCA) performed on a correlation matrix between the 32 landscape structure metrics (separately for NCA and grid square landscapes). PCA reduces the dimensionality of the data and as such facilitates the identification of independent variables that account for as much of the variation possible within the original data (Jongman, *et al.*, 1995). Furthermore, PCA enables the identification of potential underlying factors explaining the variation between landscapes, facilitating the identification of landscapes with similar structure.

Acronym	Metric	Units (range)
AREA_MN	Mean Patch Size	ha (>0)
AREA_RA	Range in Patch Size	ha
CIRCLE_AM	Area-weighted Mean Related Circumscribing Circle	None (0-1)
CIRCLE_MN	Mean Related Circumscribing Circle	None (0-1)
CIRCLE_RA	Range in Related Circumscribing Circle	None (0-1)
COHESION	Patch Cohesion Index	None
CONTAG	Contagion index	% (0-100)
CONTIG_AM	Area-weighted Mean Contiguity Index	None (0-1)
CONTIG_MN	Mean Contiguity Index	None (0-1)
CONTIG_RA	Range in Contiguity Index	None (0-1)
CWED	Contrast weighted edge density	m/ha (0>)
ECON_AM	Area-weighted Mean Edge Contrast Index	% (0-100)
ECON_CV	Edge Contrast Index coefficient of variation	%
ENN_AM	Area-weighted Mean Euclidean Nearest-Neighbour Distance	m (>0)
ENN_CV	Euclidean Nearest-Neighbour Distance co-efficient of variation	%
ENN_MN	Mean Euclidean Nearest-Neighbour Distance	m (>0)
FRAC_AM	Area-weighted Mean Fractal Dimension Index	None (1-2)
FRAC_CV	Fractal Dimension Coefficient of Variation	%
GYRATE_AM	Area-weighted Mean Radius of Gyration	m (≥ 0)
GYRATE_CV	Radius of Gyration Coefficient of Variation	**%**
GYRATE_MN	Mean Radius of Gyration	m (≥ 0)
IJI	Interspersion and juxtaposition index	% (0-100)
LSI	Landscape Shape Index	None (≥1)
MESH	Effective Mesh Size	ha (≥0.06)
PRD	Patch richness density	No/100 ha (>0)
PROX_AM	Area-weighted Mean Proximity Index	None (≥ 0)
PROX_CV	Proximity Index co-efficient of variation	%
SHAPE_CV	Shape Index Coefficient of Variation	%
SHAPE_MN	Mean Shape Index	None (≥1)
SIDI	Simpson's diversity index	None (0-1)
SIMI_AM	Area-weighted Mean Similarity Index	None (≥ 0)
SIMI_CV	Similarity Index co-efficient of variation	%

Table 2.3: The 32 landscape structure metrics selected for further analysis. The Acronym, metric name, units and range are provided for each metric, and are ordered alphabetically. For a full description of each metrics see Appendix A.4.

2.1.5 Influence of spatial scale on metric discrimination

The ability of metrics to discriminate between landscapes at different grain sizes was investigated. Following the methodology of Baldwin *et al.*, (2004) the original landscape data sets (LCM 2000 and PH1 2000) with a 25 m grain size were re-sampled using a majority rule method in ArcGIS v10 to create five resolution levels; 50 m, 100 m, 250 m, 500 m and 1000 m (Figure 2.3a-f). The majority method groups neighbouring pixels together, reclassifying pixels based on area (Simova and Gdulova, 2012). To avoid cumulative aggregation error re-sampling was performed using the original 25 m grain size (Baldwin, *et al.*, 2004). At each scale the 32 landscape structure metrics identified in section 2.1.4 were calculated (Table 2.3).

A One-way ANOVA was calculated at each scale to examine the impacts of scale on the discrimination between landscapes by the 32 landscape structure metrics on the basis of the characteristics TA, MPS, DLC and NLC. For three metrics (CIRCLE_RA, CONTIG_RA, and ENN_MN) derived from the PH1 2000 data set for the 1 km grid square landscapes the non-parametric Kruskal-Wallis Test was performed due to non-normality of these three metrics at scales of 50 m – 250 m.

A PCA was conducted at each scale in order to detect which metrics discriminate between landscapes facilitating the identification of landscapes with similar landscape structure. Procrustes rotation was conducted on the scores (landscape extent) and loadings (landscape structure metrics) of the first four principal components from each PCA to assess the relative similarities of associations between landscapes and metric values across the scales.

Hierarchical Cluster Analysis (HCA) was also conducted on scores and loadings from the first four components from the PCA to identify clusters of landscapes and metrics at each spatial scale. The similarity coefficient between scores for each landscape extent was calculated using Euclidean Distances, and landscapes were then clustered using the Complete Link method, which considers the minimum similarity between two samples in a cluster to define the similarity between two clusters. This process was then repeated considering the similarities between the loadings of the 32 landscape structure metrics. Clusters were identified at different similarity levels, based on maximising homogeneity of clusters (average similarity of the members) and separation of the clusters (average dissimilarity of each cluster to its nearest

neighbour) (Jongman *et al.*, 1995). Mantel Product Moment Correlation were calculated to assess the degree of correlation between the Euclidean Distance similarity matrices at each spatial scale.

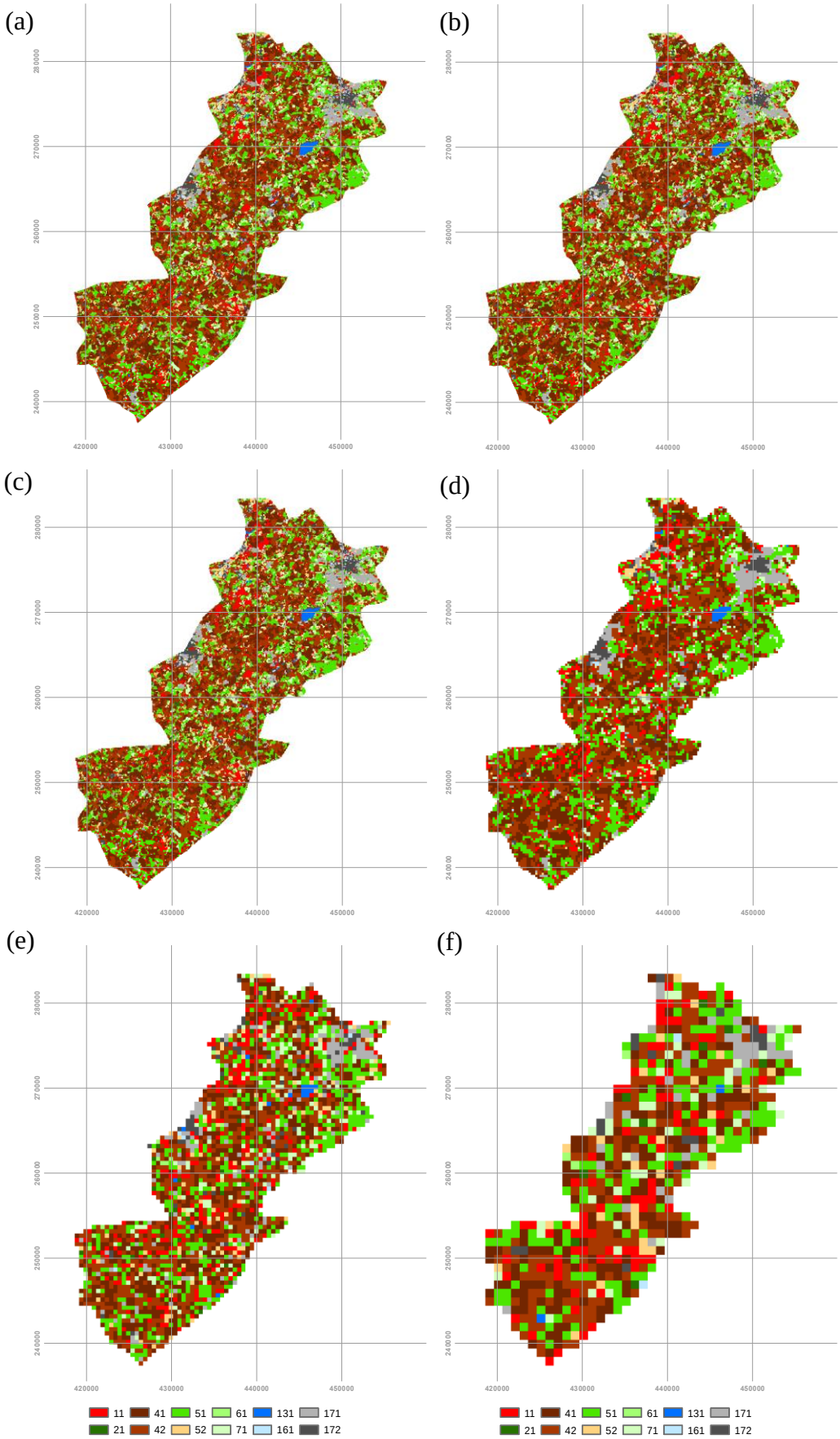


Figure 2.3a-f: The progressive increase in grain size for National Character Area 96 (Dunsmore and Feldon) from 25 m to 1000 m obtained by resampling the original 25 m using the majority method. The five resolution levels used in further analysis are shown; (a) 25 m (b) 50 m (c) 100 m (d) 250 m (e) 500 m and (f) 1000 m.

2.2 Model development (Chapter 4)

2.2.1 Landscape extent

The location of this study was Warwickshire, and the surrounding Coventry and Solihull metropolitan districts, referred to collectively as Warwickshire hereon. Warwickshire, located within the centre of England, UK (52 18'N, 1 34'W; Ordnance Survey, 10km squares SK and SP) is entirely land locked and is bordered by the extensively urbanised city of Birmingham to the North West (Figure 2.1). Landscape extent was defined by imposing a 1 km grid across Warwickshire, resulting in a total of 2467 individual landscapes, referred to hereafter as grid squares (Figure 2.1).

Warwickshire is characterised by a varied landscape, with its boundaries falling within eight National Character Areas (NCAs), two of which are major NCAs of lowland Britain, 'Arden' and 'Dunsmore and Feldon' (Falk, 2009). Arden, located to the north-west of Warwickshire, is characterised by ancient countryside, with networks of small and irregular field systems bordered by distinctive, ancient hedgerows (Falk, 2009; Natural England, 2012b). This ancient countryside is coupled with poor soils, sand and gravel (Falk, 2009). Within Warwickshire, Arden is also characterised and influenced by the urbanised centres of Coventry and Nuneaton (Natural England, 2012b). The land use across 'Dunsmore and Feldon', located to the south-east of Warwickshire, is dominated by grassland and arable cereals and is characterised by large regular hedged fields, fertile land, and cleared woodland (Falk, 2009, Natural England, 2012c). This NCA comprises mostly calcareous clays, producing fertile alkaline soils within Feldon, whilst Dunsmore is associated with poor acidic soils with bracken-dominated sites (Falk, 2009, Natural England, 2012c).

2.2.2 Butterfly data: 1990-1999

Butterfly species records were obtained for the years 1990-1999 from the UK Butterfly Monitoring Scheme (UKBMS) and Warwick Biological Records Centre (WBRC) (Warwickshire County Council), which collated butterfly data from the Butterfly Conservation General Recording Scheme. The species data from the WBRC, was recorded with an average spatial accuracy of 1 km, a standard data resolution for wide-scale monitoring schemes (UKBMS, 2014). Data from these two

sources were combined to provide a greater spatial distribution of data across Warwickshire, with records occurring within 515 1 km grid squares. Data from WBRC with unknown or 'ongoing' years were removed, as the reliability of these records is questionable. The skippers, small skipper (*Thymelicus sylvestris* Poda) and Essex skipper (*Thymelicus lineola* Ochsenheimer) are often confused on transect surveys, hence records of these species were combined within transects where they co-occur (Cowley, *et al.*, 2000).

Determining species presence-absence

Aggregated records for all species were used to classify the presence or absence of butterflies within each grid square. For the combined species data, a species record within a grid square was taken as 'presence' and no record within a square as 'absence' (LCM 2000 n = 2467; PH1 2000 n = 2080).

Individual species were grouped into four Ecological Attribute Groups (EAG1-4) identified by Shreeve *et al.*, (2001) (see sections 1.5.3 and 4.2.1). A fifth group was created to categorise migrants from outside of the British Isles (EAG0). This migrant group, however, was not considered during statistical analysis because these species were found in very low numbers and their behavioural response to landscape variables is less likely to be captured due to their migratory abilities (Shreeve & Dennis, 2011). Additionally EAG4 comprised a low spatial distribution across Warwickshire and as such was not considered during the statistical analysis (Table 4.2). Within the 515 occupied squares, presence of species classified within a particular EAG was taken as a 'presence' for that group within the grid square and if a grid square did not comprise a record for any species within a particular EAG, this was taken to represent an 'inferred absence' for that EAG (n = 515).

Determining species dispersal distances

Dispersal was considered to include any movement between habitat patches in accordance with Sekar (2012). The mean dispersal distances identified from mark release recapture of 12 Warwickshire butterfly species were identified from the literature (Sekar, 2012; Stevens, *et al.*, 2010). For all butterfly species the median dispersal distance of 323 m was obtained, and for species comprising EAG 1-3 a median dispersal distance of 262 m was obtained. Butterflies are generally

characterised by daily movements ranging from 200 – 600 m (Davis, *et al.*, 2007). The median dispersal distance of 325 m for all butterfly species and 250 m for EAG1-3 species was therefore used in further analysis in sections 2.2.4.

2.2.3 Landscape data

The rasterised LCM 2000 and the PH1 2000 data sets, with a grain size of 25 m resolution, were used for the model development and these data sets are described fully in section 2.1.2. The area of each land cover/ habitat, measured in hectares (ha), was extracted from the LCM 2000 and PH1 2000 data for each 1 km grid square, equivalent to the scale of species data summaries, using ArcMap v10. Landscape diversity, calculated using the Simpsons Diversity Index (LSIDI) based on land cover area, and landscape heterogeneity (NLAND), based on the number of classified habitats, were also identified within each 1 km grid square.

The LCM 2000 identifies 26 land cover classes corresponding to 20 terrestrial broad habitats (Fuller, 2002), with 13 land cover classes identified within Warwickshire (Table 2.4). When considering all grid squares within Warwickshire, arable non-rotational (LCM-5) has a small spatial distribution, occurring within only 0.29 % of squares (Table 2.4). Within the 515 grid squares containing butterfly records this land cover is no longer present, and as such is not considered in further analysis for EAG species (Table 2.4). Within Warwickshire the PH1 2000 identifies 44 semi-natural habitats including the habitats ‘built/ hard standing’ and ‘infrastructure’ (Table 2.5). There are several PH1 habitats which occur in only 0.05 % of grid square and have an average coverage across Warwickshire of less than 0.001 ha. These are basin mire (PH1-43), acid/ neutral flush (PH-26), and coniferous semi-natural woodland (PH-3) (Table 2.5). These were omitted prior to the analysis for all grid squares and the 515 grid squares. When considering the 515 grid squares, an additional two habitats, ‘non-ruderal’ (PH-24) and ‘spoil’ (PH-32) were characterised by small spatial distribution and average coverage and as such were omitted prior to further analysis (Table 2.5).

Land Cover Map 2000	Code	All butterflies (*n* = 2467)					EAGs (*n* = 515)				
		Prop (%)	**Mean (ha)**	**±SE**	**Min (ha)**	**Max (ha)**	**Prop (%)**	**Mean (ha)**	**±SE**	**Min (ha)**	**Max (ha)**
Broad-leaved/ mixed woodland	LCM-11	95.51	9.81	0.17	0.00	60.13	97.09	12.26	0.46	0.00	60.13
Coniferous woodland	LCM-21	41.16	1.28	0.07	0.00	45.75	47.96	1.58	0.17	0.00	43.81
Arable cereals	LCM-41	87.31	17.21	0.32	0.00	83.13	83.69	13.19	0.58	0.00	59.25
Arable horticulture	LCM-42	97.90	26.37	0.31	0.00	86.06	98.45	22.66	0.62	0.00	72.06
Arable non-rotational	LCM-43	0.29	0.04	0.02	0.00	23.13	0.00	0.00	0.00	0.00	0.00
Improved grassland	LCM-51	97.65	19.31	0.30	0.00	96.88	97.67	18.89	0.62	0.00	66.88
Set-aside grassland	LCM-52	79.89	3.74	0.09	0.00	46.00	87.57	4.66	0.22	0.00	28.38
Neutral grassland	LCM-61	23.90	1.38	0.08	0.00	63.13	25.24	1.62	0.18	0.00	31.50
Calcareous grassland	LCM-71	79.27	6.69	0.15	0.00	53.19	80.97	7.13	0.32	0.00	43.81
Water (inland)	LCM-131	20.73	0.60	0.06	0.00	80.44	27.77	1.44	0.26	0.00	80.44
Inland bare ground	LCM-161	17.63	0.57	0.04	0.00	24.88	23.50	0.77	0.10	0.00	18.19
Suburban/ rural developed	LCM-171	80.63	9.74	0.33	0.00	94.81	84.08	11.83	0.73	0.00	92.38
Continuous urban	LCM-172	46.60	3.27	0.17	0.00	85.81	54.17	3.97	0.38	0.00	70.63
Landscape diversity	LSIDI	-	0.72	0.00	0.06	0.88	-	0.74	0.00	0.14	0.88
Landscape heterogeneity	NLAND	-	7.68	0.03	3.00	12.00	-	8.08	0.06	4.00	12.00

Table 2.4: The distribution in area of the 13 land covers derived from the Land Cover Map 2000 (LCM 2000) across Warwickshire, considering the grid squares used for determining presence-absence of all butterfly species (*n* = 2467) and species grouped by their ecological attributes (EAGs) (*n* = 515). The average coverage, range and proportion of grid squares comprising each land cover is provided. LCM land covers are ordered by the code for each cover.

Phase 1 Habitat	Code	All butterflies (n = 2079)					EAGs (n = 466)				
		Prop (%)	Mean (ha)	±SE	Min (ha)	Max (ha)	Prop (%)	Mean (ha)	±SE	Min (ha)	Max (ha)
Broad-leaved semi-natural woodland	PH-1	74.60	2.36	0.11	0.00	58.44	76.82	3.68	0.35	0.00	58.44
Broad-leaved plantation	PH-2	65.90	1.39	0.06	0.00	35.81	70.39	1.81	0.17	0.00	31.63
Coniferous semi-natural woodland	PH-3	0.05	0.00	0.00	0.00	0.13	0.21	0.00	0.00	0.00	0.13
Coniferous plantation	PH-4	23.86	0.65	0.06	0.00	35.50	25.32	0.85	0.15	0.00	30.94
Mixed semi-natural woodland	PH-5	1.59	0.02	0.01	0.00	7.19	1.29	0.01	0.00	0.00	1.38
Mixed plantation	PH-6	24.29	0.64	0.06	0.00	42.88	25.32	0.77	0.16	0.00	40.25
Dense/continuous scrub	PH-7	60.41	0.54	0.02	0.00	13.31	73.39	0.99	0.08	0.00	13.31
Scattered scrub	PH-8	2.16	0.01	0.00	0.00	2.25	4.29	0.02	0.01	0.00	1.56
Broad-leaved parkland/scattered trees	PH-9	3.85	0.01	0.00	0.00	1.25	4.29	0.02	0.00	0.00	1.19
Recently felled woodland	PH-11	0.67	0.02	0.01	0.00	14.19	1.07	0.05	0.03	0.00	14.19
Orchard (commercial)	PH-12	3.08	0.06	0.02	0.00	23.81	1.72	0.03	0.02	0.00	10.38
Unimproved acidic grassland	PH-13	0.63	0.01	0.00	0.00	5.75	1.72	0.03	0.02	0.00	5.75
semi-improved acidic grassland	PH-14	0.87	0.01	0.00	0.00	4.19	2.36	0.02	0.01	0.00	2.88
Unimproved neutral grassland	PH-15	9.52	0.11	0.02	0.00	17.13	12.45	0.21	0.05	0.00	17.13
Semi-improved neutral grassland	PH-16	84.32	4.33	0.14	0.00	83.13	92.49	5.98	0.35	0.00	60.19
Unimproved calcareous grassland	PH-17	0.43	0.01	0.01	0.00	7.25	1.29	0.06	0.03	0.00	7.25
semi-improved calcareous grassland	PH-18	0.82	0.02	0.01	0.00	7.81	1.93	0.05	0.02	0.00	7.81
Improved grassland	PH-19	95.91	25.34	0.43	0.00	93.88	96.78	24.91	0.90	0.00	90.63
Marsh/marshy grassland	PH-20	14.05	0.12	0.01	0.00	13.25	20.17	0.18	0.03	0.00	7.25
Continuous bracken	PH-22	3.27	0.03	0.01	0.00	18.69	4.94	0.05	0.02	0.00	7.25
Tall ruderal	PH-23	40.45	0.32	0.02	0.00	14.38	48.07	0.45	0.04	0.00	7.44
Non-ruderal	PH-24	0.19	0.00	0.00	0.00	0.50	0.43	0.00	0.00	0.00	0.19
Dry heath/acid grassland mosaic	PH-25	0.10	0.00	0.00	0.00	3.56	0.43	0.01	0.01	0.00	3.56
Acid/neutral flush	PH-26	0.05	0.00	0.00	0.00	0.31	0.21	0.00	0.00	0.00	0.31
Swamp	PH-27	10.92	0.05	0.01	0.00	7.63	18.03	0.12	0.03	0.00	7.63
Inundation vegetation	PH-28	1.01	0.00	0.00	0.00	2.13	1.93	0.01	0.01	0.00	2.13
Standing water	PH-29	54.83	0.70	0.08	0.00	85.19	61.80	1.60	0.31	0.00	85.19
Running water	PH-30	17.27	0.19	0.02	0.00	12.00	23.18	0.27	0.04	0.00	12.00
Quarry	PH-31	2.50	0.26	0.05	0.00	42.25	6.22	0.67	0.17	0.00	42.25
Spoil	PH-32	0.05	0.00	0.00	0.00	1.13	0.00	0.00	0.00	0.00	0.00
Refuse tip	PH-33	0.58	0.03	0.01	0.00	25.56	1.29	0.04	0.02	0.00	8.38
Arable	PH-34	93.03	44.50	0.59	0.00	100.00	92.27	35.06	1.16	0.00	94.31
Allotments	PH-35	11.01	0.18	0.02	0.00	14.88	13.52	0.25	0.05	0.00	14.88
Set-aside	PH-36	11.21	0.46	0.05	0.00	63.38	10.30	0.41	0.09	0.00	20.56
Amenity grassland	PH-37	67.29	3.50	0.15	0.00	62.25	69.96	4.26	0.37	0.00	59.06
Ephemeral/short perennial	PH-38	6.16	0.08	0.01	0.00	14.81	10.09	0.18	0.05	0.00	14.81
Introduced shrub	PH-39	1.88	0.01	0.00	0.00	2.31	2.58	0.01	0.01	0.00	2.31
Buildings	PH-40	81.05	10.83	0.43	0.00	90.81	84.33	13.08	0.91	0.00	90.19
Bare ground	PH-41	26.55	0.40	0.04	0.00	32.94	32.62	0.70	0.12	0.00	32.94
Basin mire	PH-43	0.05	0.00	0.00	0.00	0.88	0.21	0.00	0.00	0.00	0.88
Road/ infrastructure	PH-44	58.44	2.80	0.08	0.00	31.19	64.16	3.18	0.18	0.00	31.19
Landscape heterogeneity	NLAND	-	9.55	0.07	1.00	20.00	-	10.60	0.13	4.00	20.00
Landscape diversity	LSIDI	-	0.56	0.00	0.00	0.91	-	0.63	0.01	0.11	0.91

Table 2.5: The distribution in area of the 44 habitats derived from the PH1 habitat map 2000 (PH1 2000) across Warwickshire, considering the grid squares used for determining presence-absence of all butterfly species (n = 2467) and species grouped by their ecological attributes (EAGs) (n = 515). The average coverage, range and proportion of grid squares comprising each land cover is provided. PH1 habitats are ordered by the code for each habitat.

2.2.4 Model development; explanatory variables

Three sets of explanatory variables were considered; landscape compositional variables, landscape connectivity variables and landscape structural variables. Correlations among each set of explanatory variables were explored prior to the formal modelling process. Correlations among the landscape compositional variables and landscape connectivity variables were assessed separately by means of spearman rank correlation (*S*). The landscape structure metrics were normalised by $\log_{10}$ transformation, unless otherwise stated, and correlations assessed by means of Pearson Product Moment Correlations (*r*). Pairs of variables with strongly significant correlations (absolute $S > 0.8$; absolute $r > 0.8$) were identified and one of these variables was removed. All statistical analysis, unless otherwise stated, was conducted using the statistical package Genstat ® 13.2.

Landscape composition

Using the LCM 2000 and PH1 2000 data, total land cover/ habitat area (ha), landscape diversity (LSIDI) and landscape heterogeneity (NLAND) within each 1km grid square (see section 2.2.3) were used for the development of models using landscape compositional data. For the PH1 2000 data set total length of hedgerow in each grid square was also included (Table 2.6). Hedgerow data was obtained in addition to the Warwickshire Phase 1 Habitat Map as part of the Warwickshire Habitat Biodiversity Audit provided by Warwickshire County Council (see sections 1.6.1 and 2.1.2). Three classifications of hedgerows were identified; intact hedgerow, defunct hedgerow and hedge with trees (Table 2.6). All hedgerow types were considered in the analysis.

Hedgerow type	PH1 Code	Total frequency	Proportion of total length (%)
Intact hedge	J21	129232.00	78.52
Defunct hedge	J22	10473.00	7.28
Hedge with trees	J23	22480.00	14.21

Table 2.6: The proportion and frequency of each hedgerow type within Warwickshire in 2001 according to the PH1 habitat classification.

Spearman rank correlations were calculated to identify significant correlations between the landscape compositional variables for the two data sets (LCM 2000 and PH1 2000). No correlations with an absolute coefficient greater than $S \geq 0.8$ were identified amongst the landscape compositional variables for both data sets. As such no variables were omitted from the analysis on the basis of high spearman-rank correlations.

Landscape connectivity metrics

Land cover classes selected within the landscape compositional models are considered to be key land cover classes (see sections 4.2.2 and 4.2.6). The final landscape compositional models included 12 LCM land covers and 21 PH1 habitats. The connectivity of these key land cover classes was determined within each 1 km grid square using the software Conefor v2.6 (Saura and Torne, 2009). Two methods of obtaining measures of connectivity were considered; (1) connectivity of key land cover in isolated squares and (2) importance of patches within key land covers for maintaining connectivity across Warwickshire. For the computation of the connectivity variables, a border was applied to each 1 km grid square equal to the median dispersal distance for all butterfly species (325 m) and for those species comprising the ecological attribute groups 1-3 (250 m) (see section 2.2.2). A border was applied in order to include habitat patches present within the periphery of the square that may influence butterfly presence.

The spatial location of hedgerows was incorporated within the connectivity analysis for the key PH1 2000 habitats, in order to consider their role in providing additional habitat as well as providing conduits for movement of butterfly species associated with woodland and grassland habitats. The shelter foot print provided by hedgerows for butterflies has been identified to extend layward to four times the height of the hedge (Dover, *et al.*, 2000; Lewis, 1969). Using data regarding hedgerow height available for the PH1 2010 data, the average height of 2.5 m was used as a surrogate for the 2000 hedgerow data set, and as such hedgerows were buffered by 5 m, providing a width of 10 m per hedgerow. Buffered hedgerows were then combined with the habitat ‘semi-natural broad-leaved woodland’ (PH-1), and with the

grassland habitats ‘unimproved neutral grassland’ (PH-15), ‘semi-improved neutral grassland’ (PH-16) and ‘calcareous grassland’ (PH-16).

Method 1

The binary connectivity metric, Integral Index of Connectivity (IIC), and the probabilistic metric, Probability of Connectivity (PC) (Pascual-Hortal and Saura, 2006; Saura and Pascual-Hortal, 2007), were computed for each key land cover class within each 1 km grid square using the median butterfly dispersal distances of 325 m and 250 m (see section 2.2.2) as thresholds for determining the network of connected patches (node component). For the LCM 2000 a total of 24 grid square connectivity metrics were computed (12 IIC metrics and 12 PC metrics) and for the PH1 2000 a total of 42 grid square connectivity metrics were computed (21 IIC metrics and 21 PC metrics).

Landscape connectivity variables were considered in addition to the compositional variables during the development of landscape connectivity models. The IIC and PC metrics exhibit weak correlations with total area for each key land cover class. However, metrics IIC and PC are strongly correlated at the 1 km scale, therefore only the metric IIC was used in further analyses (Eq1), as this is the simplest metric available. For the LCM 2000 a total of 12 IIC metrics were considered during the modeling procedure (Table 4.6) and for the PH1 2000 a total of 21 IIC metrics were considered during the modeling procedure (Table 4.18).

Eq 1

$$IIC = \frac{IIC_{num}}{A_L^2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{a_i * a_j}{1 + nl_{ij}}\right)}{A_L^2}$$

Where:

a_i and a_j = area of pacthes i and j,

nl_{ij} = number of connections in the shortest path between patches i and j

A_L^2 = total class area

See Pascual-Hortal & Saura (2006)

Method 2

In addition to computing IIC for each grid square for each key LCM/ PH1 land cover class, the importance of each patch for maintaining connectivity across the whole landscape of Warwickshire was calculated. Patch importance is obtained from calculating the change in IIC value with the removal of each patch in the landscape in turn (varIIC – Eq2) (Saura and Rubio, 2010). The metric varIIC was calculated using the distance thresholds of 325 m and 250 m to determine the network of connected patches for each key land cover class (see section 2.2.2). The importance of the nodes located within each 1 km grid square, as captured by varIIC, was then summed for each square providing a complementary assessment of habitat connectivity at the 1 km scale.

Eq 2

$$varIIC = \frac{IIC - IIC_{after}}{A_L^2}$$

Where:

A_L^2 = total class area

See Saura & Rubio (2010)

The varIIC metric can be partitioned into three components that contribute towards the overall connectivity value obtained (Saura and Rubio, 2010). The components of varIIC are (1) connectivity (varIICconn), which measures the importance of patches for maintaining overall connectivity (2) intra-connectivity (varIICintra) which considers connectivity within the patches and (3) dispersal flux (varIICflux) which measures how well connected a patch is to other patches (Saura and Rubio, 2010). The three components were computed for each key land cover class in addition to the overall varIIC metric.

For the LCM 2000 dataset a total of 48 metrics (12 classes x 4 metrics) were computed which measure the importance of patches in that landscape for maintaining connectivity across Warwickshire. Several of these metrics exhibited strong correlations (absolute $S \geq 0.8$) with the area of the corresponding land cover class, resulting in a final selection of only six additional varIIC metrics to be considered during the modelling procedure (Table 4.6). For the PH1 2000 a total of 84 metrics

(21 classes x 4 metrics) were computed which measure the importance of patches in that landscape for maintaining connectivity across Warwickshire. The majority of these 84 metrics exhibited strong correlations (absolute $S \geq 0.8$) with the area of the corresponding habitat, resulting in a final selection of only two additional varIIC metrics to be considered during the modelling procedure (Table 4.18).

Landscape Structure

Using the software FRAGSTATS (version 4) (McGarigal, *et al.*, 2012) landscape structure metrics were calculated for each grid square landscape using land cover data derived from LCM 2000 and PH1 2000. A total of 69 metrics were initially computed which measured the landscape aspects area/ edge, shape, aggregation, and diversity (see section 1.6.2). The computation of landscape structure metrics followed the same procedure outlined in section 2.1.3.

Pearson Product Moment Correlations were used to assess the relationship between all landscape structure metrics, and pairs of metrics with an absolute correlation coefficient greater than $r \geq 0.8$ were investigated further (0.01 % significance level). Representative metrics of each significant pairwise correlation combination were chosen, with the simplest metric of the two chosen. A high correlation coefficient was chosen as in contrast to chapter 3 the number of data points was not a limiting factor (section 2.1.3) and as such a wider range of metrics could be considered. For all grid squares within the LCM 2000 dataset (n =2427) a total of 38 landscape structure metrics remained after removal of highly correlated metrics, and for the occupied squares used for the EAG models (n = 515) a total of 36 metrics remained (41 different metrics in total) (Table 4.10). For the PH1 2000, a total of 35 landscape structure metrics were selected from the correlation analysis for all grid squares (n = 2079) and 37 metrics were selected from the correlation analysis for the occupied squares (n = 466) (39 different metrics in total) (Table 4.22).

2.2.5 Model development; model type

Logistic regression analysis (GLM), assuming a binomial distribution and a logit link function, was used to model the presence-absence of all butterfly species and species comprising the four ecological attribute groups (EAG) as a function of landscape composition, landscape connectivity and landscape structure, resulting in a total of

15 separate models. Logistic regression was used to identify relationships between species distribution and the key explanatory variables (variables selected in the final models), as well as obtaining the predicted probability of butterfly presence within a grid square as a function of the key landscape variables.

To select the best model from a set of independent variables, forwards selection and backwards elimination procedures were conducted, with selection/ elimination thresholds ranging from 1 (default) to 3.86, based on X^2 distribution for 1 degree of freedom at the 5 % significance level. With each increase in the selection/elimination threshold deviance tests were conducted in order to identify significant changes in residual deviance with the loss of independent variables. Variables were retained within the model if a significant increase in the residual deviance occurred when the variable was dropped.

2.2.6 Model development: Assessment of model performance

Model significance was determined using the regression deviance and the strength of the associations assessed using a t-test, to determine whether the slope parameter for each variable was significantly different from zero. The variance inflation factor (VIF) was calculated for each variable in the model to check for co-linearity between independent variables. Model goodness of fit was assessed with the Hosmer-Lemeshow (H-L) test statistic, which identifies sub groups within the data set and assesses the degree to which observed values match model-predicted values. Good model fits are associated with non-significant H-L test statistic values (p >0.05), which indicate no evidence for differences between observed and model predicted values. The discriminating ability of the model was assessed by calculating the proportion of correctly predicted presence (sensitivity) and absence (specificity) squares by the generation of a confusion matrix (Allouche, *et al.*, 2006). Thresholds for determining presence-absence were equal to the prevalence of the data (proportion of presence within the data set) (Lobo, *et al.*, 2008). The Receiver Operating Characteristic (ROC) area under the curve (AUC) was calculated for each model, with significant AUC values (p <0.05) indicating that the discriminating ability of the model is significantly better than that obtained by chance (AUC > 0.5). AUC values were also assessed against the criteria: excellent AUC ≥ 0.90; good 0.80

≥ AUC < 0.90; fair 0.70 ≥ AUC < 0.80; poor 0.60 ≥ AUC < 0.70; and fail 0.50 ≥ AUC < 0.60 (Araujo, *et al.*, 2005).

2.2.7 Combining models of landscape composition, connectivity and structure

For both the PH1 and LCM data sets the explanatory variables which measured the landscape components of composition, connectivity and structure were combined. Only those variables which were included in the individual landscape component models were considered during this stage of model development. During the development of the connectivity models, the connectivity metrics were considered in addition to the compositional variables, providing complementary assessments of composition. As such only variables included within the final landscape connectivity models (which include both connectivity and compositional variables) and structural models were considering during the forwards selection and backwards elimination procedure (section 2.2.5) to select the best-combined model (see sections 4.2.5 and 4.2.9).

2.3 Model validation (Chapter 5)

2.3.1 Landscape extent and landscape data

The combined landscape based models developed to predict presence-absence of butterfly species across Warwickshire (see section 2.2 and Chapter 4) were validated for the county of Warwickshire. The original models were developed using the landscape data sets LCM 2000 and PH1 2000 and butterfly data aggregated across 1990-1999 (training data). In order to validate these models, temporally independent landscape data was obtained (PH1 only). The Warwickshire Habitat Biodiversity Audit for the year 2010 (PH1 2010) was provided by Warwickshire County Council. The PH1 2010 map was rasterised to 25 m grain size (resolution) following the same procedure outlined in section 2.1.2 in order for consistency in the generation of the two raster data sets. The LCM 2000 (training data) was used during the model validation procedure because the classifications for landscape covers in the LCM 2007 differ from those in the LCM 2000. The PH1 2000 combined models (see section 2.2 and Chapter 4) were applied to the PH1 2010 data set in order to obtain the predictive values for each grid square based on the temporally different landscape data, to be used during model validation. The predictive values obtained from the LCM 2000 combined models were used during model validation.

Model validation was conducted considering two different spatial extents within Warwickshire. Firstly model validation was conducted considering all 1 km grid square landscapes, in addition to occupied grid square landscapes as determined from the butterfly records obtained from the UK BMS and the WBRC (section 5.2). Secondly, 19 sample sites (1 km grid squares) in Warwickshire with no previous records of butterflies between 1990 and 1999 were selected for collecting butterfly and vegetation data for model validation (Figure 2.4) (section 5.3). Sites were selected at a range of predictive probabilities across all eight models and as such ranged in their habitat suitability and landscape characteristics. Sample sites contrasting in predictive probabilities were arranged in three clusters in order to minimise travel distance and in turn ensure sites with contrasting suitability are visited on the same day, in accordance with the method used by Cozzi *et al.*, (2008). Data collected from the 19 sites were temporally independent from the training data

(collected 13 years previous) considering the short life span, and rapid generation of UK butterfly species (Kumar, *et al.*, 2009; Thomas, 2005). Additionally, the 19 sites were spatially independent from the combined landscape models built using the Ecological Attribute Group (EAG) data (occupied grid squares) (see section 2.2.2 and Chapter 4).

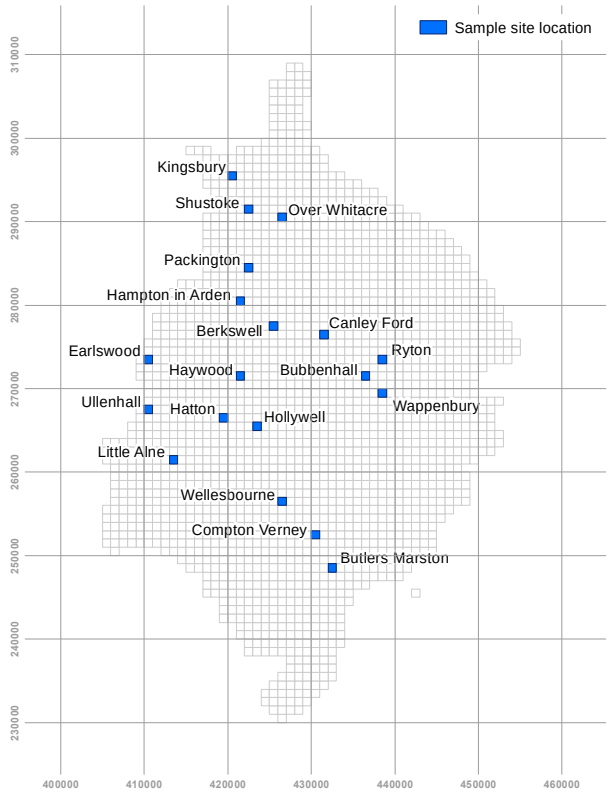


Figure 2.4: Location of the 19 sample (validation) sites across Warwickshire.

2.3.2 Butterfly data: Warwickshire 2000 - 2009

Butterfly data was obtained from the UKBMS and Warwick Biological Records Centre (WBRC) (Warwickshire County Council) for the years 2000 – 2009. Data was filtered following the same process outlined in section 2.2.2 to remove potentially unreliable and erroneous records.

Determining species presence-absence

Aggregated records for all species were used to classify the presence or absence of butterflies within each grid square. For the combined species data (all butterfly

species), a species record within a grid square was taken as ‘presence’ and no record within a square as ‘absence’ (LCM 2000 *n* = 2467; PH1 2010 *n* = 2080). Following the same procedure outlined in section 2.2.2 individual species were grouped into four Ecological Attribute Groups (EAGs) identified by Shreeve *et al.,* (2001) (see section 2.2.2). Within the 262 occupied squares (2000-2009), presence of species classified within a particular EAG was taken as a ‘presence’ for that group within the grid square and if a grid square did not comprise a record for any species within a particular EAG, this was taken to represent an ‘inferred absence’ for that EAG (*n* = 262). Presence-absence of all species and for each EAG group was used as observational data for model validation.

Butterfly community characteristics – Species richness, abundance and diversity

Using this data the species richness, total abundance and diversity was obtained for all species within a grid square and for all EAGs. Total abundance for all butterfly species was standardised by total number of surveying weeks per year. The diversity of butterfly species was calculated using the Simpsons Diversity Index (1-D). The Reciprocal Diversity Index (1/D) was also considered in order to distinguish between butterfly absent sites and those with only one species of butterfly recorded. Butterfly community measurements of standardised abundance, species richness and diversity were used in further analysis and log transformed to meet the assumptions of normality and homoscedasticity. When considering all butterfly species combined, the community measurements of standardised abundance, species richness and diversity (Simpsons and Reciprocal Diversity Index) of ‘all species’ could not be normalised thus non parametric tests were used.

2.3.3 Butterfly data: Sample sites

Butterfly transect sampling was conducted for the 19 sample sites between 22nd April 2013 and 8th September 2013 covering the entire flight period of the 33 species resident within Warwickshire (as identified in section 4.2). A total of eight visits were made to each site with two weeks between visits, with exception of June when there was three weeks between visits due to unsuitable weather conditions. Groups of sites were visited in random order during the sampling period.

Within the 19 sample sites transect routes were chosen to sample each habitat type present, and exact route position was dictated by terrain and access (following visual assessment of each sample site). Transect length varied between sites due to differences in habitat composition and the heterogeneous nature of some of the sites. Transect length averaged 2518 m ± 87 m, and each transect was split into sections representing each major habitat type sampled along the route (Pollard, 1977).

Sampling of butterflies was conducted along each transect route in accordance with the UKBMS method developed by Pollard (1977), for the quick monitoring of butterfly numbers on fixed sites (Thomas, 1983). The surveyor walked at a constant speed recording all adult butterflies which flew or were observed within 2.5 m each side of the transect route and 5 m ahead. Surveys were conducted between 10:45 and 15:45, during weather conditions suitable for butterfly activity: Temp >17 °C or 13 – 17 °C in sunny weather; Wind < 3 on the Beaufort scale; and no complete cloud cover. Butterflies were recorded and visually identified to species level in the field using identification keys and references of Hoffman and Marktanner (1995), Tolman (2004) and Riley (2007). Individuals of each species were counted, however, distinction between two species, Essex skipper (*Thymelicus lineola* Ochsenheimer) and small skipper (*Thymelicus sylvestris* Poda), during flight is difficult, with differences only occurring in the colour of the antenna tip (Cowley, 2000), and as such counts of these two species were grouped. Butterfly species and counts were recorded separately in each transect section and a six figure grid reference of the mid-point for each section was obtained using ArcGIS v10. Double counts are considered unlikely along transect routes due to the large area covered across the site and the route of the transects (Cozzi, *et al.*, 2008). Furthermore within each transect section butterflies which had been counted were kept track of to avoid double counting (Flick, *et al.*, 2012).

Species presence within a site was determined if at least one individual was recorded in one of the eight surveys (Cozzi, *et al.*, 2008). For each site three response variables were measured for all species and each species EAG; species abundance, species richness and species diversity. Species and EAG occurrence (presence and absence) was determined across all sites.

Species abundance was obtained by summing estimates of abundance across all eight surveys, this total was standardised by transect length to provide a measure of abundance per 100 m (density) using 100*N/L where N is the total abundance within a site and L is the transect length (Gross, *et al.*, 2007; Thomas, 1983). Butterfly species richness was the total number of butterfly species seen across all eight visits. The diversity of butterfly species within each site was calculated using the Shannon's Diversity Index in order to capture the influence of rare butterfly species observed during the sampling. Butterfly abundance (standardised abundance), species richness and diversity were used in further analysis and transformed to meet the assumptions of normality and homoscedasticity. Butterfly abundance data was square root transformed, species richness data was log transformed and species diversity data was squared transformed.

2.3.4 Vegetation data: Sample sites

Vegetation structure and flora surveys were conducted at each site, in spring (May) and in summer (August). The sampling methodology, adapted from the National Vegetation Classification (NVC) (Rodwell, 2006), consisted of a 4 m x 4 m survey of the ground layer (≤ 0.2 m), field layer (> 0.2 m, ≤ 2 m), and understorey layer (≤ 8 m), with sampling effort repeated four times within each transect section. Plant species present within each layer were visually identified to species level using identification keys, most notably Rose (1989; 2006), and Sterry (2008) (see Appendix A5 for species list). Samples of unidentified species in the field were taken for subsequent identification. The percentage cover of each species was recorded based on the Domin scale of cover/ abundance (Rodwell, 2006). The total Domin value for species within a vegetation layer can exceed 100 % due to the structural overlap of plants (Rodwell, 2006).

Within each vegetation layer the identified plant species were grouped into the following categories of vegetation type: bare ground, short herbs, short grass (ground layer); tall herb, tall grass, low shrub, ferns (field layer); tall shrub/ understorey tree (understorey layer) (Table 2.7). Classification of structural vegetation was in accordance with methods implemented by Dennis (2004). The percentage cover data for each vegetation type was averaged across the total number of samples within a site (Table 2.7). Species richness of each vegetation layer (field, ground and

understorey) was obtained as the total number of species within that layer across all samples for each site. Species diversity was obtained for each vegetation layer using the Simpsons and Shannon's Diversity Index. The two diversity indices were highly correlated and as such the Simpsons Diversity Index (1-D) of each vegetation layer was used in further analysis. This index places more weight on the most abundant species within the site, with little changes to the index occurring with the addition of rare species (Maurer and McGill, 2011). As such, it is applicable for representing the diversity of the overall composition of each layer, reflecting the sampling strategy of the vegetation.

Additionally plant species were classified into larval and nectar food plant groups, according to associations with the resident butterfly species identified within Warwickshire, and the total species richness and average percentage cover of each food plant group were determined.

For each vegetation layer/ type and food plant group the average percentage cover data, species richness and species diversity (Simpsons Index) were used in further analyses. Vegetation data was log transformed, with exception to the percentage cover data which was arc-sine transformed to meet the assumptions of normality and homoscedasticity. Pearson Product Moment Correlations (r) were calculated to identify strongly significant correlations (absolute $r \geq 0.8$) among the vegetation data considering percentage cover, species richness and diversity. The percentage cover and species richness of larval and nectar food plants correlated strongly with the vegetation composition of the field, and ground layers. As such, the species richness and percentage cover of larval and nectar food plants are not considered in further analysis and are represented by the species richness of the ground layer in spring, and the diversity of the ground layer and species richness of the field layer in summer.

Acronym	Metric	Range
Field layer		
FGRASS	Percentage of field layer consisting of grass species	0.0 – 46.2 %
FHERB	Percentage of field layer consisting of herb species	0.0 – 99.5 %
FSHRUB	Percentage of field layer consisting of low shrub species	0.9 – 54.2 %
F_SD	Simpsons Diversity Index for field layer species	0.1 – 1.0
F_SR	Species richness of field layer	2.0 – 32.0
FERN	Percentage of low lying fern species	0.0 – 54.9 %
Ground layer		
GBARE	Percentage of ground layer consisting of bare ground	0.6 – 35.9 %
GGRASS	Percentage of ground layer consisting of grass species	39.9 – 100.5 %
GHERB	Percentage of ground layer consisting of herb species	23.8 – 150.5 %
G_SD	Simpsons Diversity Index for ground layer species	0.6 – 1.0
G_SR	Species richness of ground layer	14.0 – 48.0
Understorey layer		
UNDER	Percentage of understorey layer consisting of tree and tall shrub species	4.3 – 68 %
U_SD	Simpsons Diversity Index for understorey layer species	0.5 – 0.9
U_SR	Species richness of understorey layer	2.0 – 15.0

Table 2.7: The vegetation variables capturing local habitat quality considered in further analysis. The acronym, metric name, range and units are provided for each metric. The ranges for the metrics are derived from the average values across all 19 sample sites. Metrics are unit less unless otherwise stated.

2.3.5 Model Validation

Using the predictive values from the PH1 2010 and LCM 2000 combined models, the ability of these models to accurately predict the Warwickshire butterfly data set 2000-2009 (see section 2.3.3) was assessed by calculating the proportion of correctly predicted presence (sensitivity) and absence (specificity) squares by the generation of a confusion matrix (Allouche, *et al.*, 2006). Thresholds for determining presence-absence were equal to the prevalence of the training data (proportion of presence within the data set) (see section 2.2.6 and Chapter 4). The Receiver Operating Characteristic (ROC) area under the curve (AUC) was calculated for each model, with significant AUC values ($p < 0.05$) indicating that the discriminating ability of the model is significantly better than that obtained by chance (AUC > 0.5). AUC values were also assessed against the criteria: excellent AUC ≥ 0.90; good 0.80 ≥

AUC < 0.90; fair 0.70 ≥ AUC < 0.80; poor 0.60 ≥ AUC < 0.70; fail 0.50 ≥ AUC < 0.60 (Araujo, *et al.*, 2005). The specificity, sensitivity and AUC associated with the PH1 2010 and LCM 2000 in relation to the Warwickshire butterfly data 2000-2009 were compared to the specificity, sensitivity and AUC obtained for the training data (PH1 2000 and LCM 2000).

2.3.6 Predictions of butterfly abundance, species richness and diversity

Predictive values obtained from the PH1 2010 and LCM 2000 models can be considered to provide an indication of suitability for supporting the corresponding butterfly species group. The relationship between the predictive values from each model and the butterfly community characteristics were assessed considering the two butterfly validation data sets (Warwickshire and 19 sample sites). Pearson Product Moment Correlations (r) were calculated to assess for relationships between grid square suitability (based on the LCM 2000 and PH1 2010 model predictive values), and the abundance, species richness and diversity for all species and species EAGs. For the 'all butterfly' species data from the Warwickshire data set, Spearman rank correlations (S) were conducted with the LCM 2000 and PH1 2010 model predictive values, due to the non-normality of the data.

Based on the distribution of the predicted values for the PH1 2010 and LCM 2000 combined models, grid squares were divided into four groups of suitability for supporting butterflies (low, low to medium, medium to high, high) based on the quartiles of the predictions produced by each model. By classifying each grid square into the four suitability groups according to each model, differences in the abundance, species richness and diversity between each suitability group were assessed using a One-Way Analysis of Variance (ANOVA). Non parametric Kruskal-Wallis H Tests were performed to asses for differences in median abundance, species richness and diversity of 'all butterfly' species between each suitability group, due to the non-normality of the data.

2.3.7 Relationship between butterfly observations and local habitat characteristics

Pearson Product Moment Correlations (r) were calculated to assess for relationships between butterfly abundance, species richness and diversity of the 19 sample sites and the local habitat characteristics (vegetation data) determined by the percentage

cover of each vegetation type, and the species richness and diversity of each vegetation layer.

2.3.8 Relationship between predicted values, butterfly community composition and local habitat characteristics

Nonmetric multidimensional scaling (NMDS), a method of multivariate ordination, was used to assess the extent to which the community structure of butterfly species differed amongst the 19 sample sites. The Bray-Curtis (B-C) measure of dissimilarity was used to determine distance values amongst sites based on butterfly abundance per species and this B-C dissimilarity matrix was used in the NMDS analysis. Using the B-C dissimilarity matrix, the variation in the butterfly community composition among habitat suitability groups (see section 2.3.7), was assessed by a One-Way ANOSIM (non-parametric analysis of similarities) using the software PRIMER v6 (Clarke, 1993). Hierarchical clustering based on the B-C dissimilarity matrix for butterfly species and the complete link method, was used to determine the clustering of sites on the basis of their butterfly community structure. Cluster groups were defined by the identification of the similarity threshold which maximised homogeneity of clusters (average similarity of the members) and separation of the clusters (average dissimilarity of each cluster to its nearest neighbour) (Jongman *et al.,* 1995).

The Bray-Curtis (B-C) measure of dissimilarity was also used to determine compositional distances between sites based on the transformed vegetation data for spring and summer (percentage cover of each vegetation type). NMDS analysis was then performed using the B-C dissimilarity matrix for spring and summer vegetation, to assess the extent to which the community structure of vegetation species differed amongst the 19 sample sites. Association between butterfly community composition and vegetation community composition for spring and summer was assed using Mantel Product Moment Correlation based on the B-C dissimilarity matrices.

Chapter 3: The influence of spatial scale on the discriminating ability of landscape structure metrics

3.1 Introduction

Landscape structure metrics are widely used for developing relationships with measures of biodiversity, for assessing landscape change and for comparing landscapes under different management strategies (Saura, *et al.*, 2008; Simova and Gdulova, 2012; Turner, *et al.*, 2001). Despite this the extensive use of metrics has been criticised, with use considered inappropriate or inaccurate in several studies (Li and Wu, 2004; Peng, *et al.*, 2010). Such a 'misuse' of landscape structure metrics has been attributed to a lack of understanding of metric calculation and associated limitations, which has arisen due to the increased availability of landscape pattern calculation software and the sheer number of available metrics (Li, *et al.*, 2005; Peng, *et al.*, 2010; Turner, 2005).

Numerous landscape structure metrics have been developed for the quantification of landscape pattern (Fortin, *et al.*, 2003; HainesYoung and Chopping, 1996; McGarigal, *et al.*, 2012). In most cases, individual landscape structure metrics describe either the composition or the configuration of the landscape, and not both (Simova and Gdulova, 2012). For example, the most widely used measurement of landscape heterogeneity is the quantification of structural diversity which considers only the composition of the landscape, through the use of the Shannon's diversity index (Fjellstad, *et al.*, 2001; Tews, *et al.*, 2004). As such, this index provides similar values of heterogeneity for landscapes despite having different configurations of land cover (Fortin, *et al.*, 2003; Li and Wu, 2004). It is argued that indices which combine multiple components of spatial pattern into a single value are difficult to interpret (Gustafson, 1998; Li and Wu, 2004), consequently several metrics are often required to capture the various aspects of landscape pattern, and in turn landscape composition and configuration (Li, *et al.*, 2005). Landscape pattern aspects summarised using these metrics include: area/edge, aggregation, diversity, and contrast (see section 1.6.2). The use of several landscape structure metrics ensures quantification of the different components of landscape structure, however many indices are redundant as they quantify the same information in different ways, e.g. total edge and edge density (when comparing landscapes of the same size) (McGarigal and Marks, 1995). Several metrics are also statistically correlated and, as such, are empirically redundant (Riitters, *et al.*, 1995; Schindler, *et al.*, 2008). Strong

correlation occurs between several landscape level metrics because they are based on the same variability in patch attributes that operate over the landscape scale (McGarigal and Marks, 1995).

In addition to the choice of metric to measure landscape pattern, another factor to be considered is spatial scale which influences metric behaviour (Baldwin, *et al.*, 2004; Schindler, *et al.*, 2008; Tischendorf, 2001). Spatial scale may refer to the extent (landscape size), grain size (pixel size) and thematic resolution of the map (classification level) (Simova and Gdulova, 2012). The calculation of landscape metrics via a moving window analysis provides another assessment of scale due to the size of the window that is specified by the user (Gaucherel, 2007; McGarigal and Marks, 1995). Many studies fail to identify the appropriate scale for the analysis of landscape pattern (Gustafson, 1998), and as such inaccurate inferences may be drawn when comparing landscapes (Simova and Gdulova, 2012).

Studies have investigated the effects of spatial scale on the behaviour of landscape structure metrics, yet often yield contrasting results and, to date, not all metrics have been studied (Baldwin, *et al.*, 2004; Li and Wu, 2004; Simova and Gdulova, 2012). Contrasting results have been obtained mainly due to differences in the classification of landscapes being used (i.e. land cover classification) (Turner, *et al.*, 2001). Identification of the most appropriate scale for capturing landscape patterns should therefore consider the effects of scale on a collection of landscapes derived from differing land cover classifications. Furthermore, metrics respond differently to changes in spatial scale depending on what attributes of the landscape they measure, adding further complexity to scale choice. Most notably, Wu *et al.*, (2004; 2002) investigated the effect of grain size and extent on metric behaviour and identified predictable responses for some metrics, particularly in response to changes in grain size. Metrics were grouped into three categories based on their responses to changes in extent and grain size: type I metrics show predictable scaling relations, type II metrics show stepwise scaling relations and type III metrics show no predictable response (Wu, 2004; Wu, *et al.*, 2002). Although these studies have identified the different responses of metrics to changes in scale, they have failed to consider how the discriminating ability of metrics changes with scale between different landscapes

or landscape types. The ability to discriminate between landscapes, particularly landscape types which differ in their intrinsic character (i.e. National Character Areas), is particularly important when using landscape structure metrics to aid landscape planning or to develop relationships with biodiversity (Garcia-Feced, *et al.*, 2010). In an attempt to bridge this gap, Garcia-Feced *et al.*, (2010) investigated whether the discriminating ability of metrics was consistent across scales, yet this considered only a limited number of metrics (eight) and only forested Mediterranean landscape types. To date no study has investigated the effect of scale on the discriminating ability of metrics between UK landscape types, using landscape data applicable to the planning and decision making made in the UK.

3.1.1 Aims

The aims of the work reported in this chapter are to: -

1. Identify metrics that discriminate between selected landscapes, and determine whether changing the grain size (resolution) impacts the ability of those metrics to discriminate between different landscapes at a national (NCA) and county (Warwickshire 1 km grid squares) scale.
2. Identify the best scale at which data needs to be collected for characterising landscapes and discriminating between them.
3. Compare the effect of spatial scale on landscape structural metrics between two different landscape data sources (LCM 2000 and PH1 2000) with different thematic resolution (land cover classification).

Work on these three aims will test the hypothesis that characterisation of landscape pattern and discrimination between landscapes by landscape structure metrics is consistent across scales.

3.2 Results: National Character Areas (NCAs – national scale)

3.2.1 Classification of landscape characteristics: 25 m

Total Area (TA)

When comparing the average values of the 32 landscape structure metrics between the four groups of NCAs classified by Total Area (TA; small, medium, large, and extra-large) significant differences were obtained for three metrics; LSI ($F_{3,28}$ = 33.244, p <0.001), CIRCLE_MN ($F_{3,28}$ = 3.143, p = 0.041) and PRD ($F_{3,28}$ = 60.935, p <0.001) (Table 3.1), thus showing 29 metrics are not impacted by TA. For the metrics LSI and PRD significant differences were obtained between each of the four TA groups (p<0.05), whereas the average metric value of CIRCLE_MN only differed significantly between the extremes of TA, small and extra-large NCAs (p = 0.033).

Mean Patch Size (MPS)

When assessing the effect of NCAs classified by MPS (small, medium, large, and extra-large), over half the metrics (18) had significantly different results when analysed using a one-way ANOVA (Table 3.1). Despite the large number of metrics differing in accordance to MPS, only one metric CONTIG_AM differed significantly between all the four groups of NCAs (p<0.05), other metrics differed between small and medium MPS in comparison to large and extra-large MPS.

Number of Land Cover classes (NLC)

The NLC were classified as high (17 - 23) and low (3 – 16). When analysing the impact between NCAs classified as high or low by NLC on metric output using a one-way ANOVA, four metrics (CIRCLE_RA, ENN_AM, GYRATE_MN and IJI) had significantly different (p < 0.05) outputs for NCAs with contrasting number of land classes (high and low) (Table 3.1).

Diversity of Land Cover (DLC)

The DLC groups ordered from one to nine reflect a reduction in diversity of different land covers for the NCAs in each group (Table 2.2a). When comparing metric values between NCAs grouped by DLC, significant differences were obtained for 21 of the

metrics (Table 3.1). For most metrics significant differences were obtained between the DLC groups 1, 2 and 3 in comparison to the groups 8 and 9 reflecting differences between the extremes of NCAs in terms of landscape diversity.

3.2.2 Discriminating ability of metrics with scale: 25 m – 1000 m

At a resolution of 25 m 16 metrics discriminated between NCAs on the basis of both MPS and DLC as revealed by significant differences in metric values between corresponding landscape characteristic groups (section 3.2.1; Table 3.2). Of these 16 metrics, eight maintained this discriminating ability across all scales from the highest resolution of 25 m to the coarsest resolution of 1000 m. The discriminating ability of the remaining metrics ceases at a variety of scales (as shown by non-significant differences between metric outputs), in particular, on the basis of MPS two metrics (GYRATE_MN and PROX_AM) are non-significant at scales equal to and larger than 50 m and the metric SIMI_CV is non-significant at scales larger than 100 m (Table 3.2).

For DLC, the lowest resolution at which metrics cannot discriminate between NCAs is at 100 m, with non-significant differences in metric output obtained for the metrics CIRCLE_AM and ENN_MN between the nine DLC groups of NCAs (Table 3.2).

Of the three metrics which discriminated between NCAs on the basis of TA, two of these (LSI and PRD) maintained this discriminating ability across all scales from 25 m to 1000 m. In contrast the metric CIRCLE_MN was unable to discriminate between NCAs based on TA at 50 m upwards, with non-significant differences in metric output between the four groups of NCAs grouped by TA.

The four metrics which discriminated between NCAs grouped by NLC, were unable to maintain this discriminating ability across all scales. The metric IJI discriminated between NCAs until 500 m, however non-significant differences were obtained in the output of GYRATE_MN between NCAs grouped by NLC at 50 m upwards and for CIRCLE_RA and ENN_AM from 100 m upwards.

Metric	Total Area (TA)		Mean Patch Size (MPS)		Number of Land Cover Classes (NLC)		Dominant Land Cover (DLC)	
	$F_{3,28}$	P	$F_{3,28}$	P	$F_{1,30}$	P	$F_{8,23}$	P
LSI	33.244	0.000	1.363	0.275	0.011	0.916	1.818	0.125
AREA_MN	0.087	0.967	74.278	0.000	0.137	0.714	4.997	0.001
AREA_RA	0.697	0.561	3.969	0.018	1.939	0.174	3.036	0.018
GYRATE_MN	1.010	0.403	3.691	0.023	3.348	0.077	5.847	0.000
GYRATE_AM	0.363	0.780	6.943	0.001	1.532	0.225	5.219	0.001
GYRATE_CV	0.956	0.427	9.583	0.000	1.438	0.240	5.878	0.000
SHAPE_MN	1.375	0.271	0.685	0.569	5.619	0.024	4.943	0.001
SHAPE_CV	1.070	0.378	5.336	0.005	1.248	0.273	4.943	0.001
FRAC_AM	0.595	0.624	5.553	0.004	1.421	0.243	4.993	0.001
FRAC_CV	1.076	0.375	1.044	0.388	1.898	0.178	1.471	0.222
CIRCLE_MN	3.143	0.041	1.652	0.200	0.287	0.596	2.514	0.040
CIRCLE_AM	0.174	0.913	6.803	0.001	2.505	0.124	3.305	0.012
CIRCLE_RA	1.00	0.407	0.58	0.982	3.024	0.092	0.687	0.699
CONTIG_MN	2.685	0.066	0.121	0.947	8.856	0.006	1.045	0.432
CONTIG_AM	0.230	0.875	80.776	0.00	0.093	0.762	5.029	0.001
CONTIG_RA	0.627	0.604	6.020	0.003	0.061	0.807	1.480	0.218
PROX_AM	0.342	0.795	3.018	0.046	0.589	0.449	3.953	0.005
PROX_CV	0.484	0.696	0.410	0.747	1.552	0.222	1.103	0.397
SIMI_AM	0.905	0.451	1.919	0.149	0.681	0.416	4.750	0.002
SIMI_CV	0.415	0.744	3.970	0.018	2.833	0.103	1.506	0.209
ENN_MN	0.406	0.750	16.525	0.000	3.491	0.072	2.975	0.019
ENN_AM	0.580	0.633	0.347	0.792	0.482	0.493	1.624	0.172
ENN_CV	1.683	0.193	0.541	0.658	3.230	0.082	0.986	0.472
CWED	0.364	0.780	28.965	0.000	0.852	0.363	5.266	0.001
ECON_AM	0.514	0.676	1.256	0.308	2.723	0.109	2.777	0.026
ECON_CV	0.949	0.430	1.499	0.236	4.655	0.039	1.323	0.282
CONTAG	0.144	0.933	6.261	0.002	4.870	0.035	6.140	0.000
IJI	0.095	0.962	2.048	0.130	7.775	0.009	3.794	0.006
COHESION	0.430	0.733	6.669	0.002	1.247	0.273	6.768	0.000
MESH	0.330	0.803	5.932	0.003	1.747	0.196	4.705	0.002
PRD	60.935	0.000	0.245	0.864	1.147	0.293	0.854	0.567
SIDI	0.999	0.408	4.015	0.017	0.269	0.608	11.009	0.000

Table 3.1: Differences in metric output between NCAs classified by Total Area (TA), Mean Patch Size (MPS), Number of Land Cover classes (NLC) and Diversity of Land Cover (DLC). Differences are assessed by means of a One-Way ANOVA with test statistic (F), degrees of freedom and significance value (P) provided

	25 m				50 m				100 m				250 m				500 m				1000 m			
Metric	A	M	N	D	A	M	N	D	A	M	N	D	A	M	N	D	A	M	N	D	A	M	N	D
AREA_MN		*		*		*		*		*		*		*		*		*		*		*		*
AREA_RA		*		*		*		*		*		*		*		*				*				*
CIRCLE_AM		*		*		*		*		*				*				*		*		*		
CIRCLE_MN	*			*				*				*				*				*				*
CIRCLE_RA			*				*		*											*				
COHESION		*		*		*		*		*		*		*		*		*		*		*		*
CONTAG		*		*		*		*		*		*		*		*		*		*		*		*
CONTIG_AM		*		*		*		*		*		*				*		*				*		*
CONTIG_MN												*				*								*
CONTIG_RA		*				*								*				*				*	*	
CWED		*		*		*		*		*		*		*		*		*		*		*		*
ECON_AM				*				*								*				*				*
ECON_CV																								*
ENN_AM			*				*																	
ENN_CV															*									
ENN_MN		*		*		*		*		*				*										
FRAC_AM		*		*		*		*		*		*		*		*		*		*		*		*
FRAC_CV								*				*				*				*				
GYRATE_AM		*		*		*		*		*		*		*		*		*		*		*		*
GYRATE_CV		*		*		*		*		*		*		*		*		*		*		*		*
GYRATE_MN		*	*	*				*				*				*				*			*	*
IJI			*	*			*	*			*	*			*	*			*	*		*		*
LSI	*				*				*				*				*				*			
MESH		*		*		*		*		*		*		*		*		*		*		*		*
PRD	*				*				*				*				*				*			
PROX_AM		*		*				*				*												
PROX_CV																*				*				*
SHAPE_CV		*		*		*		*		*		*		*		*		*		*				*
SHAPE_MN				*				*				*				*				*				
SIDI		*		*		*		*		*		*		*		*		*						*
SIMI_AM				*				*				*				*				*				*
SIMI_CV		*				*									*				*					

Table 3.2: Discriminating ability of metrics with increasing grain size from a resolution of 25 m to 1000 m, based on significant differences in average metric output between NCAs grouped by Total Area (A), Mean Patch Size (M), Number of land cover classes (N) and Diversity of Land Cover (D). Results are based on a One-Way Analysis of Variance, and * indicate significant differences ($p<0.05$).

3.2.3 Discriminating ability of metrics combined: 25 m

At a 25 m grain size the first four Principal Components (PC) of the Principal Component Analysis (PCA) explain 72.1 % of the variation in the 32 landscape metrics amongst the NCAs (Table 3.3). Component 1 (PC-1) comprises 18 metrics with high loadings which are associated with the landscape aspects area, shape, aggregation, contrast and diversity (Table 3.3). The 18 metrics when combined provide measures of patch dispersion, interspersion and subdivision and in turn represent landscape continuity or landscape fragmentation (Table 2.3; Appendix A4).

NCAs can be separated on PC-1 by the landscape characteristics MPS and DLC (Figures 3.1a and 3.1b). NCAs characterised by small and medium MPS can be distinguished from those characterised by large and extra-large MPS on PC-1 (Figure 3.1a). Additionally on PC-1 NCAs characterised by the DLC groups with high landscape diversity (DLC groups 1 and 2) can be distinguished from those NCAs characterised by lower landscape diversity (DLC groups 7, 8, and 9) (Figure 3.1b). Of the 18 metrics associated with PC-1, 15 metrics discriminated between NCAs based on the classifiers of MPS and DLC, as indicated by significant differences in metric output between corresponding landscape classification groups (Table 3.1).

Principal Component 2 (PC-2) comprised four metrics with high loadings (Table 3.3) which measure patch shape and extent. PC-3 comprised five metrics associated with patch boundary configuration and PC-4 comprised five metrics associated with patch shape irregularity (Table 3.3; Appendix A4).

Metrics with the highest loadings on the PCs contribute towards discrimination between NCAs. When considering the 2D configuration of the scores, the NCAs 96 and 42 (Figure 3.1a), provide an example of extreme discrimination on PC-1 with a high negative score for NCA 96 (score = - 4) on PC-1 and a high positive score for NCA 42 (score >8) on PC-1, despite similar scores on PC-2 (NCA 96 score = 1.5; NCA 42 score = 1.9) (see Figure 3.1a). Differences in patch size, aggregation and heterogeneity are evident from visual comparison of the composition and configuration of the two NCAs (Figures 3.2a,b; Table 3.4). Discrimination between NCA 96 and 42 on PC-1 by landscape structure metrics is most pronounced for four metrics; GYRATE_AM, AREA_RA, CWED and IJI (Figure 3.2c; Table 3.4). NCA

42 is characterised by extra-large MPS, with a significantly higher area-weighted mean patch extent (GYRATE_AM) ($F_{3,28}$ = 6.943, p<0.001), and range in patch size (AREA_RA) ($F_{3,28}$ = 3.969, p=0.018) in comparison to NCA 96, which is characterised by medium MPS (Figures 3.2a,b; Figure 3.1a; Table 3.1; Table 3.4). NCA 42 and 96 are also characterised within different DLC groups, NCA 42 in DLC-7 (low landscape diversity) and NCA 96 in DLC-1 (high landscape diversity) (Figure 3.1b). When considering the DLC groups, NCA 96 was characterised by significantly higher degree of edge contrast between neighbouring patches (CWED) ($F_{8,23}$ = 5.266, p<0.001), and interspersion of patch types (IJI) ($F_{8,23}$ = 3.794, p=0.006) (Figures 3.2a,b; Table 3.1; Table 3.4). Despite these dissimilarities, NCAs 96 and 42 are similar on PC-2 in terms of the landscape structure metrics associated with patch extent and shape; GYRATE_MN, SHAPE_MN, CIRCLE_MN and CONTIG_RA (Figure 3.2a,b; Table 3.4).

Landscape metric	PC1	PC2	PC3	PC4	Landscape metric *(cont.)*	PC1	PC2	PC3	PC4
Area					**Shape**				
AREA_MN	.791	.430	.207	-.299	CIRCLE_MN	-.021	.799	.185	.223
AREA_RA	.842	-.203	.340	.249	CIRCLE_AM	.865	.004	-.085	.249
GYRATE_MN	.033	.846	.323	-.319	CIRCLE_RA	.048	.224	-.067	.627
GYRATE_AM	.923	-.174	.218	.057	CONTIG_MN	-.114	.203	.762	-.459
GYRATE_CV	.962	-.138	-.050	-.057	CONTIG_AM	.863	.350	-.005	-.249
					CONTIG_RA	.219	.692	.073	.208
Aggregation					FRAC_AM	.932	-.278	.108	.026
COHESION	.917	-.234	.062	.007	FRAC_CV	.171	.448	-.567	.314
CONTAG	.888	.023	-.184	-.113	SHAPE_MN	-.233	.806	.381	-.001
ENN_MN	.792	.268	-.271	-.120	SHAPE_CV	.855	-.163	.019	-.083
ENN_AM	-.435	.260	-.357	-.079					
ENN_CV	-.021	.292	-.183	-.054	**Contrast**				
IJI	-.625	-.066	.164	.125	ECON_AM	-.262	-.442	.249	-.289
LSI	-.405	.019	.629	.447	ECON_CV	.247	.411	-.500	.419
MESH	.869	-.154	.274	.082	CWED	-.812	-.478	.139	.074
PROX_AM	.592	-.304	.311	.134					
PROX_CV	.306	-.251	.174	.688	**Diversity**				
SIMI_AM	.533	.028	.259	.115	PRD	-.022	-.246	-.828	-.266
SIMI_CV	-.615	-.323	.170	.146	SIDI	-.871	.217	.109	.156

Table 3.3: The loadings of 32 landscape metrics on the principal components, grouped by landscape aspect. Note: Variables with the highest loadings on each component have been highlighted.

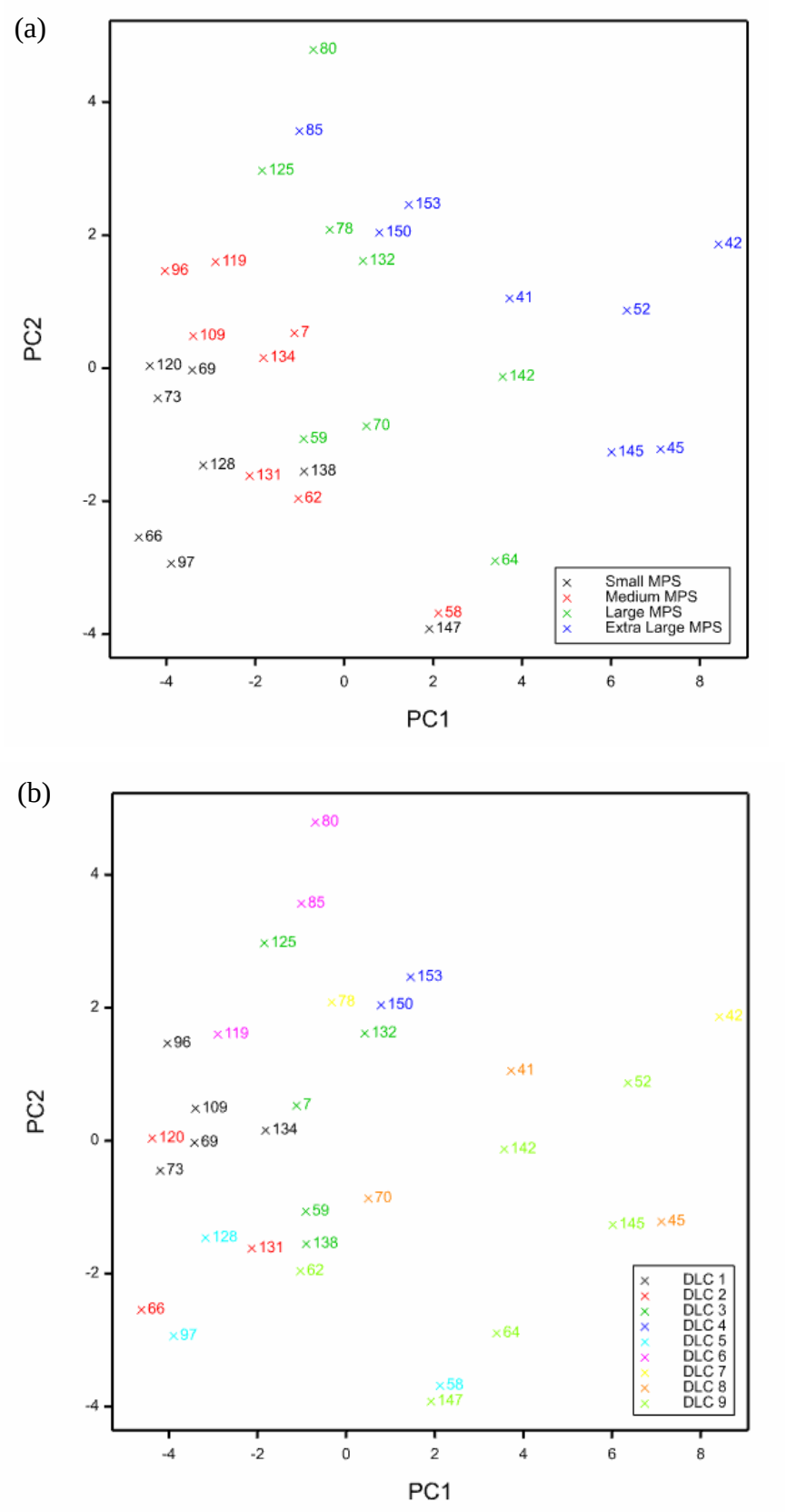


Figure 3.1: Comparison of the first two principal components (PC-1 and PC-2) of the configuration of the scores from the Principal Component Analysis (PCA), grouped by (a) Mean Patch Size (MPS) and (b) Diversity of Land Cover (DLC). Labels represent NCA ID number.

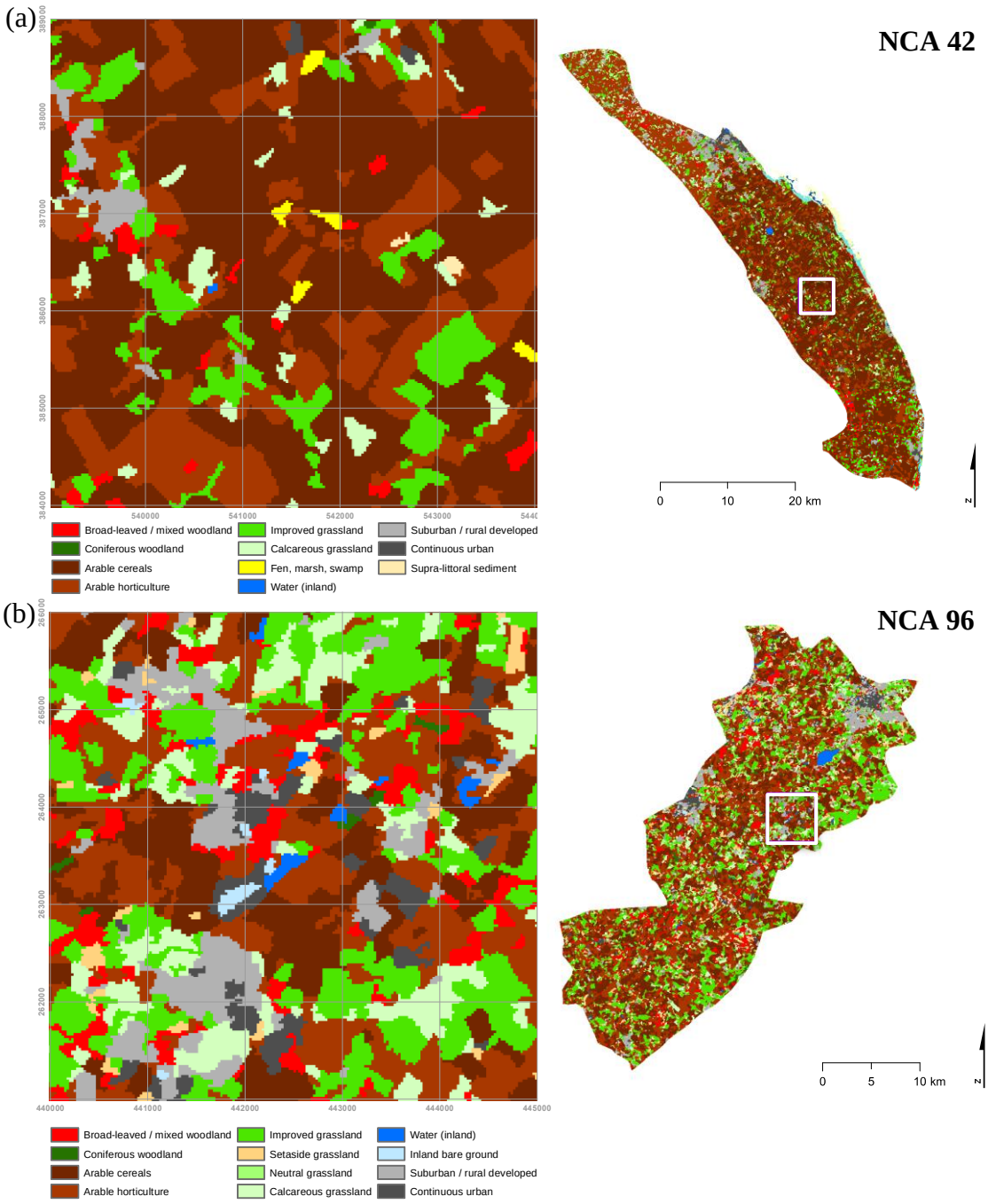


Figure 3.2: Landscape composition and configuration of two of the 32 NCAs; (a) NCA 42 and (b) NCA 96. Landscape composition and configuration is derived from the LCM 2000. Extent indicators (white) refer to the location of the 5 km x 5 km sections of landscape shown within the corresponding NCA.

	PC-1			PC-2	
Metric	NCA 96	NCA 42	Metric	NCA 96	NCA 42
GYRATE_AM	367.54 m	3781.13 m	GYRATE_MN	102.82 m	115.49 m
AREA_RA	645.88 ha	26292.19 ha	SHAPE_MN	1.59	1.59
CWED	74.86 %	45.06 %	CIRCLE_MN	0.60	0.60
IJI	75.15 %	53.43 %	CONTIG_RA	0.97	0.96

Table 3.4: Selection of landscape structure metrics with high loadings on PC-1 and PC-2 derived from the composition and configuration of NCAs 42 and 96 (LCM 2000).

3.2.4 Discriminating between NCAs with scale: 25 m – 1000 m

When comparing the 4-dimension (4-D) configuration of the scores from the PCA at 25 m (section 3.2.3) with that obtained at the different grain sizes (50 m, 100 m, 250 m, 500 m and 1000 m) by means of a Procrustes Rotation, the Residual Sums of Squares (RSS) between the pairs of configurations increases as the difference in grain sizes being compared increases (Table 3.5). As such, the RSS are greatest when comparing the 4-D configuration of each scale to 25 m (Table 3.5), and RSS are lowest when comparing the 4-D configuration obtained at 25 m to 50 m, 50 m to 100 m, 100 m to 250 m, 250 m to 500 m, and 500 m to 1000 m (Table 3.5). Relative similarity between NCAs is greatest when comparing the 4-D configuration obtained at a scale of 50 m to that obtained at a scale of 100 m, and when comparing the 4-D configuration obtained at a scale of 25 m to that obtained at a scale of 50 m, as indicated by the lowest RSS (Table 3.5).

When comparing the 4-D configurations from scales 50 m – 1000 m to the 4-D configuration obtained at 25 m, a number of NCAs are consistently dissimilar relative to that NCA at a scale of 25 m, with high Procrustes residuals obtained (Appendix A6). The projected Procrustes residuals (distance) between NCA scores for the first two PCs plotted at different scales in Figures 3.3a-e highlight the separation between NCAs 42 and 41 at each scale comparison. There are several NCAs with small residuals (obtained from the Procrustes Rotation) that are similar to each other relative to the other NCAs, when comparing 25 m to 50 m, and 100 m (Figures 3.3a,b). NCAs with small residuals at these scale comparisons include NCAs 66, 70, 78, 80, 85, 97, 120, 125 and 132 (Figures 3.3a,b – as indicated by the orange groups). With exception to NCA 132, the relative similarity between these NCAs decreases (with larger residuals) when comparing the configurations from the scale 25 m to the configurations obtained at scales 250 m upwards (Figures 3.3c-e).

Despite the relative dissimilarity between individual NCAs at smaller scales (25 m, 50 m and 100 m) compared to larger scales (250 m, 500 m and 1000 m), similarities between NCAs (obtained from the 4D output from the PCA at each scale) are significantly correlated with each other at each scale combination (Table 3.6). Pearson Product Moment Correlation coefficients decrease, however, when comparing the Euclidean similarity matrices between NCAs obtained at 25 m to 250

m upwards (Table 3.6). This pattern is similar to that obtained when considering the RSS from the Procrustes Residuals (Table 3.5), with highest similarity occurring between 50 m and 100 m (r = 0.9386, p <0.001) (Table 3.6).

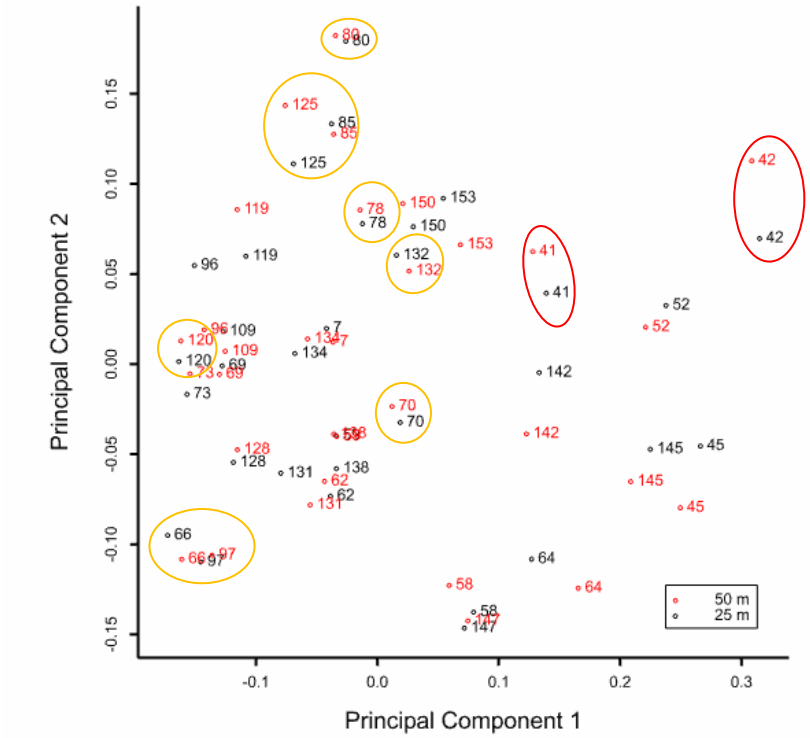
25	-					
50	0.0588	-				
100	0.1114	0.0453	-			
250	0.1668	0.1469	0.1198	-		
500	0.2240	0.1974	0.1839	0.1156	-	
1000	0.3453	0.3359	0.3246	0.2656	0.2108	-
	25	**50**	**100**	**250**	**500**	**1000**

Table 3.5: Residual Sums of Squares (RSS) from the Procrustes Rotation comparison of the 4-dimension configuration of the scores from the Principal Component Analysis for each scale.

25	-					
50	0.8557	-				
100	0.7863	0.9386	-			
250	0.7126	0.7631	0.8055	-		
500	0.5975	0.6194	0.6826	0.8640	-	
1000	0.4681	0.4825	0.5512	0.6635	0.6965	-
	25	**50**	**100**	**250**	**500**	**1000**

Table 3.6: Pearson product-moment correlations between the Euclidean similarity matrices derived from the 4-dimension configuration of the scores from the Principal Component Analysis for each scale. Pearson product-moment correlations are obtained by Mantel Tests for each pairwise comparison and all correlations are significant (p<0.001).

(a)



(b)

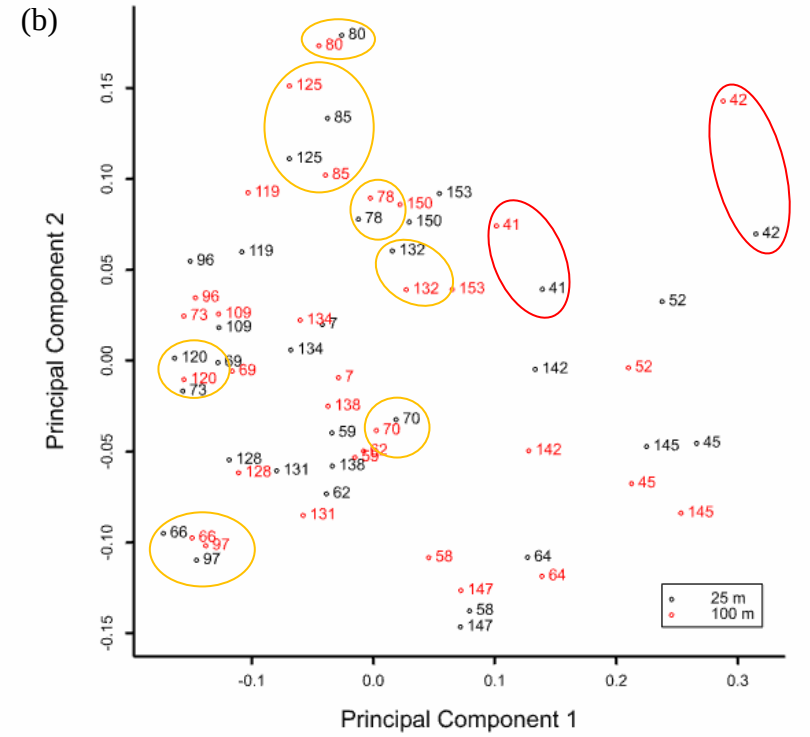
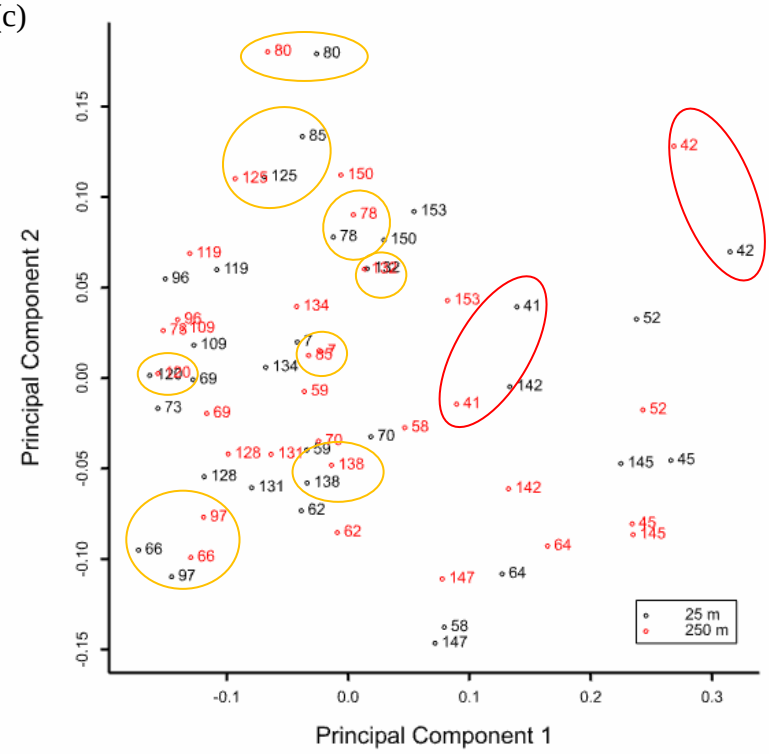


Figure 3.3 *(cont.)*

(c)



(d)

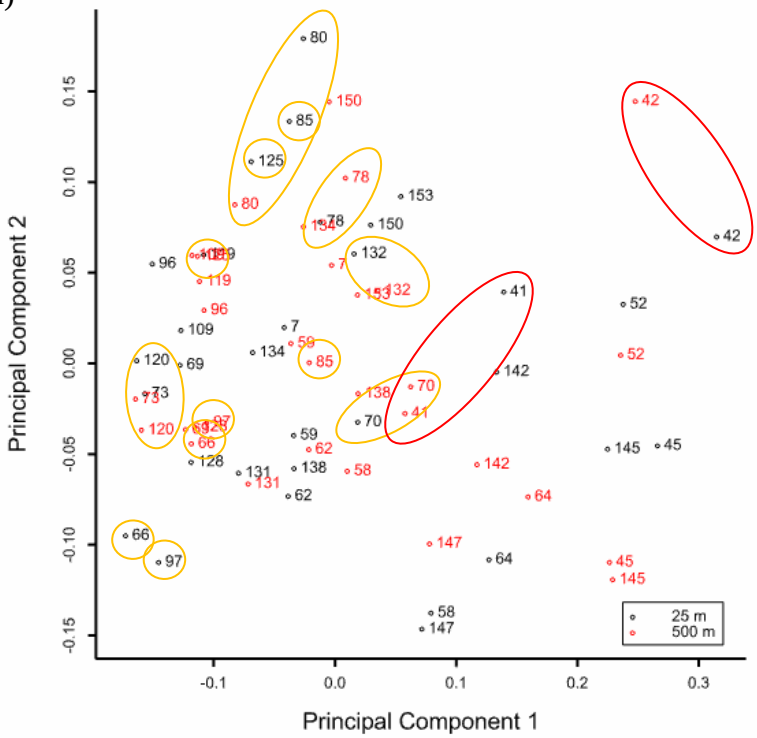


Figure 3.3 *(cont.)*

(e)

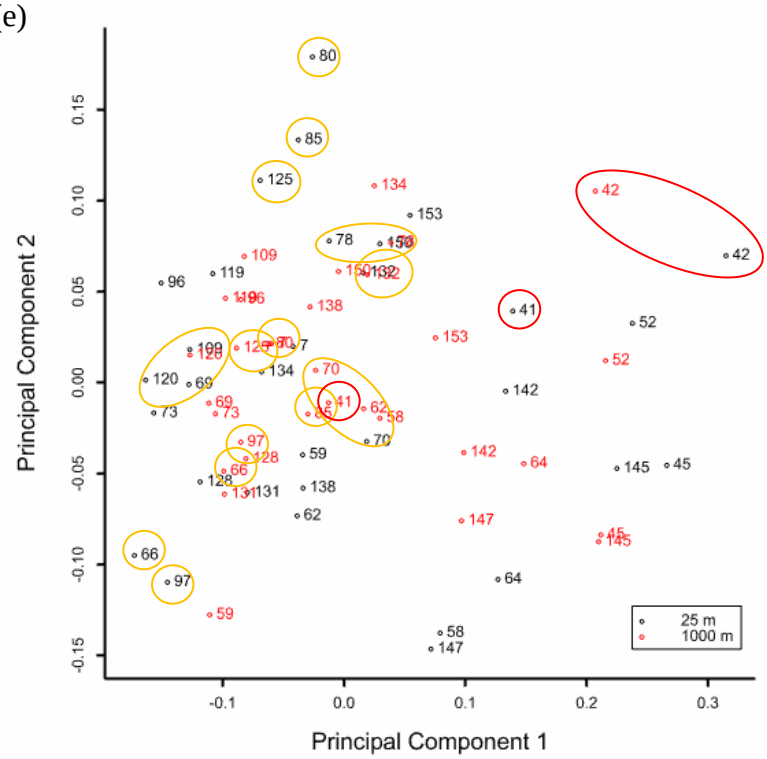


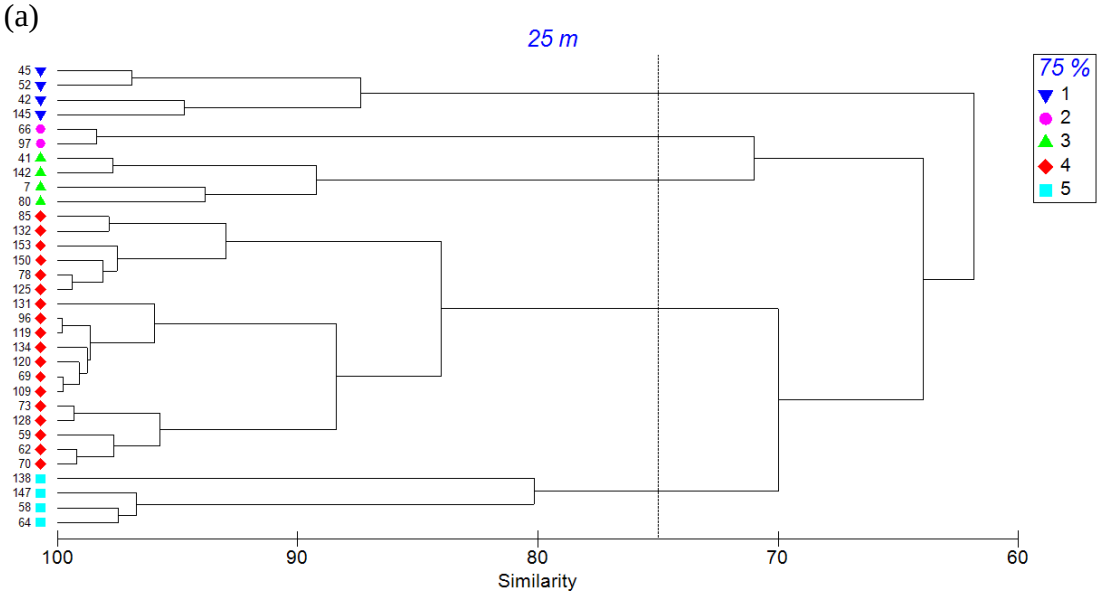
Figure 3.3: Comparison of the first two principal components (PC-1 and PC-2) of the configuration of the scores from the Principal Component Analysis (PCA) between scales; (a) 25 m to 50 m (b) 25 m to 100 m (c) 25 m to 250 m (d) 25 m to 500 m and (e) 25 m to 1000 m. Distance between the same NCA scores plotted at two different scales represents the NCA (scores) projected residuals from the Procrustes Rotation for the first two PCs. NCAs with low Procrustes residuals across scale comparisons from 25 m to 250 m are circled in orange and those with high Procrustes residuals across all scale comparisons are circled in red. See Appendix A6 for 4D Procrustes Residuals.

The 4D output from the PCA at each of the five scales was used to cluster NCAs according to their similarities to each other at each scale (Figure 3.4a-f). At the initial scale of 25 m five clusters are evident at a similarity threshold of 75% (Figure 3.4a). Cluster group 1 comprised four NCAs which had similarly high positive loadings on PC-1 (Figure 3.3a). These four NCAs were characterised by extra-large MPS, and low landscape diversity (DLC groups 8 and 9) (Figure 3.4a; Figure 3.1a,b). Cluster group 2 comprised two NCAs, both of which were characterised by small MPS (Figure 3.4a; Figure 3.1a). Group 3 comprised four NCAs which were dissimilar in terms of MPS and DLC groups (Figure 3.4a; Figure 3.1a). These four NCAs, however, were characterised by low negative scores on PC-3 and were similar in terms of the metrics which measured patch boundary configuration (Table 3.3).

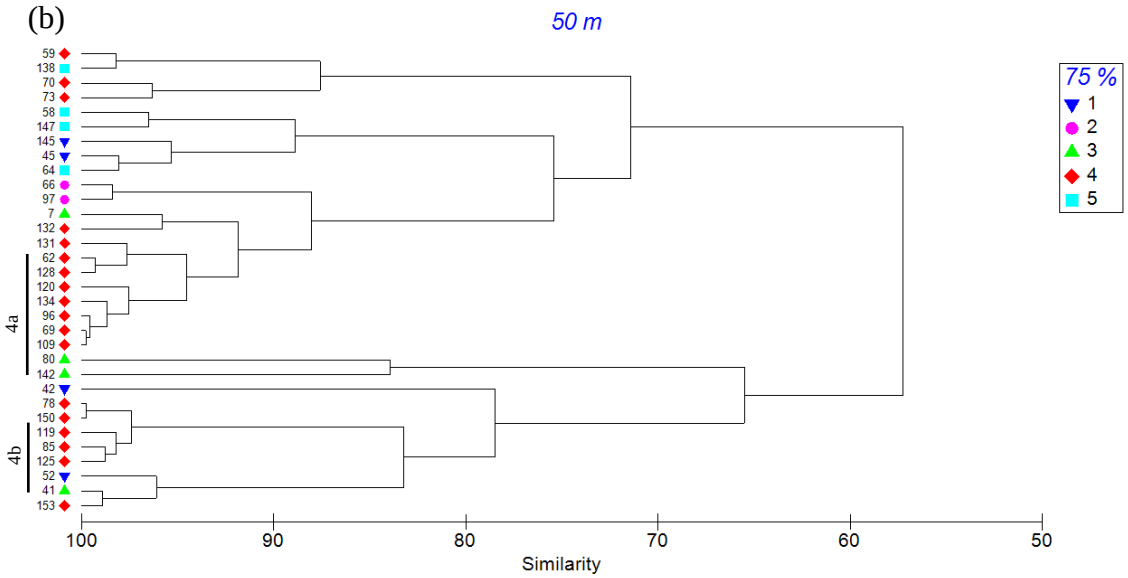
Cluster group 4 comprised 18 NCAs which had similarly low scores on PC-1 (Figure 3.4a; Figure 3.3a). Of these NCAs, ten were characterised by small and medium sized MPS (Figure 3.1a) and 12 by high landscape diversity (DLC groups 1-4) (Figure 3.1b) in comparison to the other NCAs. Cluster group 5 comprised four NCAs with similarly low negative scores on PC-2, and were similar in terms of the metrics which measured patch shape and extent (Figure 3.4a; Figure 3.3a; Table 3.3).

As the scale increases groups initially clustered at 25 m are separated (Figure 3.4b-f). Most notably, group 4 is separated at 50 m into three clusters at a similarity threshold of 75 % (Figure 3.4b). Once separated at 50 m a number of individual NCAs within group 4 (group 4A: 62, 128, 131, 69, 109, and 96 and group 4B: 78, 150, 85, and 125) remain clustered at 50 m, 100 m and 250 m at similarity thresholds of 70 % to 90 % (Figures 3.4b-d). At 500 m these individual metrics comprising group 4A and 4B are no longer clustered at a similarity threshold of 90 %, however, all members of group 4 are re-clustered at this scale at a similarity threshold of 70 % (Figure 3.4e). Groups 1, 3 and 5 are also separated into several different clusters from 50 m onwards, however group 2 remains clustered at all scales.

(a)



(b)



(c)

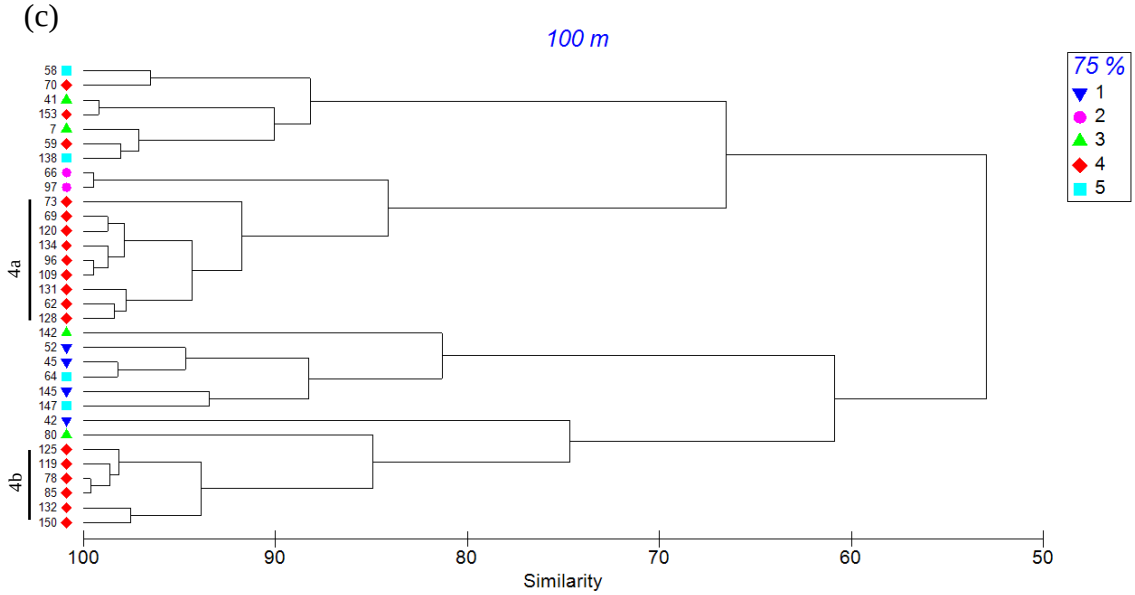


Figure 3.4 *(cont.)*

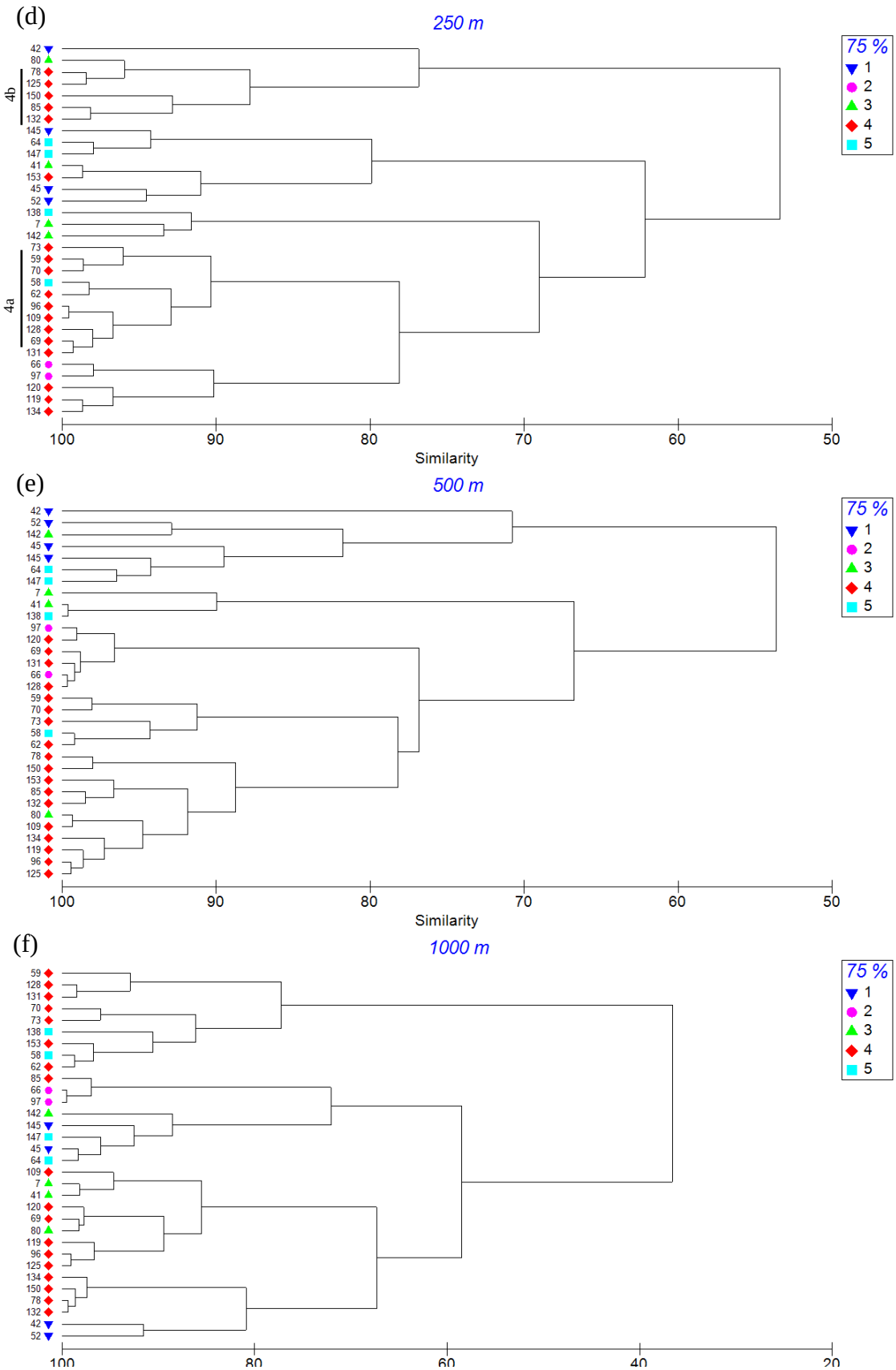


Figure 3.4: Clustering of National Character Areas (NCAs) using complete link algorithm based on Euclidean similarity matrix. The clusters at 25 m (a) are defined at a similarity level of 75 % and the members of these clusters are then identified at the scales (b) 50 m, (c)100 m, (d) 250 m, (e) 500 m and (f)1000 m.

3.2.5 Associations between metrics with scale: 25m to 1000m

When comparing the configuration of the loadings of the 4-D output from the PCA at 25 m (section 3.2.3) with that obtained at the different grain sizes (50 m, 100 m, 250 m, 500 m and 1000 m) by means of a Procrustes Rotation, the residual sums of squares (RSS) between the pair of configurations increases as the difference in grain sizes being compared increases (Table 3.7). The RSS obtained for each pairwise comparison to 25 m is higher than the RSS obtained from the pairwise comparisons starting from any other scale (Table 3.7). This increase in RSS with scale and high RSS when comparing scales to 25 m is the same pattern to that obtained when comparing the 4D configuration of the NCAs (scores) (see section 3.2.4; Table 3.5). Metric values obtained for grain sizes 250 m and 500 m are most similar, and, relative similarity is also high when comparing 25 m to 50 m, 50 m to 100 m, and 500 m to1000 m (Table 3.7).

When considering the behaviour of individual metrics, as grain size increases from 25 m, a number of metrics including AREA_RA, COHESION, CONTAG, ECON_AM, FRAC_AM, GYRATE_AM, and GYRATE_CV maintain similar values relative to the other metrics with consistently small Procrustes residuals (Figure 3.5a-e – highlighted in green; Table 3.9). Procrustes residuals are obtained when comparing the 4-D configuration of the loadings at two different scales (Appendix A6), and the projected Procrustes residuals (distance) between metric loadings for the first two PCs at each scale comparison are highlighted in Figures 3.5a-e. Several metrics also maintain small Procrustes residuals when comparing 25 m to 50 m and 100 m but not when comparing to larger scales; CWED, LSI, SIMI_AM, SIMI_CV and SIDI (Figure 3.5a-e – highlighted in orange; Table 3.9). In contrast the metrics FRAC_CV, CONTIG_MN, PROX_CV and CONTIG_RA maintain consistently high Procrustes residuals indicating variation in metric value with scale (Figure3.5a-e – highlighted in red; Table 3.9).

Relative Euclidean similarities between the metrics obtained from the 4D output from the PCA at each scale are significantly correlated with each other at each scale combination (Table 3.8). Similarly to the pattern observed from the Procrustes Rotation RSS (Table 3.7) the correlation coefficient decreases as the difference in the scale comparison increases (Table 3.8). Correlation coefficients are lowest when

comparing each scale to 25 m in comparison to any other starting scale. For example, the correlation when comparing 100 m to 25 m is lower than that obtained when comparing 100 m to 50 m (r = 0.8418, p<0.001; r = 0.9472, p<0.001 respectively).

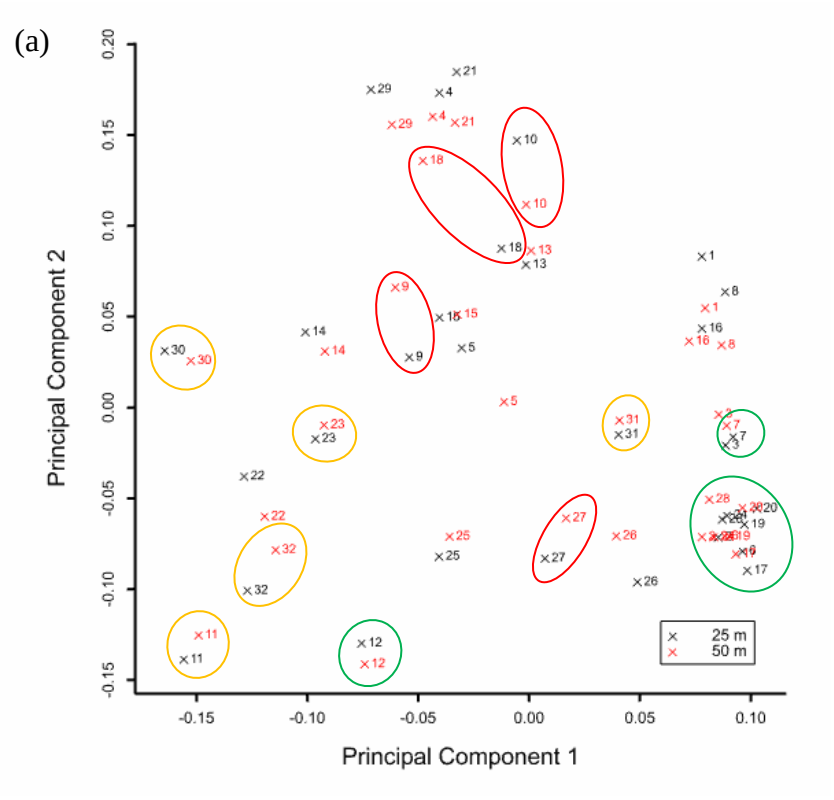
25	-					
50	0.1130	-				
100	0.2596	0.1038	-			
250	0.3630	0.3206	0.2177	-		
500	0.4408	0.3284	0.2098	0.0915	-	
1000	0.5036	0.4182	0.3167	0.1421	0.1186	-
	25	**50**	**100**	**250**	**500**	**1000**

Table 3.7: Residual Sums of Squares (RSS) from the Procrustes Rotation comparison of the 4-dimension configuration of the loadings from the Principal Component Analysis for each scale.

25	-					
50	0.9177	-				
100	0.8418	0.9472	-			
250	0.7244	0.7911	0.8574	-		
500	0.6512	0.7538	0.8228	0.9285	-	
1000	0.5904	0.6823	0.7403	0.8543	0.8929	-
	25	**50**	**100**	**250**	**500**	**1000**

Table 3.8: Pearson product-moment correlations between the Euclidean similarity matrices derived from the 4-dimension configuration of the loadings from the Principal Component Analysis for each scale. Pearson product-moment correlations are obtained by Mantel Tests for each pairwise comparison and all correlations are significant (p<0.001).

(a)



(b)

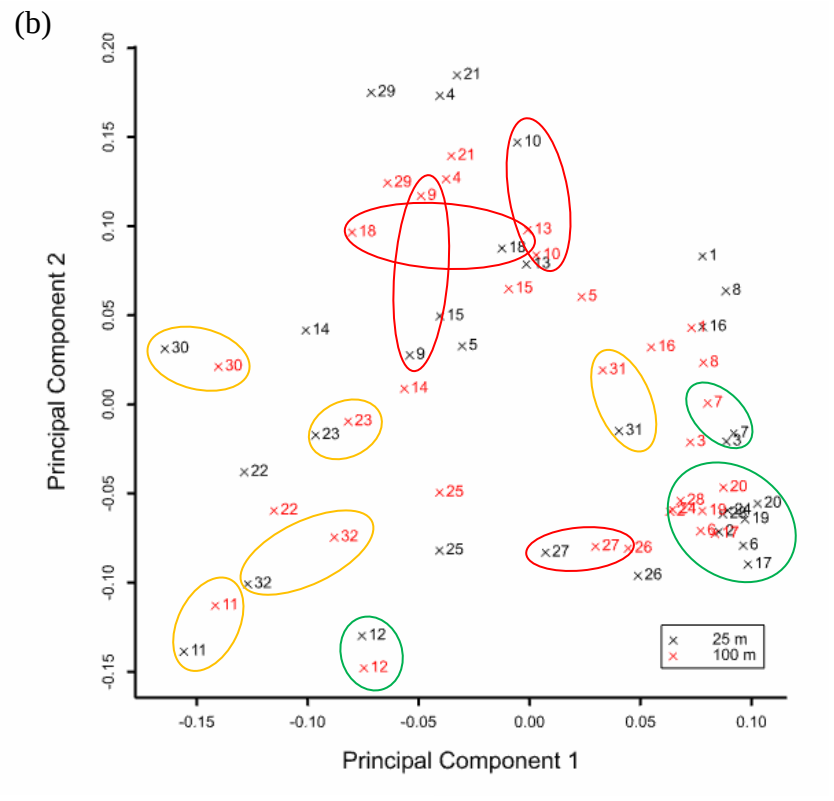
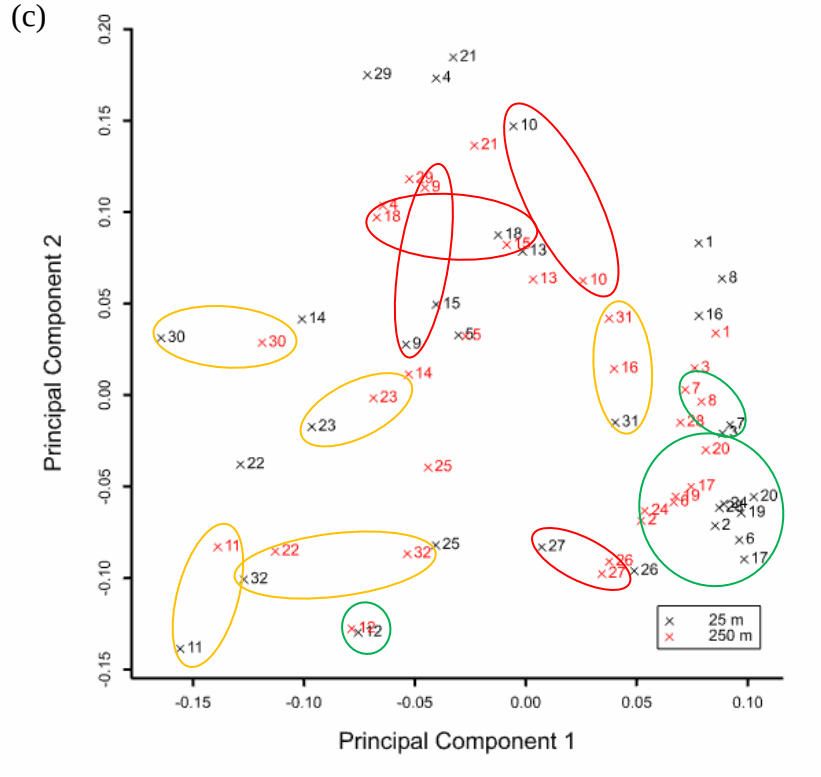


Figure 3.5 *(cont.)*

(c)



(d)

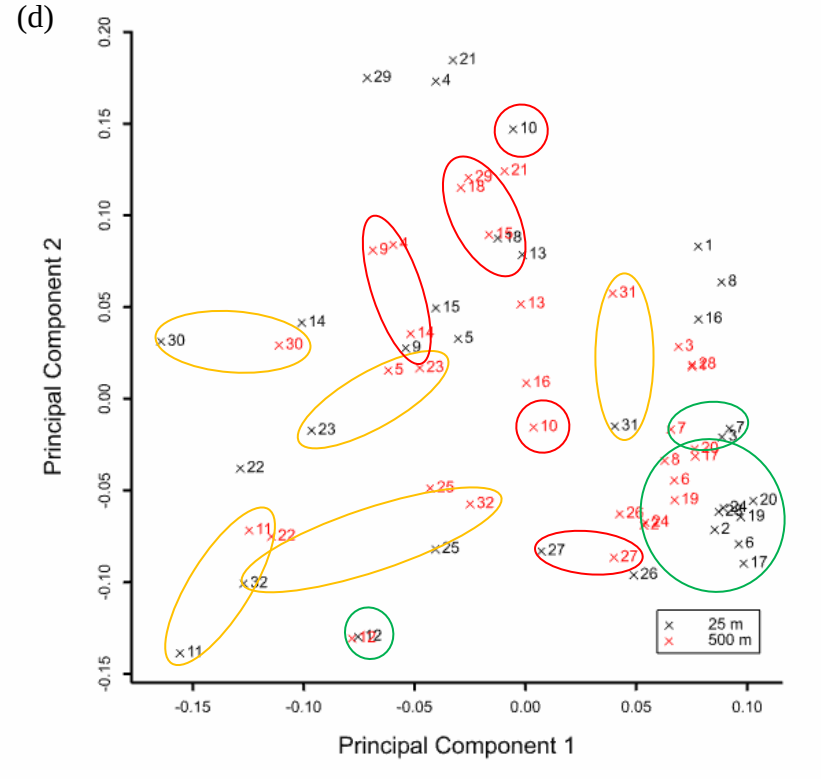


Figure 3.5 *(cont.)*

(e)

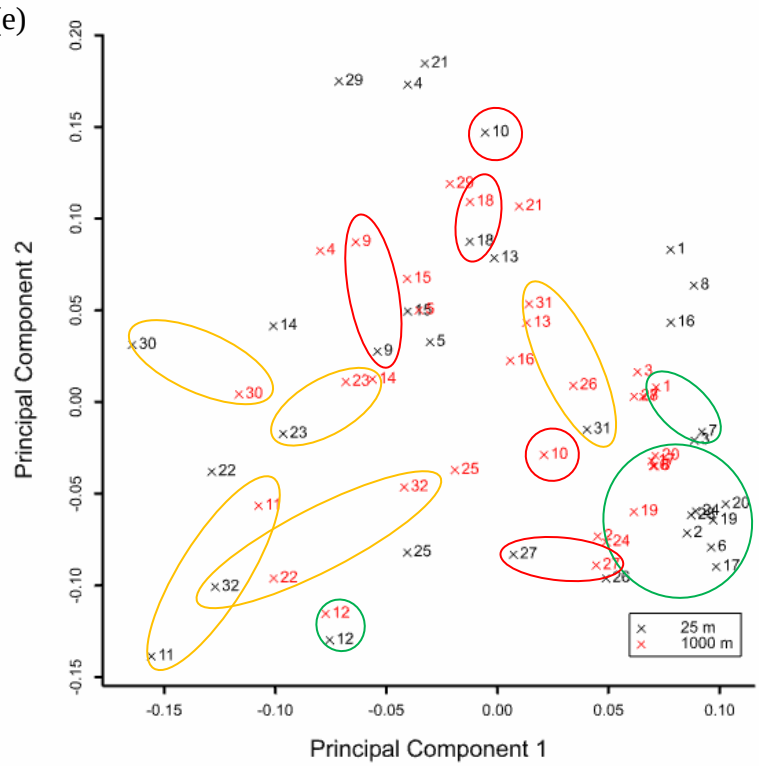


Figure 3.5 a-e: Comparison of the first two principal components (PC-1 and PC-2) of the 4-D configuration of the loadings from the Principal Component Analysis (PCA) between scales; (a) 25 m to 50 m (b) 25 m to 100 m (c) 25 m to 250 m (d) 25 m to 500 m and (e) 25 m to 1000 m. Distance between the same metric loadings plotted at two different scales represents the metric (loadings) projected residuals from the Procrustes Rotation for the first two PCs. Metrics with high Procrustes residuals across scales are circled in red; small Procrustes residuals across scales in green; and small Procrustes residuals from 25 m to 100 m only in orange. See Appendix A6 for 4D Procrustes Residuals.

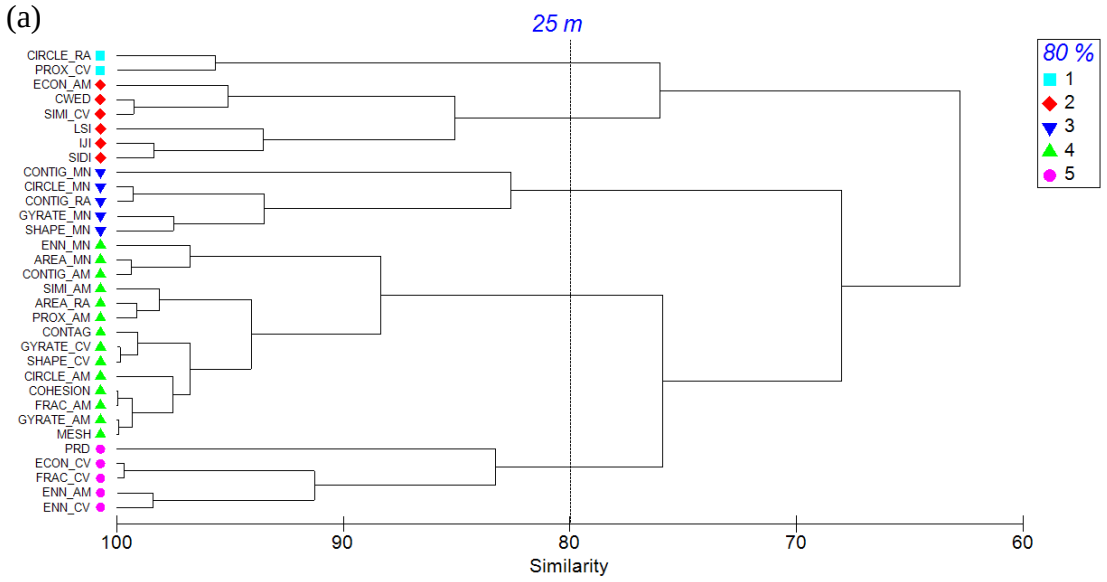
Metric	Code	Metric *(cont.)*	Code
AREA_MN	1	FRAC_AM	17
AREA_RA	2	FRAC_CV	18
CIRCLE_AM	3	GYRATE_AM	19
CIRCLE_MN	4	GYRATE_CV	20
CIRCLE_RA	5	GYRATE_MN	21
COHESION	6	IJI	22
CONTAG	7	LSI	23
CONTIG_AM	8	MESH	24
CONTIG_MN	9	PRD	25
CONTIG_RA	10	PROX_AM	26
CWED	11	PROX_CV	27
ECON_AM	12	SHAPE_CV	28
ECON_CV	13	SHAPE_MN	29
ENN_AM	14	SIDI	30
ENN_CV	15	SIMI_AM	31
ENN_MN	16	SIMI_CV	32

Table 3.9: Metric codes for Procrustes Residual Plots.

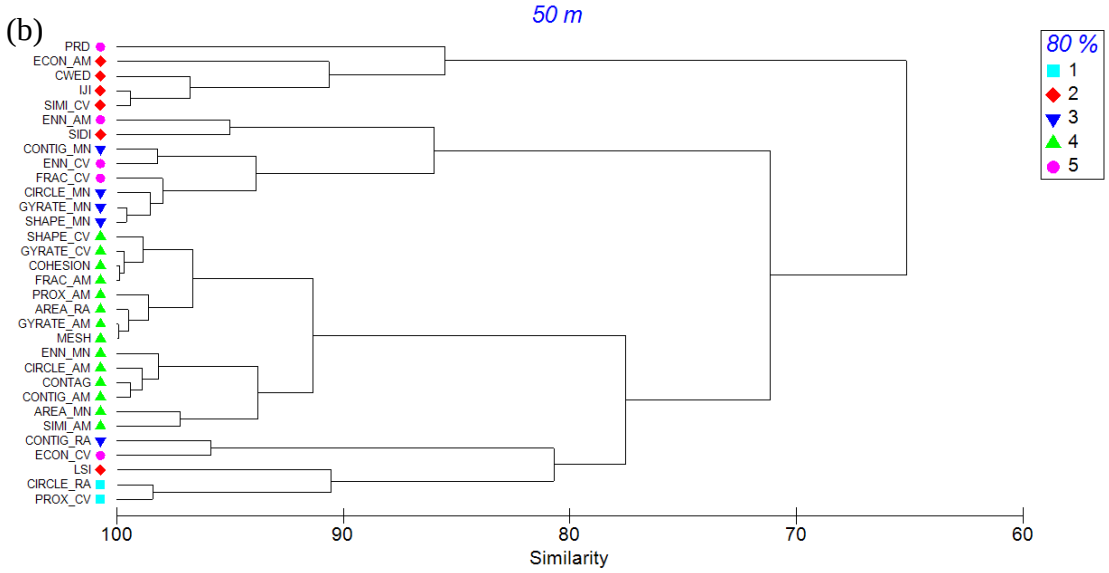
The hierarchal cluster analysis of NCA landscape structure according to the metric values obtained at a scale of 25 m suggests the partitioning of five clusters of metrics at a similarity level of 80% (Figure 3.6a). Cluster group 1 comprises two metrics associated with PC-4 providing measures of patch shape irregularity (Table 3.3). Cluster group 2 comprises six metrics, four of which are associated with PC-1 with high negative loadings on this component, and collectively provide measures of landscape fragmentation (Table 3.3; Figure 3.5a; Figure 3.6a). Cluster group 3 comprises five metrics which are associated with PC-2, and measure patch shape and extent. Cluster group 4 included 14 metrics, all which have the highest positive loadings on PC-1 and collectively measure landscape fragmentation (Table 3.3; Figure 3.5a; Figure 3.6a). Cluster group 5 comprises five metrics, three of which are associated with PC-3 and provide measures of patch boundary configuration (Table 3.3).

Relationships between metrics are maintained across scales most notably for members of group 4, which remain clustered at 50 m (92% similarity) and 100 m (82% similarity), despite addition of some metrics from the other groups at 100 m (Figure 3.6b,c). From 250 m upwards the metrics ENN_MN and SIMI_AM are no longer clustered with group 4, but the similarity between the remaining members of group 4 increases to 82%, 89% and 90% at the scales 250 m, 500 m and 1000 m respectively (Figure 3.6d-f). Group 3 also maintain similarity across scales, with only the metric CONTIG_RA becoming separated at 50 m (Figure 3.6b). The remaining groups change considerably with increasing grain size, most notably group 1 metrics (CIRCLE_RA and PROX_CV), group 5 metrics (ECON_CV, PRD, FRAC_CV and ENN_CV) and the metric CONTIG_RA from group 3 exhibited greatest variability, joining and re-joining groups at different scales (Figures 3.6a-e).

(a)



(b)



(c)

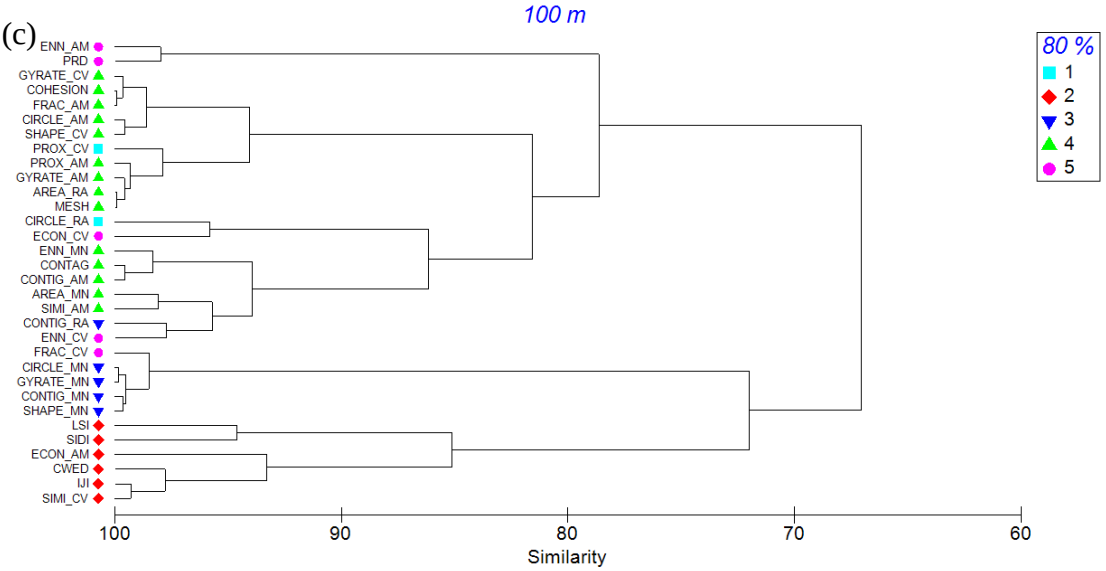


Figure 3.6 *(cont.)*

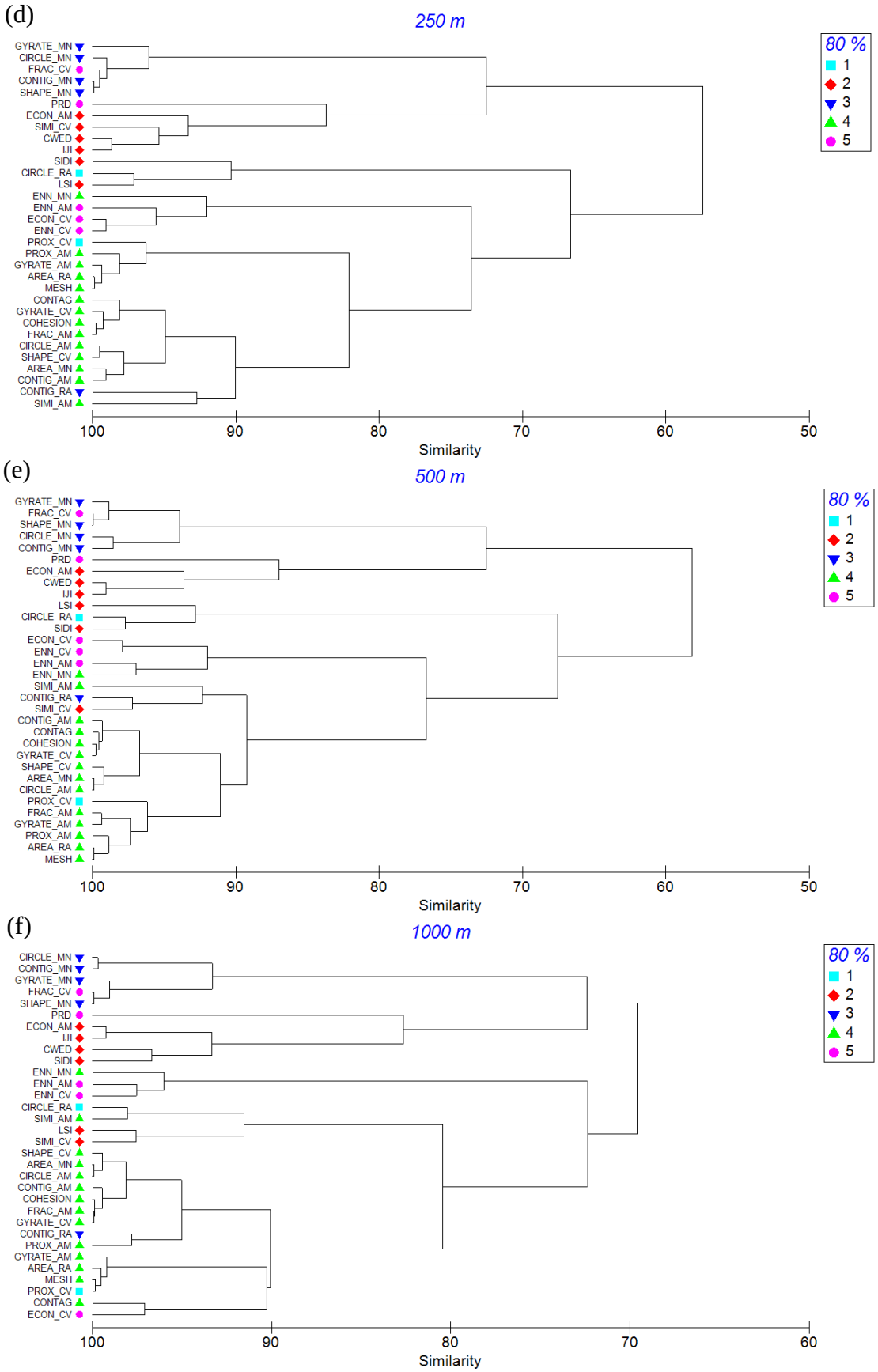


Figure 3.6: Clustering of landscape structure metrics using complete link algorithm based on Euclidean similarity matrix. The clusters at 25 m (a) are defined at a similarity level of 80 % and the members of these clusters are then identified at the scales (b) 50 m, (c) 100 m, (d) 250 m, (e) 500 m and (f) 1000 m.

3.3 Results: Warwickshire grid square landscapes (county scale)

3.3.1 Classification of landscape characteristics: 25 m

Mean Patch Size (MPS)

For the metrics derived from the LCM 2000, all metrics with the exception of PROX_AM were able to distinguish between grid squares grouped by their mean patch size (MPS), with significant differences observed in metric value (Table 3.10). For 22 out of the 31 metrics which discriminated between grid squares on the basis of MPS, significant differences were observed between all four groups of grid squares (small, medium, large and extra-large). For the remaining nine metrics significant differences were observed between the grid squares with smaller patch size (small and medium) in comparison to the grid squares with larger patch size (large and extra-large). For the metrics derived from the PH1 2000, all metrics were able to discriminate between grid squares grouped by MPS (Table 3.11). The majority of these metrics (23) discriminated between all four groups of grid squares by MPS.

Number of land cover classes (NLC)

When comparing the metric values between grid squares grouped by the number of land cover classes (NLC) significant differences were observed for 27 of the metrics derived from the LCM 2000 (Table 3.10). The values for the metrics CIRCLE_MN, ENN_MN, FRAC_CV, PROX_CV and SHAPE_CV did not differ significantly between the grid squares grouped by NLC (Table 3.10). For the metrics which did discriminate between grid squares based on NLC, significant differences were observed between grid squares characterised by the smallest number of land cover classes, and the other three groups. For the metrics based on the PH1 2000, all 32 discriminated between landscapes grouped by NLC (Table 3.11), with 16 of these discriminating between all four groups and the rest between those grid squares characterised by the smallest number of land cover classes, in comparison to the other three groups.

Diversity of Land Cover (DLC)

With exception of the metrics CONTIG_RA and ENN_CV, landscape structure metrics derived from the LCM 2000 discriminated between grid squares grouped by Diversity of Land Cover (DLC) (Table 3.10). For nine of these metrics, significant differences occurred between grid squares characterised by low landscape diversity (DLC groups 12, 13 and 14), in comparison to those characterised by higher landscape diversity (DLC groups 1-8). For the metric SIDI, grid squares grouped by DLC 12-14 differed significantly to all other groups and for the metric CIRCLE_AM, only DLC group 14 differed to all other groups. For the metrics derived from PH1 2000, all 32 discriminated between DLC (Table 3.11), with nine metrics detecting significant differences between grid squares characterised by high landscape diversity (DLC group 1), in comparison to those characterised by lower landscape diversity (DLC groups 15-19). In particular, for the metric SIDI significant differences were obtained between the grid squares with lowest landscape diversity (DLC 19) from highest diversity (DLC 1-5).

3.3.2 Discriminating ability of metrics with scale: 25 m – 250 m

For the grid square landscapes derived from the LCM 2000, significant differences were observed between grid squares grouped by MPS for all metrics apart from PROX_ AM at 25 m and SHAPE_CV at 100 m and 250 m (Table 3.12). When considering the grouping of the grid squares by NLC, the five metrics which were non-significant at 25 m were also unable to discriminate between landscapes at 50 m (Table 3.12). However, only the metrics ENN_MN, FRAC_CV and PROX_CV were unable to discriminate between landscapes at 100 m, and a further three metrics at 250 m. For the grouping of grid squares by DLC, all 32 metrics discriminated between DLC groups at 50 m and all apart from SHAPE_CV at 100 m and 250 m.

For the grid square landscapes derived from the PH1 2000, metrics which significantly discriminated between landscapes based on MPS and DLC maintained this discriminating ability across scales from 25 m to 250 m, with the exception of the metric CIRCLE_MN, and SHAPE_MN at 50 m, and ECON_CV at 250 m (Table 3.12). Most metrics were also able to discriminate by NLC across scales, with exception of the metric ENN_MN at 50 m and ECON_CV at 100 m.

Metric	Mean Patch Size (MPS)		Number of Land Cover Classes (NLC)		Dominant Land Cover (DLC)	
	$F_{3,2423}$	P	$F_{3,2423}$	P	$F_{13,2413}$	P
AREA_MN	4166.12	0.000	317.20	0.000	35.38	0.000
AREA_RA	473.15	0.000	88.72	0.000	73.62	0.000
CIRCLE_AM	41.28	0.000	12.43	0.000	25.38	0.000
CIRCLE_MN	5.30	0.000	0.22	0.882	4.95	0.000
CIRCLE_RA	121.06	0.000	38.81	0.000	6.21	0.000
COHESION	1706.75	0.000	215.76	0.000	52.28	0.000
CONTAG	568.50	0.000	34.96	0.000	88.70	0.000
CONTIG_AM	1858.68	0.000	227.64	0.000	51.40	0.000
CONTIG_MN	59.45	0.000	17.70	0.000	2.37	0.004
CONTIG_RA	48.48	0.000	14.51	0.000	0.87	0.583
CWED	1149.43	0.000	326.61	0.000	49.85	0.000
ECON_AM	413.73	0.000	219.08	0.000	59.87	0.000
ECON_CV	77.97	0.000	83.76	0.000	33.50	0.000
ENN_AM	21.01	0.000	12.50	0.000	11.88	0.000
ENN_CV	28.80	0.000	14.46	0.000	1.12	0.333
ENN_MN	67.86	0.000	0.17	0.915	9.49	0.000
FRAC_AM	13.05	0.000	5.06	0.002	6.24	0.000
FRAC_CV	7.27	0.000	1.48	0.218	1.86	0.030
GYRATE_AM	880.78	0.000	144.65	0.000	75.40	0.000
GYRATE_CV	159.24	0.000	36.63	0.000	8.83	0.000
GYRATE_MN	986.77	0.000	195.36	0.000	8.83	0.000
IJI	236.48	0.000	11.66	0.000	26.90	0.000
LSI	1892.38	0.000	227.71	0.000	51.65	0.000
MESH	799.27	0.000	131.13	0.000	81.35	0.000
PRD	410.11	0.000	4357.72	0.000	15.65	0.000
PROX_AM	1.84	0.138	5.51	0.001	2.60	0.001
PROX_CV	15.98	0.000	2.12	0.096	13.67	0.000
SHAPE_CV	8.08	0.000	2.54	0.055	2.33	0.004
SHAPE_MN	55.49	0.000	24.33	0.000	3.13	0.000
SIDI	545.88	0.000	168.68	0.000	108.76	0.000
SIMI_AM	212.45	0.000	56.15	0.000	119.78	0.000
SIMI_CV	31.61	0.000	41.29	0.000	8.95	0.000

Table 3.10: Differences in metric output derived from the LCM 2000 between grid squares classified by Mean Patch Size (MPS), Number of Land Cover classes (NLC) and Diversity of Land Cover (DLC). Differences are assessed by means of a One-Way ANOVA with test statistic (*F*), degrees of freedom and significance value (*P*) provided.

Metric	Mean Patch Size (MPS)		Number of Land Cover Classes (NLC)		Dominant Land Cover (DLC)	
	$F_{3,2423}$	P	$F_{3,2423}$	P	$F_{18,2423}$	P
AREA_MN	3826.36	0.000	1115.97	0.000	16.23	0.000
AREA_RA	358.48	0.000	214.28	0.000	21.26	0.000
CIRCLE_AM	148.61	0.000	94.07	0.000	11.95	0.000
CIRCLE_MN	38.02	0.000	18.72	0.000	2.85	0.000
CIRCLE_RA*	409.28	0.000	273.80	0.000	102.96	0.000
COHESION	1195.05	0.000	633.11	0.000	22.58	0.000
CONTAG	396.30	0.000	171.45	0.000	32.45	0.000
CONTIG_AM	1717.43	0.000	706.10	0.000	31.48	0.000
CONTIG_MN	329.94	0.000	195.06	0.000	2.85	0.000
CONTIG_RA*	274.28	0.000	175.01	0.000	206.36	0.000
CWED	1512.90	0.000	630.00	0.000	33.31	0.000
ECON_AM	524.38	0.000	277.36	0.000	27.53	0.000
ECON_CV	49.06	0.000	21.28	0.000	13.30	0.000
ENN_AM	66.79	0.000	12.07	0.000	4.87	0.000
ENN_CV	161.74	0.000	93.75	0.000	5.11	0.000
ENN_MN*	184.01	0.000	26.74	0.000	110.68	0.000
FRAC_AM	158.53	0.000	83.49	0.000	13.02	0.000
FRAC_CV	101.84	0.000	53.25	0.000	8.05	0.000
GYRATE_AM	616.32	0.000	564.42	0.000	22.07	0.000
GYRATE_CV	7.26	0.000	9.99	0.000	3.67	0.000
GYRATE_MN	1000.13	0.000	564.42	0.000	5.57	0.000
IJI	208.29	0.000	92.27	0.000	8.73	0.000
LSI	1679.44	0.000	711.92	0.000	31.30	0.000
MESH	459.71	0.000	264.00	0.000	19.64	0.000
PRD	1175.40	0.000	5267.62	0.000	13.17	0.000
PROX_AM	64.16	0.000	31.27	0.000	6.98	0.000
PROX_CV	59.58	0.000	83.83	0.000	2.71	0.000
SHAPE_CV	83.65	0.000	37.83	0.000	13.97	0.000
SHAPE_MN	17.82	0.000	16.83	0.000	3.78	0.000
SIDI	436.83	0.000	322.13	0.000	33.80	0.000
SIMI_AM	25.76	0.000	17.53	0.000	25.00	0.000
SIMI_CV	47.61	0.000	36.16	0.000	10.25	0.000

Table 3.11: Differences in metric output derived from the PH1 2000 between grid squares classified by Mean Patch Size (MPS), Number of Land Cover classes (NLC) and Diversity of Land Cover (DLC). Differences are assessed by means of a One-Way ANOVA with test statistic (*F*), degrees of freedom and significance value (*P*) provided. * indicates results for Kruskal-Wallis Test are provided.

Metric	LCM 2000												PH1 2000											
	25 m			50 m			100 m			250 m			25 m			50 m			100 m			250 m		
	M	N	D	M	N	D	M	N	D	M	N	D	M	N	D	M	N	D	M	N	D	M	N	D
AREA_MN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
AREA_RA	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CIRCLE_AM	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CIRCLE_MN	*		*	*		*	*	*	*	*	*	*	*	*	*		*		*	*	*	*	*	*
CIRCLE_RA	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
COHESION	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CONTAG	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CONTIG_AM	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CONTIG_MN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CONTIG_RA	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CWED	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ECON_AM	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ECON_CV	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*			*
ENN_AM	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*
ENN_CV	*	*		*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*
ENN_MN	*		*	*		*	*		*	*		*	*	*	*				*	*	*	*	*	*
FRAC_AM	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
FRAC_CV	*		*	*		*	*		*	*		*	*	*	*	*	*	*	*	*	*	*	*	*
GYRATE_AM	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
GYRATE_CV	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
GYRATE_MN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
IJI	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
LSI	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
MESH	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
PRD	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
PROX_AM			*	*		*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*
PROX_CV	*		*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
SHAPE_CV	*		*	*		*		*					*	*	*	*	*	*	*	*	*	*	*	*
SHAPE_MN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*		*	*	*	*	*	*
SIDI	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
SIMI_AM	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
SIMI_CV	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

Table 3.12: Discriminating ability of metrics with increasing grain size from a resolution of 25 m to 250 m, based on significant differences in average metric output between grid square landscapes derived from LCM 2000 and PH1 2000 grouped by Total Area (A), Mean Patch Size (M), Number of land cover classes (N) and Diversty of Land Cover (D). Results are based on a One-Way Analysis of Variance, and * indicate significant differences ($p<0.05$).

3.3.3 Discriminating ability of metrics combined – comparison between LCM 2000 and PH1 2000: 25 m

When comparing the output from the Principal Component Analysis of the 32 landscape structure metrics derived from the LCM 2000 and PH1 2000, the first four Principal Components (PCs) explain a similar amount of variation in the metrics amongst the grid squares (64.37 % and 67.02 % respectively).

Metrics with high loadings on PC-1 and PC-2 are similar between the PCA outputs for the two data sets (Table 3.13 and 3.14). For the LCM 2000 there are 13 metrics with relatively high loadings on PC-1 and for the PH1 2000 there are 12 metrics, 11 of which are the same for the two data sets. PC-1 for the LCM 2000 includes the additional metrics GYRATE_CV and IJI and for the PH1 2000 PC-1 additionally includes PRD. Considering the 11 metrics associated with PC-1 for both data sets, nine metrics are also associated with PC-1 of the PCA considering variability between the NCAs (Table 3.3), and as such provide measures of landscape fragmentation.

There are seven metrics with high loadings on PC-2, which are common to the two data sets (Table 3.13 and 3.14). Of these, four metrics are also associated with PC-2 derived from the NCA PCA and PC-2 was considered to provide a measure of patch shape and extent (Table 3.3). The additional metrics associated with PC-2 for both LCM 2000 and PH1 2000 provide additional measures of patch extent (e.g. CONTIG_MN) as well as measures of variability in patch size, for example, GYRATE_CV measures the variability in mean patch extent.

There are five metrics strongly associated with PC-3 for the LCM 2000, and seven metrics for the PH1 2000, with only three metrics (FRAC_CV, SHAPE_CV and SHAPE_MN) in common (Table 3.13 and 3.14). For the LCM 2000, only one metric associated with PC-3 is the same as the metrics with highest loadings on PC-3 in the NCA PCA (Table 3.3). The six metrics associated with PC-3 for the LCM 2000 can be considered to be associated with patch shape complexity, with the five metrics directly measuring measure shape and variability (Table 3.13; Appendix A4). For the PH1 2000 the seven metrics with highest loadings together provide measures of landscape functional connectivity, with metrics measuring not only the complexity of

the boundary (SHAPE_MN), but also the variability in degree of contrast with the matrix (ECON_CV), and the isolation to patches of the same class type (ENN_AM and PROX_AM) (Table 3.14; Appendix A4).

For the LCM 2000, there are nine metrics with high loadings on PC-4, and for the PH1 2000, there are five metrics, with only two in common between the two data sets (Table 3.13 and 3.14). For the LCM 2000, PC-4 shows greatest similarity to the PH1 2000 PC-3, comprising metrics which together provide measures of landscape connectivity, considering the contrast between class types and isolation between patches of the same or similar class. The five metrics associated with PC-4 for the PH1 2000 provide additional measures of patch shape complexity and landscape connectivity.

Metrics with high loadings can discriminate between grid square landscapes. For example the grid squares ID 3664 and ID 3047 are considered dissimilar on PC-1 based on the PCA from the LCM 2000 and PH1 2000 but similar on PC-3 based on the PCA from the PH1 2000 and similar on PC-2 based on the PCA from the LCM 2000 (Table 3.15a,b). On visual inspection the two grid square landscapes can be seen to be dissimilar in terms of patch size and aggregation for both LCM 2000 and PH1 2000 data sets (Figure 3.7a-d). In particular for both data sets, the two grid squares differ in terms of the following four metrics with high loadings on PC-1; GYRATE_AM, MESH, CWED and SIDI (Figure 3.7a-d; Table 3.15a,b). Differences are observed for both data sets in terms of patch extent and aggregation, with higher area-weighted mean radius of gyration (GYRATE_AM) and area-weighted mean patch size (MESH) for grid square 3664 in comparison to 3047 (Table 3.15a,b). The higher levels of subdivision within grid square 3047 is associated with greater edge contrast (CWED) between neighbouring patches and higher landscape diversity (SIDI) in comparison to grid square 3664 (Table 3.15a,b).

Similarities are observed in the two grid square landscapes when considering metrics with high loadings on different components between the two data sources (Table 3.13; Table 3.14). For the two LCM 2000 grid squares, similarities are observed on PC-2 in terms of average patch shape and range in patch shape as measured by the metrics SHAPE_MN, CIRCLE_MN, CONTIG_MN and CONTIG_RA (Table 3.13;

Table 3.15a). For the two PH1 2000 grid squares, similarities are observed on PC-3 in terms of average patch shape complexity and variability, as measured by the metrics SHAPE_MN, SHAPE_CV and FRAC_CV (Table 3.14; Table 3.15b). The straight-line distance between patches of the same type, weighted by patch area (ENN_AM), is also similar between the two landscapes (Table 3.15b).

Landscape metric	PC1	PC2	PC3	PC4	Landscape metric *(cont.)*	PC1	PC2	PC3	PC4
Area					**Shape**				
AREA_MN	0.857	-0.366	-0.026	0.054	CIRCLE_MN	-0.124	-0.537	0.263	0.284
AREA_RA	0.872	0.278	-0.003	0.214	CIRCLE_AM	-0.417	-0.260	0.554	0.105
GYRATE_MN	0.489	-0.768	0.150	0.094	CIRCLE_RA	-0.404	0.210	0.171	-0.114
GYRATE_AM	0.934	0.153	0.163	0.195	CONTIG_MN	-0.017	-0.795	-0.193	0.244
GYRATE_CV	0.717	0.567	0.206	0.062	CONTIG_AM	0.924	-0.064	-0.251	-0.012
					CONTIG_RA	-0.106	0.592	0.050	-0.383
Aggregation					FRAC_AM	0.106	0.145	0.851	0.294
COHESION	0.937	0.039	0.193	0.151	FRAC_CV	-0.103	0.224	0.780	-0.093
CONTAG	0.887	0.190	-0.037	0.016	SHAPE_MN	-0.023	-0.640	0.589	0.240
ENN_MN	0.326	0.124	-0.226	0.474	SHAPE_CV	0.049	0.191	0.866	0.083
ENN_AM	-0.278	-0.014	-0.275	0.317					
ENN_CV	-0.108	0.258	0.178	-0.216	**Contrast**				
IJI	-0.639	-0.064	-0.334	0.156	ECON_AM	-0.684	0.264	-0.007	0.374
LSI	-0.924	0.071	0.253	0.010	ECON_CV	0.374	-0.178	0.145	-0.435
MESH	0.911	0.224	-0.046	0.164	CWED	-0.876	0.204	0.127	0.212
PROX_AM	-0.027	-0.164	0.237	-0.398					
PROX_CV	0.281	0.370	0.030	0.321	**Diversity**				
SIMI_AM	0.553	-0.114	0.209	-0.270	PRD	-0.583	0.346	-0.060	0.215
SIMI_CV	-0.103	0.430	-0.111	0.384	SIDI	-0.907	-0.138	0.002	0.006

Table 3.13: The loadings of 32 landscape metrics derived from the LCM 2000 on the principal components, grouped by landscape aspect. Note: variables with the highest loadings on each component have been highlighted.

Landscape metric	PC1	PC2	PC3	PC4	Landscape metric *(cont.)*	PC1	PC2	PC3	PC4
Area					**Shape**				
AREA_MN	-0.749	-0.410	0.040	-0.156	CIRCLE_MN	-0.019	-0.650	0.137	-0.096
AREA_RA	-0.836	0.341	0.184	-0.089	CIRCLE_AM	0.697	-0.285	0.005	0.360
GYRATE_MN	-0.526	-0.761	0.197	-0.022	CIRCLE_RA	0.624	0.281	0.317	0.170
GYRATE_AM	-0.883	0.248	0.217	0.081	CONTIG_MN	-0.380	-0.810	-0.078	-0.136
GYRATE_CV	-0.222	0.870	0.053	0.169	CONTIG_AM	-0.947	0.007	0.070	0.060
					CONTIG_RA	0.103	0.529	0.121	0.225
Aggregation					FRAC_AM	0.500	0.142	0.137	0.364
COHESION	-0.902	0.079	0.153	0.220	FRAC_CV	0.578	-0.029	0.656	0.162
CONTAG	-0.814	0.343	0.304	0.073	SHAPE_MN	0.207	-0.643	0.559	0.109
ENN_MN	-0.429	0.023	-0.164	-0.004	SHAPE_CV	0.577	-0.173	0.629	0.204
ENN_AM	0.220	-0.090	-0.433	0.225					
ENN_CV	0.495	0.236	0.164	0.053	**Contrast**				
IJI	0.522	0.075	-0.204	-0.340	ECON_AM	0.843	0.036	0.194	-0.187
LSI	0.944	0.000	-0.064	-0.058	ECON_CV	-0.349	-0.144	-0.441	0.483
MESH	-0.890	0.298	0.171	-0.139	CWED	0.939	0.038	0.042	-0.179
PROX_AM	0.168	-0.191	0.621	0.160					
PROX_CV	0.135	0.452	0.012	0.082	**Diversity**				
SIMI_AM	-0.127	-0.158	-0.313	0.777	PRD	0.733	0.320	-0.196	-0.089
SIMI_CV	0.154	0.543	0.162	-0.224	SIDI	0.829	-0.205	-0.418	0.068

Table 3.14: The loadings of 32 landscape metrics derived from the PH1 2000 on the principal components, grouped by landscape aspect. Note: variables with the highest loadings on each component have been highlighted.

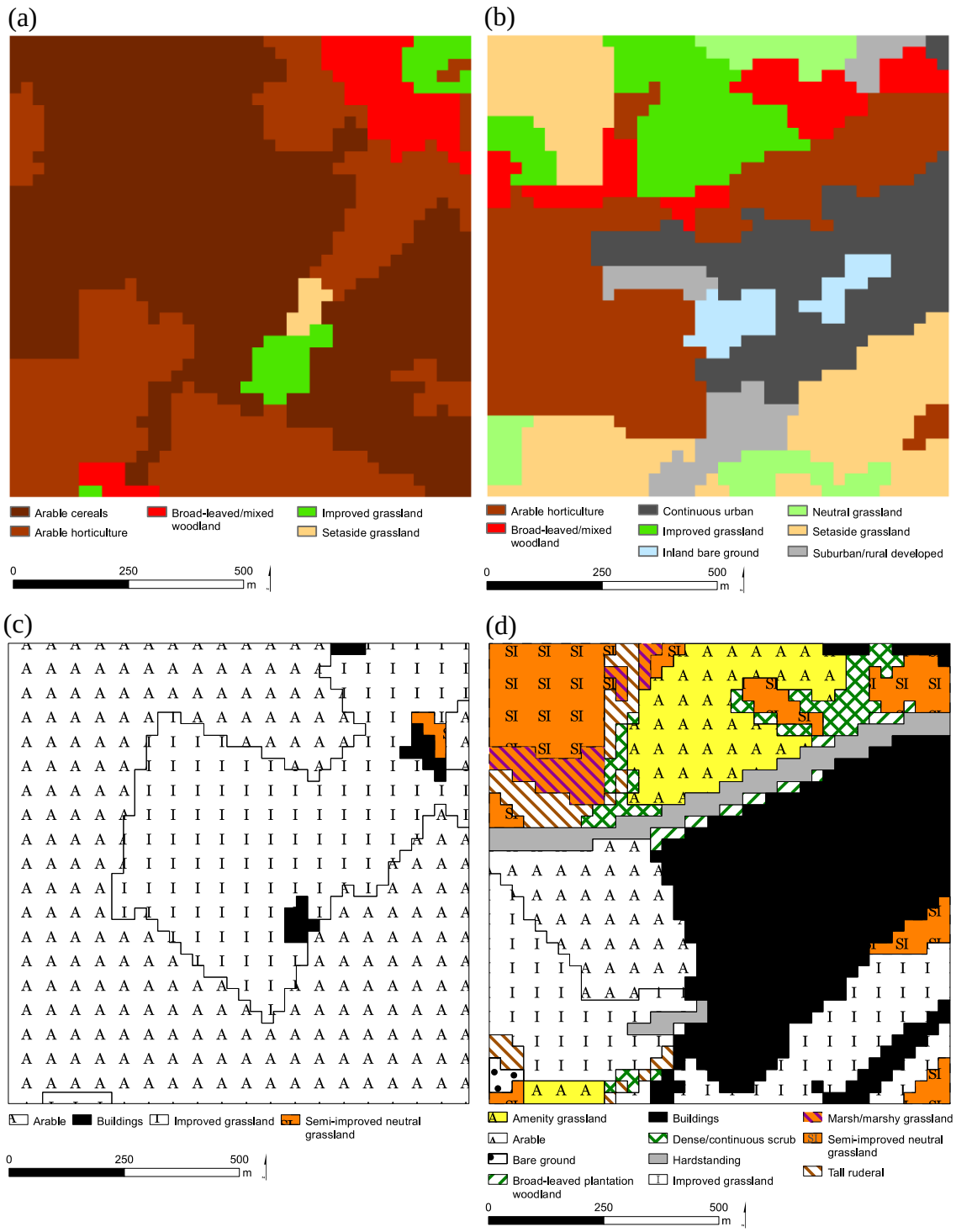


Figure 3.7: Landscape composition and configuration of two 1 km grid square landscapes, (a,c) GRID ID 3664 and (b,d) GRID ID 3047, derived from two different data sources, (a,b) LCM 2000 and (c,d) PH1 2000.

(a) LCM 2000

PC-1			PC-2		
Metric	ID 3664	ID 3047	Metric	ID 3664	ID 3047
GYRATE_AM (m)	183.02	132.19	SHAPE_MN	1.43	1.40
MESH (ha)	23.26	9.98	CIRCLE_MN	0.59	0.61
CWED (m/ha)	44.80	99.93	CONTIG_MN	0.65	0.67
SIDI	0.56	0.82	CONTIG_RA	0.75	0.56

(b) PH1 2000

PC-1			PC-3		
Metric	ID 3664	ID 3047	Metric	ID 3664	ID 3047
GYRATE_AM (m)	355.36	148.49	SHAPE_MN	1.45	1.44
MESH (ha)	52.91	13.11	SHAPE_CV (%)	23.94	26.76
CWED (m/ha)	34.60	104.81	FRAC_CV (%)	2.48	4.29
SIDI	0.46	0.82	ENN_AM (%)	186.65	161.89

Table 3.15: The values for the landscape structure metrics derived from the composition and configuration of the two 1 km grid square landscapes, GRID ID 3664 and GRID ID 3047, for a selection of metrics with high loadings which discriminate between the two landscapes. Metrics derived from two data sources are shown for (a) LCM 2000 with high loadings on PC-1 and PC-2 and (b) PH1 2000 with high loadings on PC-1 and PC-3.

3.3.4 Associations between metrics at different scales: LCM 2000 25 m – 250 m

When comparing the configurations of the loadings of the 4-D output from the PCA at 25 m (see section 3.3.3; Table 3.13) with that obtained at the three larger grain sizes (50 m, 100 m and 250 m) by means of a Procrustes Rotation, metric values become increasingly dissimilar, as reflected by increasing residual sums of squares (RSS) as the difference in grain sizes being compared increases (Table 3.16). The RSS is higher when comparing the 4D configuration from 25 m to larger grain sizes, than the RSS obtained when comparing the 4D configuration from 50 m.

Relative Euclidean similarities between the metrics obtained from the 4D output from the PCA at each scale are significantly correlated with each other at each scale combination (Table 3.17). Similarly to the pattern derived when comparing the RSS from the Procrustes rotation, the correlation coefficient decreases as the difference in the scales being compared increases (Table 3.16; Table 3.17). Correlation coefficients are slightly weaker when comparing each scale to 25 m than to 50 m or 100 m (Table 3.17).

When considering the behaviour of individual metrics, as grain size increases from 25 m, only two metrics, COHESION and GYRATE_CV, maintain relatively similar values across all scale comparisons, as indicated by consistently small projected Procrustes residuals when comparing the 4-D configuration of the loadings from each scale (Figure 3.8a-c – highlighted in green). A small number of metrics (AREA_RA, CIRCLE_AM, GYRATE_AM, and MESH) maintain similar values relative to the other metrics when comparing 25 m to 50 m, but similarity decreases with increases with scale (Figure 3.8a-c – highlighted in orange). With the exception of CIRCLE_AM, these metrics (COHESION, GYRATE_CV, AREA_RA, GYRATE_AM, and MESH) maintained small Procrustes residuals when comparing the 4-D configuration of the loadings from the NCA PCA (see section 3.2.5). A number of metrics exhibit relative dissimilarity when comparing 25 m to 50 m; FRAC_CV, IJI, PRD, PROX_AM, and PROX_CV (Figure 3.8a,b). The metrics CIRCLE_RA, CONTIG_RA and ECON_CV were relatively dissimilar when comparing 25 m to each scale (Figure 3.8a-c – highlighted in red). Of these, the metrics FRAC_CV, PROX_CV and CONTIG_RA also exhibit large residuals across scale comparisons from the NCA analysis (see section 3.2.5).

25	-			
50	0.0386	-		
100	0.2892	0.2597	-	
250	0.4284	0.3909	0.3437	-
	25	**50**	**100**	**250**

Table 3.16: Residual Sums of Squares (RSS) from the Procrustes Rotation comparison of the 4-dimension configuration of the loadings from the Principal Component Analysis (PCA) for each scale. The PCA is based on landscape structure metrics derived from the LCM 2000 for grid square landscapes.

25	-			
50	0.9242	-		
100	0.7081	0.7567	-	
250	0.5866	0.6837	0.7292	-
	25	**50**	**100**	**250**

Table 3.17: Pearson product-moment correlations between the Euclidean similarity matrices derived from the 4-dimension configuration of the loadings from the Principal Component Analysis for each scale. The PCA is based on landscape structure metrics derived from the LCM 2000 for grid square landscapes. Pearson product-moment correlations are obtained by Mantel Tests for each pairwise comparison and all correlations are significant (p<0.001).

(a)

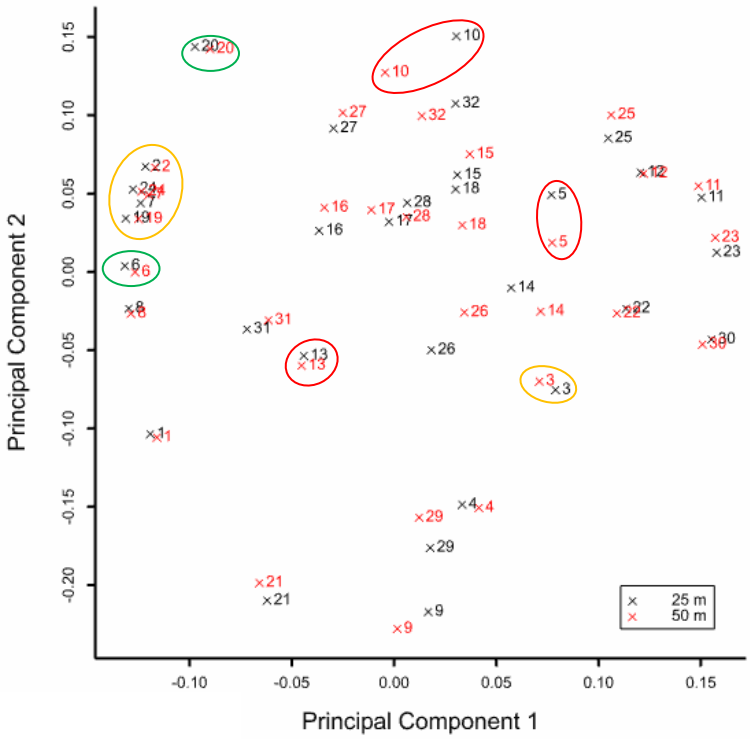
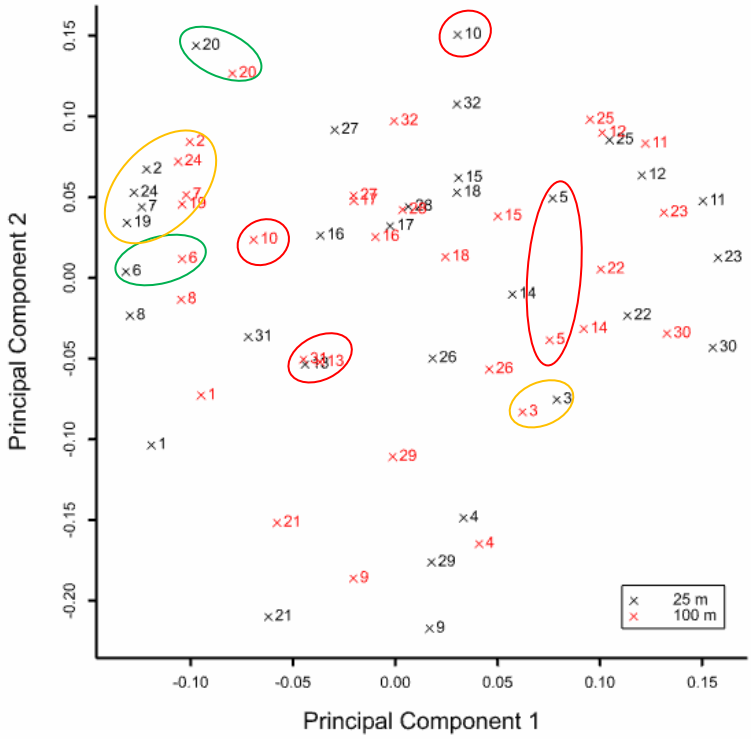


Figure 3.8 *(cont.)*

(b)



(c)

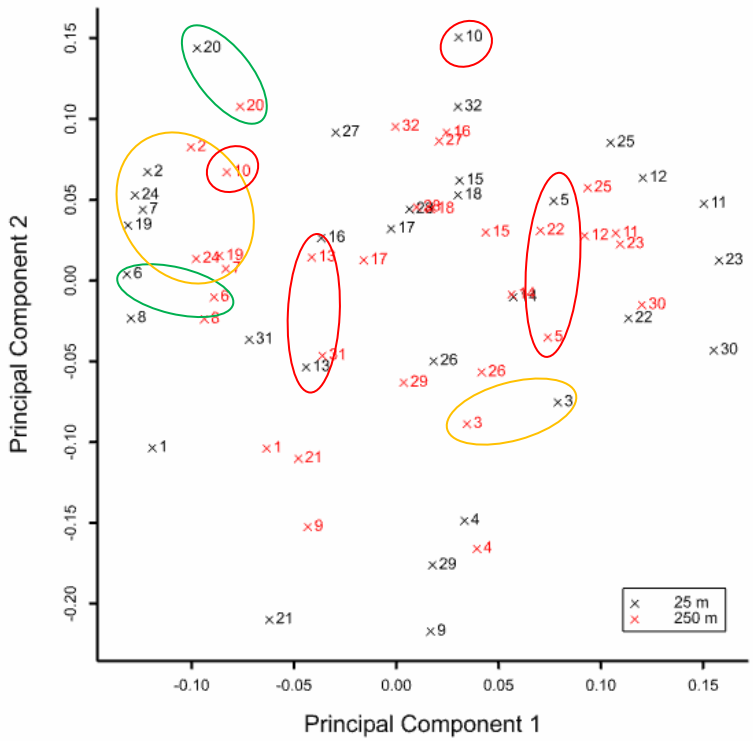
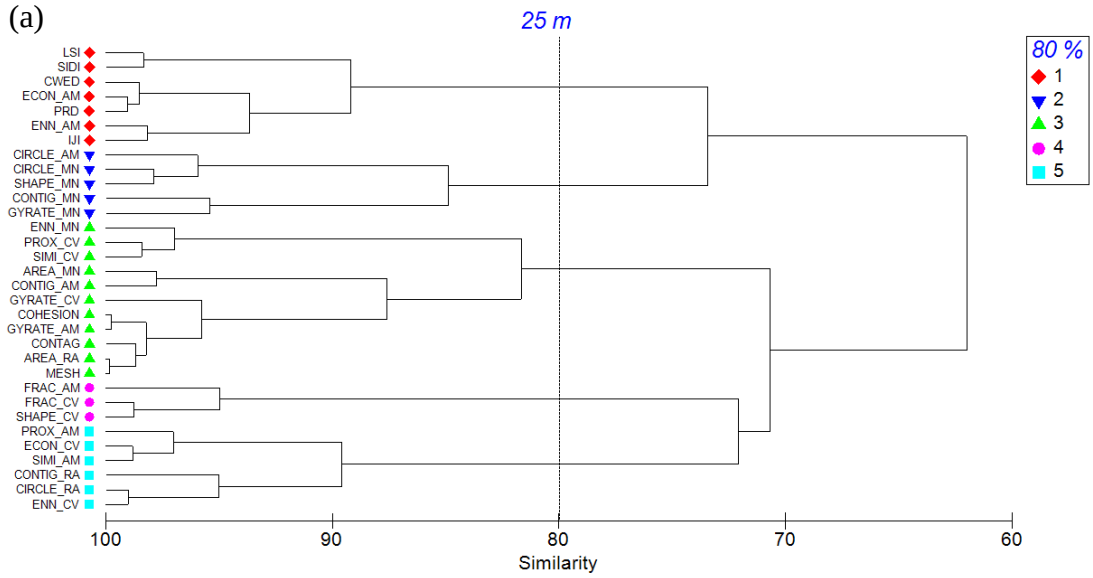


Figure 3.8 a-c: Comparison of the first two principal components (PC-1 and PC-2) of the configuration of the loadings from the LCM 2000 Principal Component Analysis (PCA) between scales; (a) 25 m compared to 50 m (b) 25 m compared to 100 m and (c) 25 m compared to 250 m. Distance between the same metric loadings plotted at two different scales represents the metric (loadings) projected residuals from the Procrustes Rotation for the first two PCs. Metrics with high Procrustes residuals across scales are circled in red; small Procrustes residuals across scales in green; and small Procrustes residuals from 25 m to 100 m only in orange. See Table 3.9 for metric codes. See Appendix A7 for 4D Procrustes Residuals.

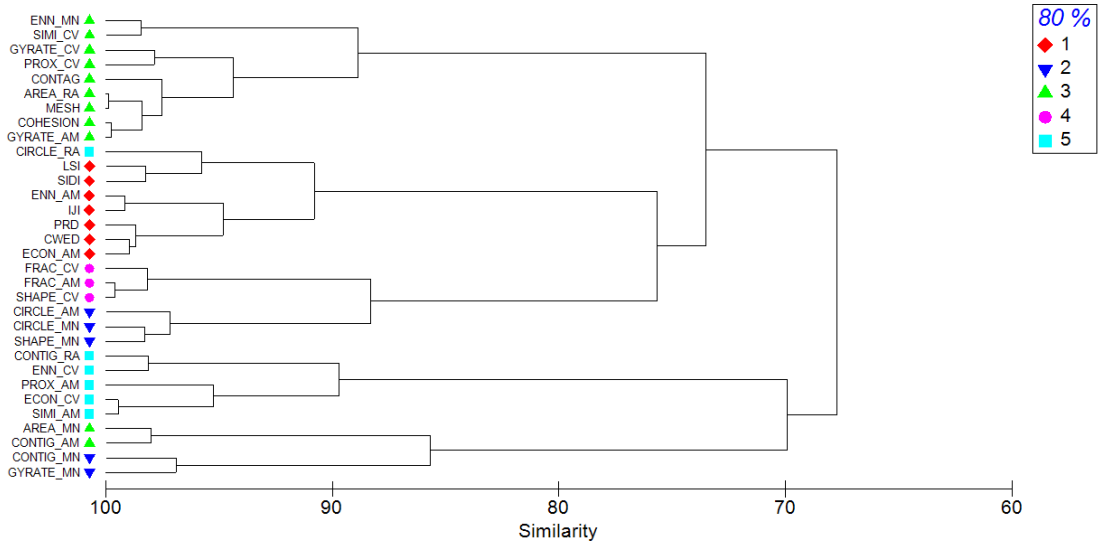
The hierarchal cluster analysis of grid square landscape structure according to the 4-D configuration of metric values obtained at a scale of 25 m suggests the partitioning of five clusters of metrics at a similarity level of 80% (Figure 3.9a). Cluster group 1 comprises seven metrics which measures patch interspersion and landscape diversity, and cluster group 2 comprises five metrics which measure average patch shape and extent. Cluster group 3 comprises 11 metrics, eight of which are metrics with high loadings on PC-1 measuring landscape fragmentation (see section 3.3.3; Table 3.13). The additional three metrics (ENN_MN, PROX_CV and SIMI_CV) further contribute towards the measure of landscape fragmentation, measuring patch aggregation and associated variability. Cluster group 4 comprises three metrics (FRAC_AM, FRAC_CV and SHAPE_CV) which directly measure patch shape complexity and variability. Cluster group 5 comprises six metrics, two of which were identified to have high loadings on PC-4 (see Section 3.3.3; Table 3.13). Metrics within this group measure variability in patch extent and patch aggregation, contributing also to measures of landscape fragmentation.

When comparing the clustering of the metrics across scales, most notably metrics comprising cluster group 1 remain grouped at each scale from 25 m (80 % similarity) to 250 m (85 % similarity) (Figure 3.9a-d). Metrics comprising cluster group 4 also remain clustered at each scale with the addition of three metrics from cluster group 2 (CIRCLE_AM, CIRCLE_MN and SHAPE_MN) at a scale of 50 m (similarity 85 %). Cluster groups 2, 3 and 5 separate at 50 m, and cluster group 2 reforms at 100 m (similarity 80 %), however clusters 3 and 5 remain separated. Within cluster group 3, however, six metrics (AREA_RA, MESH, COHESION, GYRATE_AM, CONTAG and GYRATE_CV) remain clustered at all scales (25m: 85 % similarity; 50m 90 % similarity; 100 m 90 % similarity; and 250m 85 % similarity).

(a)



(b)



(c)

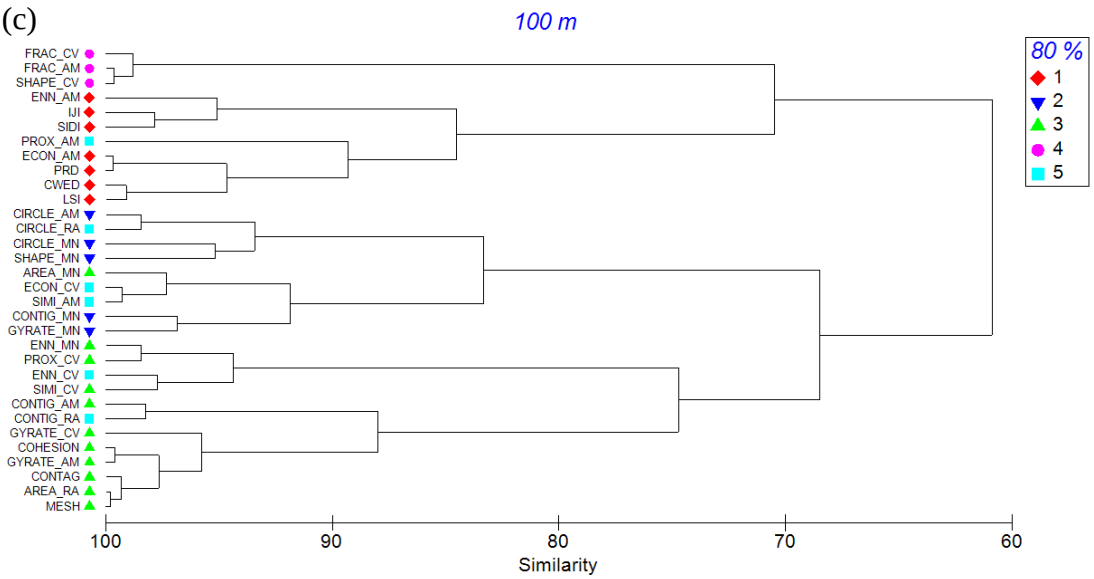


Figure 3.9 *(cont.)*

(d)

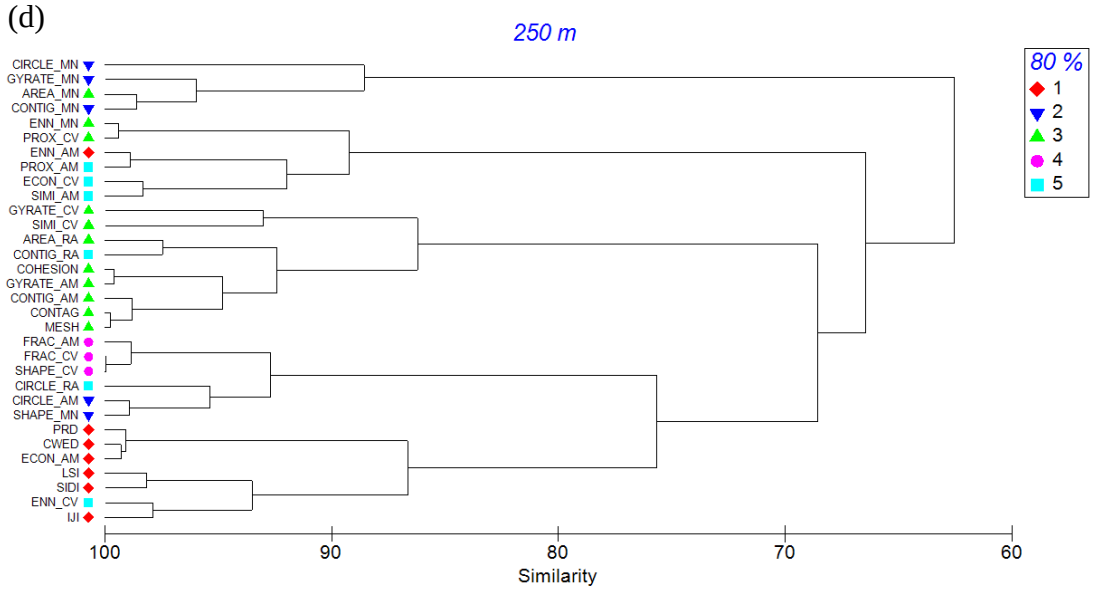


Figure 3.9a-d: Clustering of landscape structure metrics derived from the LCM 2000 using the complete link algorithm based on Euclidean similarity matrix of metric values. The clusters at 25 m (a) are defined at a similarity level of 80 % and these clusters are identified at the scales (b) 50 m, (c) 100 m and (d) 250 m.

3.3.5 Associations between metrics at different scales: PH1 2000 25 m – 250 m

For the PH1 2000, similar patterns are observed from the Procrustes Rotation comparison of the 4-D PCA configurations to that obtained when using the LCM 2000 (see section 3.3.4; Table 3.16). Metric values become increasingly dissimilar, as reflected by increasing residual sums of squares (RSS) as the difference in grain sizes being compared increases (Table 3.18). In contrast to the LCM 2000, metric values are most similar between the scales 50 m and 100 m (RSS = 0.0265).

Relative similarities between the metrics obtained from the 4D output from the PCA at each scale are significantly correlated with each other at each scale combination, with the strongest correlation between 50 m and 100 m (r = 0.9598, p <0.001) (Table 3.19). Correlation coefficients are weaker when comparing each scale to 25 m than any other scale. When considering the behaviour of metrics across scales, only four metrics (AREA_MN, CONTIG_AM, CWED, and ECON_AM) maintain relatively similar values, with small projected Procrustes residuals, when comparing the 4-D output of the PCA from each scale to that obtained at 25 m (Figure 3.10a-c – highlighted in green). In contrast the metrics ENN_CV, PROX_AM, FRAC_AM, CONTAG, IJI and SIDI maintain high Procrustes residuals when comparing each scale to 25 m (Figure 3.10a-c – highlighted in red).

25	-			
50	0.1543	-		
100	0.2390	0.0265	-	
250	0.4508	0.2458	0.1849	-
	25	**50**	**100**	**250**

Table 3.18: Residual Sums of Squares (RSS) from the Procrustes Rotation comparison of the 4-dimension configuration of the loadings from the Principal Component Analysis (PCA) for each scale. The PCA is based on landscape structure metrics derived from the PH1 2000 for grid square landscapes.

25	-			
50	0.8196	-		
100	0.7442	0.9598	-	
250	0.4639	0.6410	0.7371	-
	25	**50**	**100**	**250**

Table 3.19: Pearson product-moment correlations between the Euclidean similarity matrices derived from the 4-dimension configuration of the loadings from the Principal Component Analysis for each scale. The PCA is based on landscape structure metrics derived from the PH1 2000 for grid square landscapes. Pearson product-moment correlations are obtained by Mantel Tests for each pairwise comparison and all correlations are significant (p<0.001).

(a)

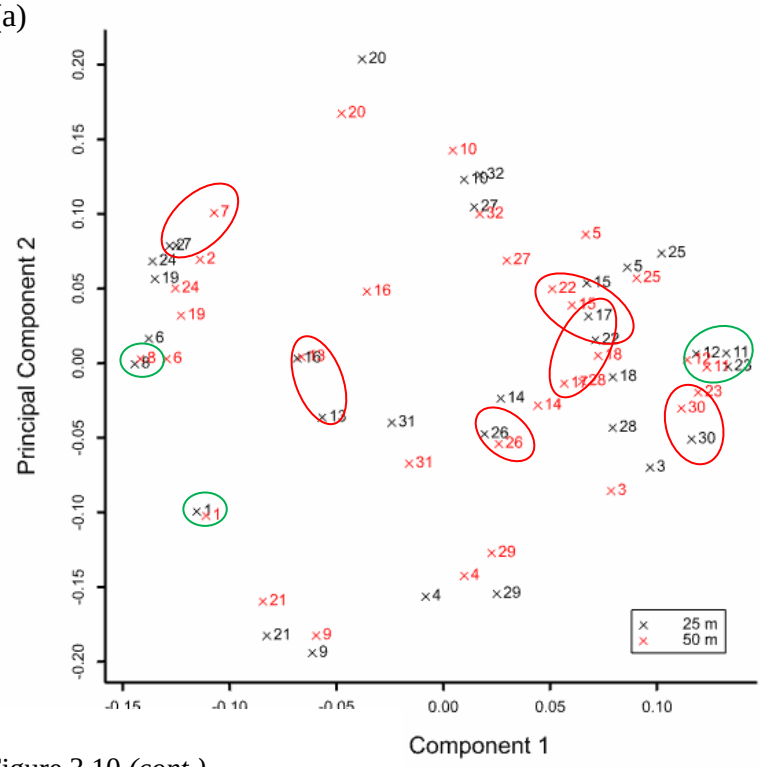
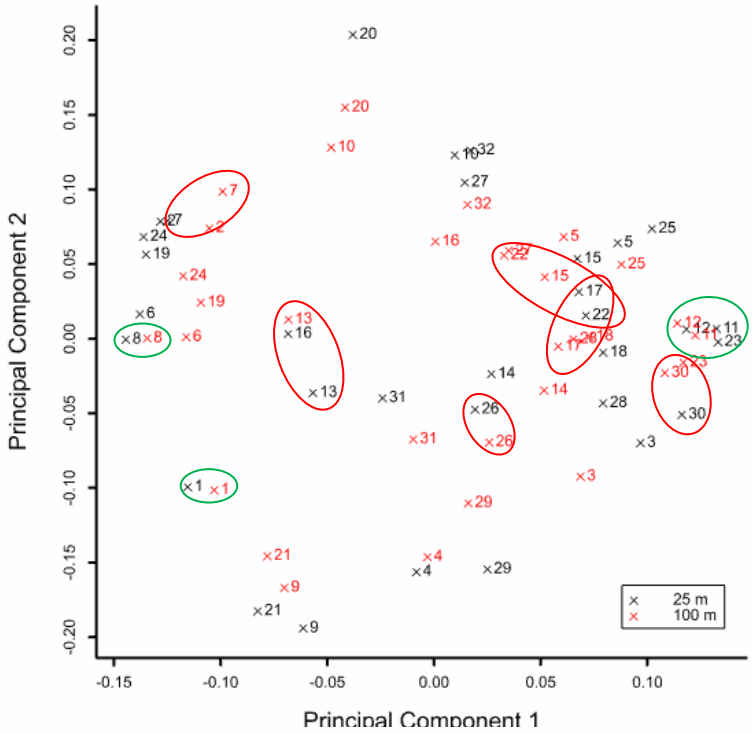


Figure 3.10 *(cont.)*

(b)



(c)

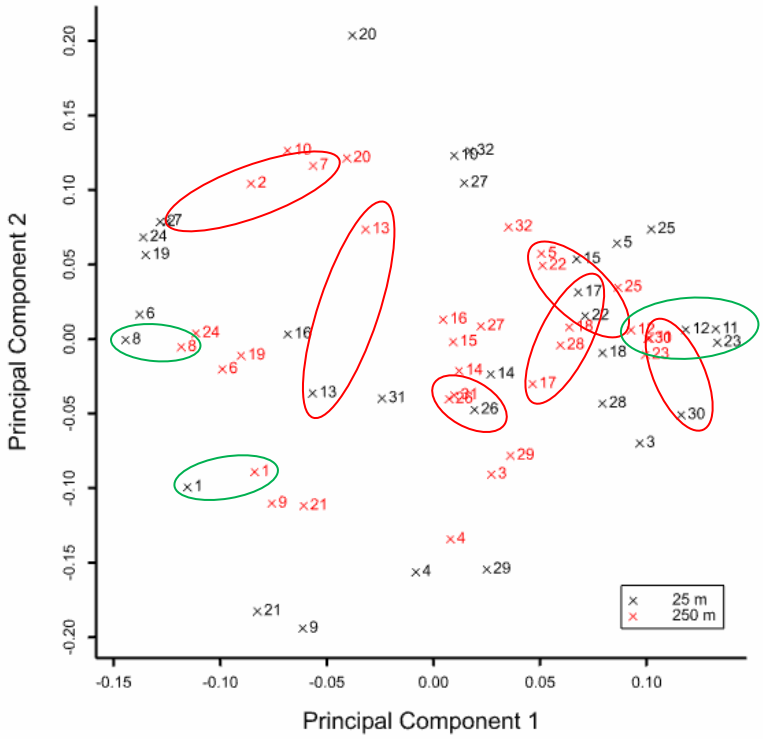
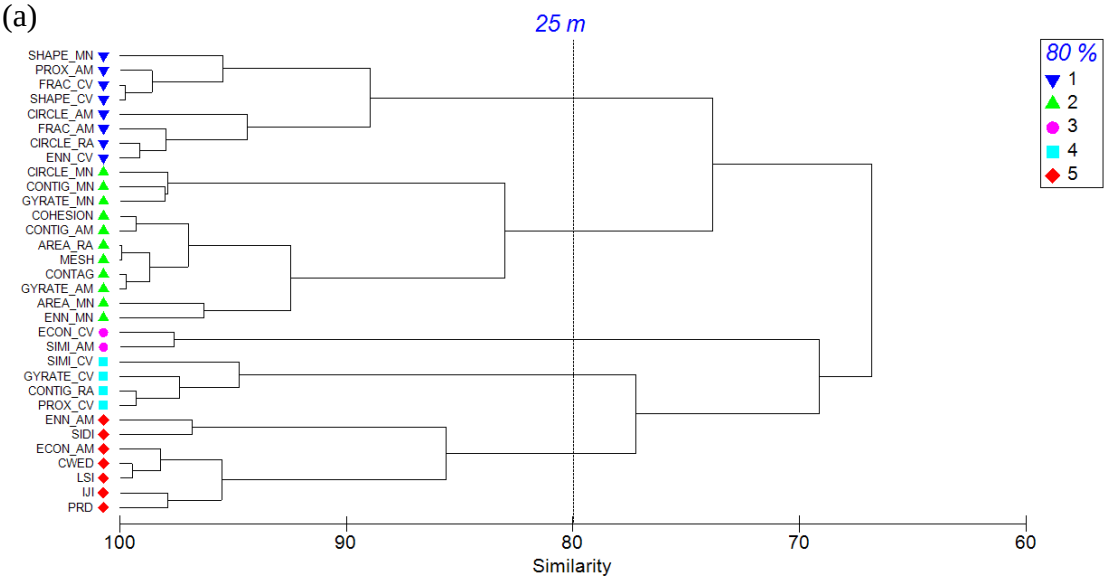


Figure 3.10 a-c: Comparison of the first two principal components (PC-1 and PC-2) of the configuration of the loadings from the PH1 2000 Principal Component Analysis (PCA) between scales; (a) 25 m compared to 50 m (b) 25 m compared to 100 m and (c) 25 m compared to 250 m. Distance between the same metric loadings plotted at two different scales represents the metric (loadings) projected residuals from the Procrustes Rotation for the first two PCs. Metrics with high Procrustes residuals across scales are circled in red; and small Procrustes residuals across scales in green. See table 3.9 for metric codes. See Appendix A7 for 4D Procrustes Residuals.

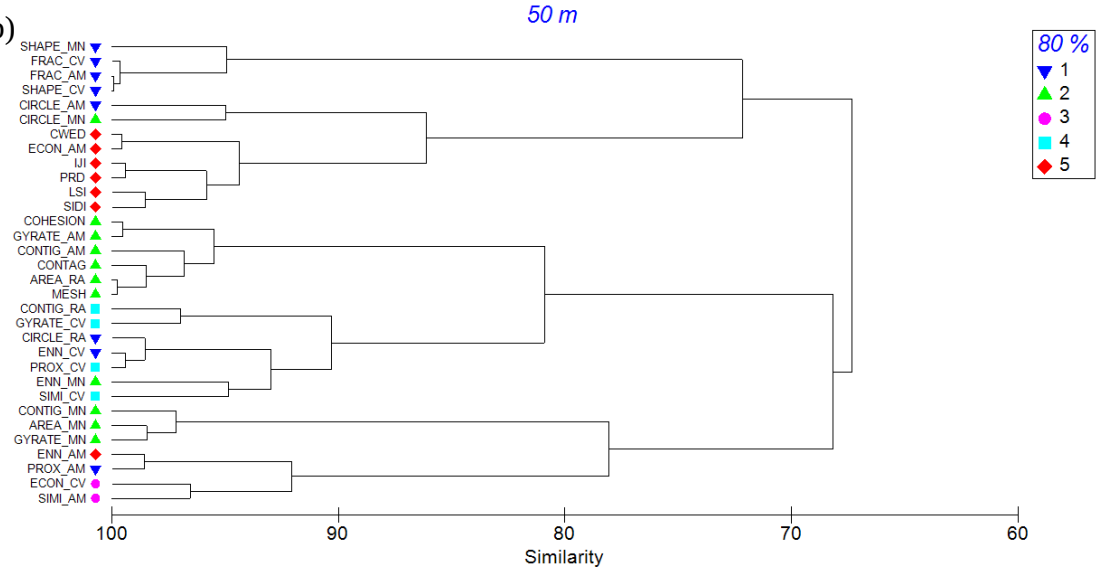
At an initial scale of 25 m the 4-D configuration of the loadings can be clustered into five groups (Figure 3.11a). Cluster group 1 comprises eight metrics, which have high loadings on PC-3 and PC-4 for the PH1 2000 and as such provide measures of patch shape complexity and variability (see section 3.3.3; Table 3.14). Cluster group 2 comprises 11 metrics, seven of which are associated with PC-1 providing measures of landscape fragmentation (Figure 3.11a; Table 3.14). The remaining metrics provide measures of patch extent (GYRATE_MN, CONTIG_MN and CIRCLE_MN) (Figure 3.11a; Appendix A4). Cluster group 3 comprises two metrics (ECON_CV and SIMI_AM) which have high loadings on PC-4, which is associated with patch shape complexity, although these two individual metrics capture the contrast between patch types (Figure 3.11a; Table 3.14; Appendix A4). Cluster group 4 comprises four metrics which have high loadings on PC-2 and as such are considered to provide measures of patch shape variability (Figure 3.11a; Table 3.14). Cluster group 5 comprises seven metrics which are associated with PC-1, PC-3 and PC-4, providing measures of landscape fragmentation and functional connectivity (Figure 3.11a; Table 3.14).

At a scale of 50 m, four of the five groups separate out joining different metrics at a similarity threshold of 80 % (Figure 3.11b). Cluster group 5 remains clustered at a similarity threshold of 90 %, with the addition of metrics from cluster group 1 and 2. In some cluster groups individual metrics remain clustered at each scale, most notably the metrics CWED, ECON_AM, IJI, PRD, LSI and SIDI from group 5; the metrics FRAC_AM, SHAPE_CV, FRAC_CV and SHAPE_MN from group 1; and the metrics MESH, CONTIG_AM, COHESION and GYRATE_AM from group 2 (Figure 3.11a-d).

(a)



(b)



(c)

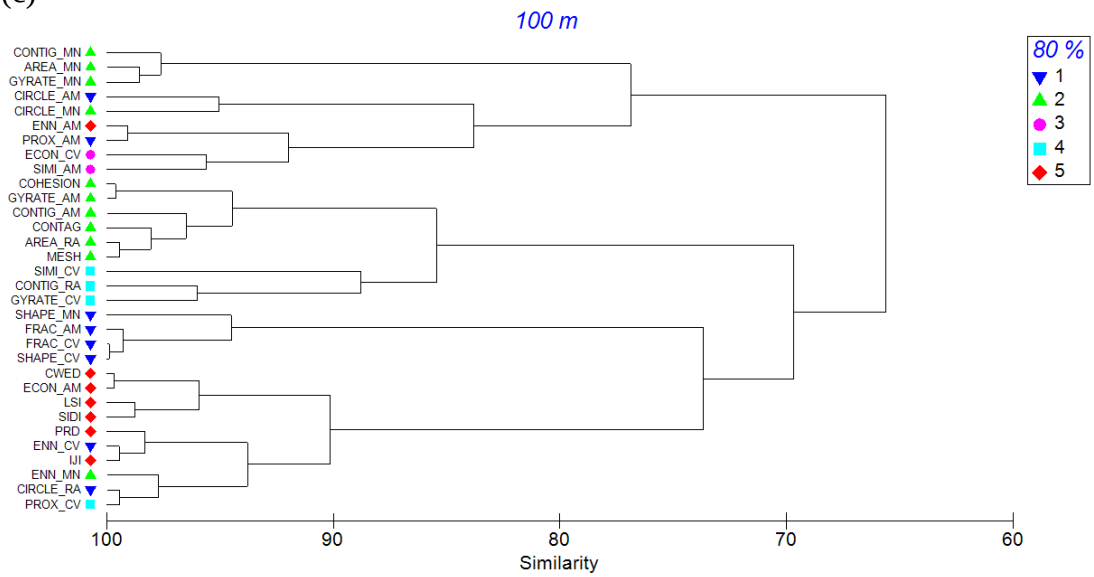


Figure 3.11 *(cont.)*

(d)

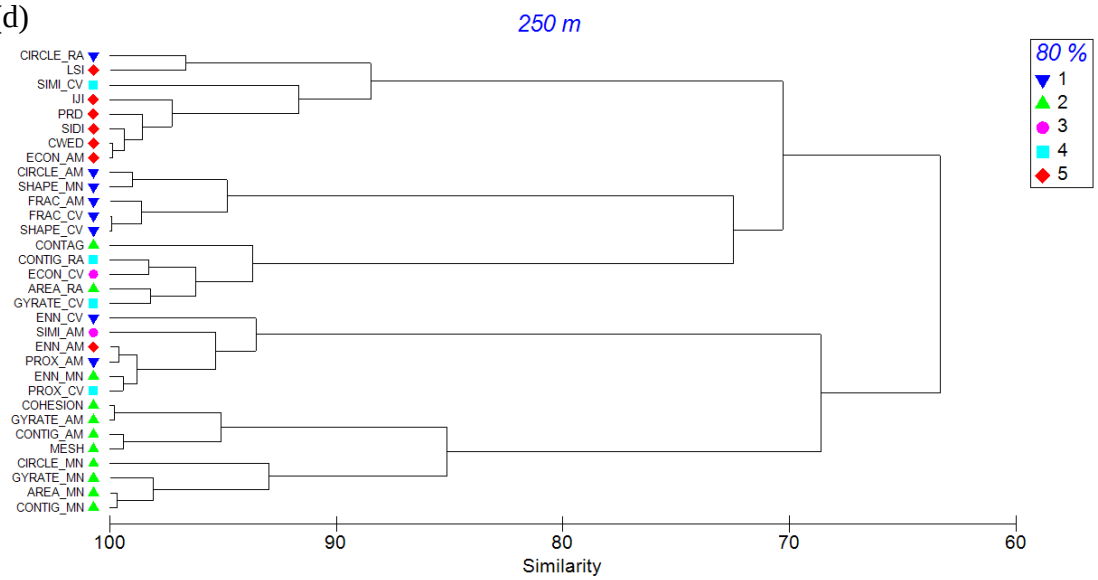


Figure 3.11a-d: Clustering of landscape structure metrics derived from the PH1 2000 using the complete link algorithm based on Euclidean similarity matrix of metric values. The clusters at 25 m (a) are defined at a similarity level of 80 % and these clusters are identified at the scales (b) 50 m, (c) 100 m and (d) 250 m.

3.4 Discussion

This chapter demonstrates the influence of spatial scale on the discrimination between landscapes delimitated by National Character Areas (NCAs) and 1 km grid squares with Warwickshire by landscape structure metrics. A number of metrics consistently discriminated between landscapes across scales, with significant differences in metric values based on the landscape characteristics mean patch size (MPS) and diversity of land covers (DLC) for both NCA and grid square landscapes derived from LCM 2000 and PH1 2000 data (Table 3.1; 3.8; 3.9). In particular, for the NCA landscapes, NCAs characterised by low landscape diversity (DLC groups 7-9) typically had larger mean patch sizes (MPS groups large and extra-large) than those NCAs with a more even distribution of area amongst land cover classes (Figure 3.1a,b). Metrics that measured landscape fragmentation were significantly different between NCAs characterised by extra-large MPS in comparison to those characterised by small and medium sized patches (Table 3.1). In particular, NCAs with extra-large MPS were characterized by a larger range in patch extent, with patches being more clumped and aggregated, and as such less fragmented. Furthermore, the distribution of area between land cover classes was significantly less equitable in these sites, as measured by the Simpsons Diversity Metric (SIDI) (Table 3.1; Figure 3.1b).

A limited number of metrics discriminated between NCAs on the basis of the landscape characteristics total area (TA) or number of land cover classes (NLC). This suggests that there was limited variability in the number of land cover classes to detect differences in metric value and that similar values were obtained across all ranges of landscape extent. The influence of MPS on metric values is therefore more important than total area, contradicting Wu *et al.,* (2002) who found that for all 20 metrics considered in their study, metric values were influenced by landscape extent, only seven of these metrics exhibiting a predictable response to changes in extent.

For the PH1 2000 grid square landscapes a higher number of metrics discriminated between landscapes on the basis of MPS, DLC and NLC, in comparison to metrics derived from the LCM 2000. This greater discriminatory power associated with the PH1 2000 metrics could be due to the higher level of detail and number of land cover classes identified during the production of this data set in comparison to the LCM

2000, and consequently a greater variation in landscape characteristics amongst grid squares. This is further supported by the significant differences observed in metric values derived from the PH1 2000 between all four NLC groups of grid square landscapes. These findings are in accordance with the literature, as thematic resolution has been found to directly influence the size distribution of patches, with increased patch density associated with increasing number of land cover classes (higher thematic resolution), and in turn a reduction in mean patch size (Castilla, *et al.*, 2009). When considering the influence of DLC on metric values for both PH1 2000 and LCM 2000, significant differences were observed for metrics between DLC groups. These significant differences, however, for most metrics could not be attributed to differences in landscape diversity between the groups and are likely due to differences in the MPS between DLC groups. The discriminatory power of the landscape metrics was maintained across all scales, in particular for the metrics derived from the PH1 2000.

Metric values change with increases in grain size, and studies by Wu *et al.*, (2002) and Wu (2004) identified that for some metrics (LSI, AREA_MN, FRAC_AM, and PRD) this response to scale can be predicted (Type I and Type II response) but for the metrics SHAPE_MN and CONTAG this response is unpredictable (Type III response). The results from this chapter demonstrate that for grid square and NCA landscapes, despite the changes in metric values, most metrics maintain their discriminatory power, even the type III metrics considered to be unpredictable by Wu (2002; 2004).

Metrics identified to be important individually for discriminating between landscapes, particularly between NCA landscapes, were also found to exhibit discriminatory power when the combined variability in landscape structure was considered for all 32 metrics. Metrics with high loadings on PC-1 from the Principal Component Analysis (PCA), were similar across the two landscape extents (NCAs and Warwickshire grid squares) and two data sources (PH1 2000 and LCM 2000), with highest correspondence between NCA landscapes and Warwickshire grid squares derived from LCM 2000.

From the PCAs, each of the first four components comprise a combination of variables with high loadings, which are associated with different aspects of landscape

structure, according to the classification of metrics by HainesYoung & Chopping (1996) and McGarigal *et al.*, (2012) (see section 1.6.2 and Table 3.3). Although metrics with highest loadings on each component are associated with different aspects of landscape structure the computation of the metric values are influenced by the same underlying variation in landscape structure. In particular, PC-1 for NCA and grid square landscapes (PH1 2000 and LCM 2000) can be considered to provide a measure of habitat fragmentation, which incorporates the subdivision, dispersion and interspersion of patches (see section 3.2.3 and 3.3.3). Of the nine metrics common to PC-1, the metrics COHESION and CONTAG provide measures of the dispersion and compaction of patches within the landscape and are influenced by patch size (Appendix A4) (McGarigal, *et al.*, 2012). The metrics SIDI and CWED provide indirect measures of patch interspersion and landscape continuity, and are influenced by neighbouring patch types, incorporating the landscape mosaic, as well as being influenced by patch size or edge density (Appendix A4) (McGarigal, *et al.*, 2012). The metric MESH directly provides a measure of subdivision, and additionally the metrics GYRATE_AM, which measures the continuity of the landscape, and CONTIG_AM, which measures the connectedness of cells within a patch, provide indirect measures of subdivision. Differences were observed between the two landscape data sets (PH1 2000 and LCM 2000) in terms of the metrics with highest loadings on PC-3 and PC-4, and consequently different grid squares were discriminated on the basis of these components.

With increases in grain size, metrics contributing to the variation measured in the first four principal components for the NCA and grid square landscapes (LCM 2000 and PH1 2000) become increasingly dissimilar relative to these for 25 m, as measured by the residual sums of squares between the 4-D configurations of the loadings from the Procrustes rotation. Furthermore, relative dissimilarity is greatest when comparing the 4-D configuration of the loadings and the scores from larger scales to 25 m, than in comparison to any other scale. This dissimilarity between 25 m and larger scales occurs because as the grain size increases the representation of the landscape becomes generalised and differs from the initial landscape pattern observed at 25 m (Simova and Gdulova, 2012). Once there is an initial loss of information with a grain size of 50 m, landscapes compared at coarser grain sizes are more similar as the patterns within them have been simplified. In particular, for the

NCA landscapes relative similarity was greatest when comparing 250 m and 500 m (Table 3.6).

High variation in metric behaviour was identified for the metrics FRAC_CV, PROX_CV and CONTIG_RA which exhibited high Procrustes residuals across all scale comparisons for the NCA landscapes and LCM 2000 grid square landscapes. Metric behaviour differed between the two sources of landscape data for the grid square landscapes, with different metrics maintaining high and low Procrustes residuals across scales. As such, the impact of scale on the behaviour of metrics differed between the two data sources.

The clustering from the 4-D configuration of the loadings, i.e. the contribution of the metrics to explaining the variability in the landscapes, is similar between the NCA and grid square landscapes, and the two data sources (PH1 2000 and LCM 2000) at a scale of 25 m. In particular the metrics SIDI, ECON_AM, CWED, LSI, and IJI are clustered together for the NCA and grid square landscapes, are associated with PC-1, and maintain consistent relationships across scales from 25 m to 250 m. This suggests that the interspersion of patches and the relationship between neighbouring patch types remains stable across different scales. The metric IJI has been previously identified to have a predictable response to an increase in grain size, and SIDI to be insensitive to changes in grain size (Baldwin, *et al.*, 2004). Metrics associated with PC-2 for the NCA landscapes measuring patch shape, and a small number of metrics associated with patch shape for the grid square landscapes (FRAC_AM and SHAPE_CV) also exhibited consistent relationships with each other across scales. Garcia-Feced *et al.*, (2010) and Saura (2004) found that even though patch shape metrics (SHAPE_MN and ED) were fine grained metrics requiring a scale of 25 m for differences to be detected in patch shape between landscapes, the shape metrics considered in their study maintained consistent discriminating ability across scales from 25 m to 500 m. The findings from this chapter support Garcia-Feced *et al.*, (2010), and contradict general consensus in the literature of an erratic response of shape metrics to increases in grain size, particularly for different landscapes (Simova and Gdulova, 2012; Wu, 2004; Wu, *et al.*, 2002).

Consistent relationships between metrics is also evident for the metrics COHESION, CONTIG_AM, MESH, CONTAG, GYRATE_AM, AREA_MN, and ENN_MN

which are clustered together for NCA and grid square landscapes at a scale of 25 m, and these metrics, with the exception of ENN_MN, are associated with PC-1, providing measures of habitat fragmentation. For the NCA landscapes, the relationships between these metrics remain consistent across scales, however, for the grid square landscapes derived from both LCM 2000 and PH1 2000 this cluster breaks at 50 m, although subsets of the metrics of this group remain clustered. This included the metrics COHESION, MESH, and GYRATE_AM for PH1 2000 and LCM 2000 and additionally the metrics AREA_RA, CONTAG and GYRATE_CV for the LCM 2000 data. These metrics are directly influenced by patch size, capturing the variability and compaction of patches and as such are likely to be influenced similarly across scales. For both data types, however, the metric AREA_MN remains separated from these metrics only re-joining the initial cluster for the PH1 2000 landscapes at a scale of 250 m. Although previously thought to exhibit a predictable response with scale (Simova and Gdulova, 2012), inconsistent behaviour with scale has been shown for AREA_MN in the study by Garcia-Feced *et al.*, (2010).

With the inconsistent relationship between some metrics at a scale of 50 m compared to 25 m, different patterns are detected amongst the landscapes. This is most evident for the NCA landscapes clustered within cluster group 4 at a scale of 25 m (Figure 3.4a). These NCAs have similarly negative scores on PC-1, and as such are characterised by low values associated with the metrics which measure patch aggregation. At a scale of 50 m different patterns are detected within the NCAs comprising cluster group 4 with the partitioning of the group according to those NCAs which differ in their patch boundary configuration as measured by the metrics with highest loadings on PC-3, which exhibit greatest variability at 50 m (Figure 3.4b; Figure 3.6b).

3.3.1 Conclusion

For the NCA and grid square landscapes, several metrics, most notably the metrics which measure landscape fragmentation with the highest loadings on PC-1, and measure patch shape (with the highest loadings on PC-2 for NCA landscapes), are robust across scales, maintaining similar relationships and discriminating ability. These results suggest landscape patterns can be detected at coarse grain sizes and that metrics can discriminate between landscapes at a range of scales. Furthermore, when

comparing the effects of scale between the two different landscape data sources, higher discriminatory power was associated with the landscape structure metrics derived from the PH1 2000 in comparison to the LCM 2000. The impact of scale on the behaviour of individual metrics differed between the two data sources; however similar consistent relationships between metric values were obtained across scales, in particular for the landscape fragmentation metrics. It is evident, however, that similarity in landscapes at larger scales occurs due to the loss of information, and the landscape patterns detected at 50 m are different to that at 25 m. Furthermore, for a range of metrics, measuring differing landscape aspects, inconsistent behaviour was detected across scales. When considering the most appropriate grain size to discriminate between landscape patterns, a range of scales should be considered to incorporate the differing response of landscape structure metrics to changes in spatial scale.

For developing relationships with biodiversity it is imperative to consider a scale which is relevant to the organism of interest and which captures the functional grain size of the landscape, i.e. spatial heterogeneity within species perceptual range (Baguette and Van Dyck, 2007; McGarigal and Marks, 1995; Tischendorf and Fahrig, 2000). For species with limited daily dispersal capabilities, such as butterflies which are generally characterised by daily movements ranging from 200 – 600 m (Davis, *et al.*, 2007), grain sizes which capture maximum landscape variability are likely to be most appropriate. Landscape elements within the range of these dispersal capabilities have been shown to determine presence-absence, as well as butterfly abundance and species richness (Flick, *et al.*, 2012; Rossi and van Halder, 2010). For example, Flick *et al.*, (2012) found that butterfly species richness was related to landscape patterns detected at a spatial extent of 250 m, with heterogeneous landscapes comprising high patch density and patch richness most important in agriculturally dominated landscapes, facilitating the use of complementary resources by different butterfly species. As such detecting a wide range of land cover types, as well as patch area and edge effects are important for developing relationships between landscape structure and butterflies (Rossi and van Halder, 2010; Schneider and Fry, 2001). The loss of information, and a reduction in the number of classes at coarse scales (Simova and Gdulova, 2012) would therefore be detrimental. Imprecision in land cover data has been found to be a limiting factor in developing

models to predict butterfly species richness and abundance (Flick, *et al.*, 2012). Overall, a scale of 25 m would not only be most appropriate for developing relationships with butterflies but also for detecting the underlying patterns within the landscape.

Chapter 4: Predicting butterfly presence-absence as a function of landscape components

4.1 Introduction

Predictive models of species distribution are valuable tools in the absence of adequate species data (Lawler, *et al.*, 2011). They are widely used for identifying suitable conservation sites and biodiversity hotspots (Baguette, *et al.*, 2000; Cabeza, *et al.*, 2004; Heikkinen, *et al.*, 2007), predicting the distribution of rare species (Heikkinen, *et al.*, 2007; Raxworthy, *et al.*, 2003), and identifying the potential impact of changes in land use and climate (Allouche, *et al.*, 2006; Araujo, *et al.*, 2005; Vaughan and Ormerod, 2005). Probabilities of occurrence derived from species distribution models can be used to identify sites of high suitability and facilitate comparison between different sites, often through the production of species distribution maps using a geographical information system (GIS) (Hirzel, *et al.*, 2006; Schroeder and Richter, 2000; Vaughan and Ormerod, 2005). Statistical models have the potential to provide insight into species-habitat associations and/or the underlying ecological processes which govern species distribution (Robinson, *et al.*, 2014). This understanding is fundamental for effective conservation and sustainable land management, particularly when considering current and projected rates of anthropogenic modification of the landscape (Manel, *et al.*, 2001; Robinson, *et al.*, 2014; Rushton, *et al.*, 2004).

Predictive models of butterfly species have been developed in response to environmental variables, such as landscape composition; however many of these studies have developed models for single species only (Schweiger *et al*, 2006, Heikkinen *et al*, 2007). Furthermore they often only consider the characteristics of habitats classified as suitable, neglecting to consider the functional connectivity of the landscape, and the landscape complementation nature of butterflies (Schweiger, *et al.*, 2006; Shreeve and Dennis, 2011). Drawing conclusions on important landscape variables between species-specific models can be difficult due to the differences in spatial scales and land cover classifications used. The importance of particular landscape elements vary depending on the thematic resolution and classification system from which they are derived (Rossi and van Halder, 2010; Turner, *et al.*, 2001). A few studies have developed a series of species-specific models using the same explanatory variables, for example, Cowley *et al.*, (2000) developed species specific landscape based models for 26 butterfly species and

Luoto *et al.*, (2006) developed separate models for 98 species. Whilst considering several different species, both these studies found that model accuracy and applicability differed between species, with varying importance of explanatory variables across species. The varying response of butterfly species to landscape pattern has been widely discussed within the literature (Dover and Settele, 2009; Shreeve and Dennis, 2011), and different landscape components have been identified as important for species with contrasting dispersal capabilities (Baguette, *et al.*, 2000; Cozzi, *et al.*, 2008).

There is a clear need for the development of landscape based models based on a standardised landscape classification system to ensure transferability to new regions and derive important conclusions on landscape management. In addition, consideration of the differing behavioural responses amongst species to landscape elements is vital. The grouping of butterfly species with similar ecological attributes as developed by Shreeve *et al.*, (2001) provides an opportunity to capture the behavioural responses of multiple species, as species grouped by their ecological attributes will respond and behave similarly to landscape pattern.

When developing a model, previous studies have identified three main components which I will introduce in relation to the current study; **(1)** data of species occurrence and explanatory variables; **(2)** mathematical model that relates the species data to the explanatory variables; and **(3)** an assessment of model performance (Rushton, *et al.*, 2004).

(1) Data for the parameterisation of distribution models

The wealth of butterfly data collected by the UK Butterfly Monitoring Scheme (UK BMS), is desirable for the development of species distribution models, particularly as the standardised methodology of Pollard transect counts ensures that data collected for differing purposes can be considered in a wide range of statistical analyses including distribution modelling. Despite the standardisation of the methodology of data collection, site location is not fully representative of the habitats across the UK, with a bias towards sites of nature conservation or statutory protection (Asher, *et al.*, 2001). As a result, wider countryside and potentially ‘unfavourable’ sites can be underrepresented in the transect count data. However, recent initiatives led by the

UK BMS and Butterfly Conservation, aim to address this issue including the 'Big Butterfly Count' and the 'Garden Butterfly Survey' (Butterfly Conservation, 2014). In particular, the Wider Countryside Butterfly Survey, launched in 2009 in partnership with the British Trust for Ornithology (BTO), involves two visits to randomly selected 1 km grid squares across the UK to provide a more representative distribution of butterfly species within different habitats (UKBMS, 2014). The Butterfly Conservations general recording scheme, 'Butterflies for the New Millennium' (BNM), facilitates casual recording of butterfly sightings, preferably with four repeat visits to a site across the flight period. Combining data sets from the UKBMS, BNM and general recording schemes available from local Biological Records Centres therefore provides a more widespread coverage of the occurrence of butterfly data. The accuracy of abundance data and the potential for predicting abundance is limited due to differences in data collection methodologies, however, increased spatial distribution of butterfly records provides an indication as to whether a site is likely to be utilised, and this occurrence data can be used in models of species presence-absence. Although data from transects counts and the general recording scheme do not necessarily provide records of species absences, pseudo-absences or inferred absences can be generated using patterns within the existing data set (Hirzel, *et al.*, 2006). The reliability of inferred absences depends on the detectability of the species, local abundance and the sampling design (Hirzel, *et al.*, 2006).

At a local scale habitat quality has been proven to be important for determining species presence-absence (Schweiger, *et al.*, 2006), however, obtaining data on habitat quality, reflecting the resource requirements of a species, can be just as costly and labour intensive as obtaining data on the species of interest itself (Fleishman, *et al.*, 2003; Schweiger, *et al.*, 2006). Advances in GIS and the increasing availability of topographic, land cover and climatic data over large spatial extents has played a major part in the rapid development of species distribution models over the last two decades (Fleishman, *et al.*, 2003; Heikkinen, *et al.*, 2007; Rushton, *et al.*, 2004; Schweiger, *et al.*, 2006).

Landscape data can be used to obtain measurements on the composition and connectivity of different land cover classes, as well as overall landscape structure. Landscape composition, connectivity and structure are widely recognised as important factors in determining species distribution (see section 1.6) (Kadoya and Washitani, 2011; Rossi and van Halder, 2010). It is important to assess how models developed using landscape components obtained from satellite derived land cover data, such as the Land Cover Map (2000), compare to more precise habitat based data which provide an indication as to the quality of those habitats, for example the Phase 1 habitat mapping technique. Such a comparison will facilitate the identification of the level of precision required to capture relationships between landscape pattern and butterfly species distribution. It is important that the landscape data used corresponds to the scale at which butterfly species respond to their environment, in terms of both data resolution and level of precision (definition of land cover types) (Dover and Settele, 2009; Kumar, *et al.*, 2009; Merckx and Van Dyck, 2007; Rossi and van Halder, 2010; Turner, *et al.*, 2001). Results from chapter 3 revealed that a grain size of 25 m was most appropriate for detecting landscape structural patterns within the perceptual range of butterfly species. The resolution and precision of landscape data should also be transferable in space and time, and both the LCM 2000 in particular and the Phase 1 habitat map are widely used and accessible across the UK.

(2) Model types

There are two main approaches to modelling species distributions: empirical models, which are based on correlative relationships, and mechanistic models, which are process based, considering species ecological and life history traits (Dormann, *et al.*, 2012; Lawler, *et al.*, 2011; Manel, *et al.*, 2001). Although gaining popularity in the field of population modelling, mechanistic models require a vast amount of data for their development which is often limited in availability, such as population dynamics, dispersal capabilities and functional traits (Dormann, *et al.*, 2012; Lawler, *et al.*, 2011). Empirical models are often developed using Generalised Linear Models (GLMs), with logistic regression the most common GLM technique for species distribution models, applying a logit-link function and a binomial error structure (Lawler, *et al.*, 2011; Manel, *et al.*, 2001; Rushton, *et al.*, 2004). Logistic regression

models are widely used to statistically assess the relationship between known environmental variables and the presence-absence of species (presence-absence models), in order to infer patterns in species-habitat associations (Dormann, *et al.*, 2012; Hirzel, *et al.*, 2006). Logistic regression is a widely used modelling approach as presence-absence data is easy to collect (Rushton, *et al.*, 2004). The probability of occurrence derived from logistic regression models can provide an indication of the suitability of sites for supporting species, and the application of a threshold can be made to predict presence-absence.

(3) Assessment of model performance

Assessment of model accuracy is important for evaluating the discriminating ability of the model, the quality and efficacy of the model, as well as identifying areas for improvement, in particular the importance of particular parameters for improving overall model fit (Allouche, *et al.*, 2006; Vaughan and Ormerod, 2005). The two main approaches for the assessment of model accuracy for logistic regression models involve construction of a confusion matrix and the use of receiver operating characteristics (ROC) (Lawler, *et al.*, 2011). The confusion matrix, a widely applied approach, includes calculating the proportion of correctly predicted presences (sensitivity) and absences (specificity) (Allouche, *et al.*, 2006). Determination of model sensitivity and specificity involves the specification of a threshold for classifying presence-absence from the predicted probabilities of occurrence. Choice of threshold has been subject to much debate, and is often subjective and arbitrary. Because of this, threshold dependent methods have been heavily criticised (Manel, *et al.*, 2001). It has been suggested, however, that threshold choice should reflect the prevalence (proportion of presence data) within the data set (Lobo, *et al.*, 2008).

The use of ROC has become a favoured approach for assessing model accuracy, particularly in relation to species distribution modelling using logistic regression (Allouche, *et al.*, 2006; Fielding and Bell, 1997; Manel, *et al.*, 2001; Rushton, *et al.*, 2004). The area under the ROC curve (AUC) is widely used as a threshold independent measure of model performance and is determined by plotting the proportion of true positives (sensitivity) against the proportion of false positives (1-specificity), calculated from the entire range of thresholds (0-1) (Allouche, *et al.*, 2006; Rushton, *et al.*, 2004). Not only is the AUC method threshold independent, but

it is also independent of species prevalence (the proportion of presence data used to derive the model), and as such provides a useful measure of model discrimination (Allouche, *et al.*, 2006; Manel, *et al.*, 2001). Vaughan & Ormerod (2005) recommend that when the presence-absence of species is to be determined from predicted probabilities of occurrence then both threshold independent and dependent measures should be used for the assessment of model performance.

4.1.1 Aims

The aims of the work reported in this chapter are to: -

1. Develop logistic regression models to predict the presence-absence of butterfly species across Warwickshire as a function of landscape composition, landscape connectivity and landscape structure.
2. Identify ecologically relevant species-landscape associations from landscape based models developed in (1) for all butterfly species and butterfly Ecological Attribute Groups (EAGs).
3. Compare the performance of the different landscape based models developed in (1); landscape composition, landscape connectivity and landscape structure.
4. Compare the predictive power of Phase 1 habitat map 2000 and the Land Cover Map 2000 for the models developed in (1).

Work on these four aims will test the hypothesis that the composition, connectivity and structure of landscapes can be used to predict the presence-absence of butterfly species and butterfly Ecological Attribute Groups (EAGs); a combination of these measures will produce the best predictive model.

4.2 Results

4.2.1 Distribution of butterfly species across Warwickshire: 1990-1999

Aggregated butterfly species data across the years 1990 – 1999 comprises a total of 36 butterfly species, of which the meadow brown (*Maniola jurtina*) has the highest total abundance (11463 individuals), and the small heath (*Coenonympha pamphilus*) has the greatest spatial distribution occurring in 330 grid squares (Table 4.1). A total of 21 species comprised a total abundance of more than 1000 individuals per species across this time period, however, only six of these species were distributed across 50 or more grid squares; Small Heath *(Coenonympha pamphilus)*, Wall *(Lasiommata megera)*, White-letter Hairstreak *(Satyrium w-album)*, Grizzled Skipper *(Pyrgus malvae)*, Dingy Skipper *(Erynnis tages)* and White Admiral *(Limenitis Camilla)* (Table 4.1). In contrast, most notably the purple hairstreak *(Neozephyrus quercus)* comprised a total abundance of 4578 individuals over this time period, with a small spatial distribution, occurring within only two grid squares.

Just under half of the butterfly species (15 species) are classified within Ecological Attribute Group 1 (EAG1), which is mainly characterised by species associated with open grasslands, and these species were distributed across a total of 412 grid squares (Table 4.1; Table 4.2). The migrant group (EAG0) comprised the smallest number of species, total abundance and spatial distribution (Table 4.2). Within the 515 occupied squares, only four of these grid squares contain species belonging to all five ecological attribute groups.

Species	EAG	Total Abundance	Distribution
Aglais io	4	3890	6
Aglais urticae	4	1205	6
Anthocharis cardamines	4	433	6
Aphantopus hyperantus	1	6080	6
Argynnis paphia	2	32	2
Aricia agestis	3	47	3
Boloria euphrosyne	1	4	2
Boloria selene	1	122	4
Callophrys rubi	2	129	4
Celastrina argiolus	2	196	5
Coenonympha pamphilus	1	9543	330
Colias croceus	0	29	2
Cupido minimus	3	1585	9
Erynnis tages	1	2830	63
Gonepteryx rhamni	2	679	5
Lasiommata megera	1	1930	159
Leptidea sinapis	4	108	6
Limenitis camilla	2	4248	56
Lycaena phlaeas	3	163	4
Maniola jurtina	1	11463	6
Melanargia galathea	1	1041	5
Melitaea cinxia	1	25	8
Neozephyrus quercus	2	4578	2
Ochlodes sylvanus	1	1257	6
Pararge aegeria	1	2738	6
Pieris brassicae	4	1232	5
Pieris napi	4	4843	6
Pieris rapae	4	1510	6
Polygonia c-album	4	518	5
Polyommatus icarus	3	2634	6
Pyrgus malvae	1	3207	69
Pyronia tithonus	1	3966	5
Satyrium w-album	2	2187	147
Thymelicus sylvestris/ lineola	1	2269	6
Vanessa atalanta	0	378	4
Vanessa cardui	0	591	4

Table 4.1. The abundance and distribution of the 36 butterfly species observed across the 515 1 km grid squares. The corresponding Ecological Attribute Group (EAG) for each species is indicated.

Group	Habitat Association	Species Richness	Total Abundance	Distribution
EAG0	Migrant	3	998	4
EAG1	Open Grassland	15	46473	412
EAG2	Woodland	7	12049	190
EAG3	Species rich short turf	4	4429	13
EAG4	Ruderal	8	13739	11

Table 4.2: Variation in Warwickshire's butterfly species richness, total abundance and spatial distribution between species ecological attribute groups (EAGs) from1990-1999. Species are grouped according to the four ecological attribute groups identified by Shreeve *et al.*, (2001) with an additional group for migratory species, which are not resident to the British Isles (EAG0). Main habitat associations for species comprising each group are provided.

4.2.2 Landscape compositional models: LCM 2000

According to the classification of land cover within Warwickshire by the Land Cover Map (LCM) 2000, Warwickshire (n = 2467) is dominated by arable horticulture (LCM-42) (26.37 ha per 1 km grid square), improved grassland (LCM-51) (19.31 ha/1 km square) and arable cereals (LCM-41) (17.21 ha/1 km square) (Table 2.4). These habitats, also, are widely distributed across Warwickshire, occurring in 80 % of the grid squares, in addition to broad-leaved/ mixed woodland (LCM-11) which occurs within 95 % of grid squares but has a smaller average coverage per grid square (9.81 ha).

Focusing on the 515 grid squares with butterfly records, the proportion and spatial distribution of land covers within those squares is somewhat different to that across all 2467 squares. This sub-section of the landscape is still dominated per grid square by arable horticulture (22.66 ha) and improved grassland (18.89 ha), but the average area per grid square of arable cereals, broad-leaved woodland and suburban/ rural developed land is now very similar (12 – 13 ha) (Table 2.4). These habitats with area per square greater than 10 ha dominate the landscape with greatest spatial distribution and average area. In addition set-aside grassland (LCM-52) and calcareous grassland (LCM-71) occur in > 80 % of grid squares, however the average area per square for these habitats is small in comparison, particularly for set aside grassland (4.66 ha).

Model development - all butterfly species presence-absence (LCM-ALL)

The forward selection and backward elimination procedures provided the same results as the selection/elimination threshold was increased from 1 to 3.86 (based on X^2 distribution for 1 degree of freedom at the 5 % significance level) with the initial selection of nine variables with a threshold of 1 and the dropping of three variables, arable non-rotational (LCM-43), set-aside grassland (LCM-52), and suburban/ rural developed (LCM-171), as the threshold increased to 3.86 (Table 4.3). Removal of these three variables in combination significantly increased the residual deviance ($\Delta D_3 = 10$, $p = 0.016$). Individually the variables LCM-43, LCM-52 and LCM-171 can be dropped from the model without significant effect on the residual deviance ($\Delta D_1 = 3$, $p = 0.064$; $\Delta D_1 = 4$, $= 0.059$; $\Delta D_1 = 3$, $p = 0.069$ respectively); however removal of any pairwise combination significantly increases the residual deviance of the model, as such all three variables were retained.

Explanatory variable	**Deviance (p)**	**In/ out ratio**	
		1-3	**3.86**
NLAND	52.884	*	*
LCM-42	37.329	*	*
LCM-11	23.204	*	*
LCM-131	20.826	*	*
LCM-41	11.390	*	*
LSIDI	5.736	*	*
LCM-43	3.780	*	
LCM-52	3.210	*	
LCM-171	3.312	*	
Model deviance (p)	-	12 (<0.001)	151 (<0.001)

Table 4.3: Landscape compositional variables identified from a forwards selection logistic regression model, using the in ratios/ out ratios of 1, 2, 3 and 3.86. See Table 2.4 for landscape compositional codes and descriptions.

Model parameterisation

Following the forwards selection and backwards elimination procedure four compositional models (referred to by the subscript 'comp' hereafter) were developed which predicted the probability of butterfly presence as a function of landscape compositional variables for all butterfly species (n = 2467) and for species comprising each ecological attribute group (n = 515) (Table 4.4). The compositional models for all butterfly species (LCM-ALL_{comp}) and for species comprising EAG2 (LCM-$EAG2_{comp}$), comprised a high number of variables, with nine variables in LCM-ALL_{comp} and 11 variables in LCM-$EAG2_{comp}$; seven are common across the two models. An increasing area of broad-leaved/ mixed semi-natural woodland (LCM-11) affected the probability of occurrence for butterfly species in all four models, with a highly significant relationship with butterfly occurrence. For the models LCM-ALL_{comp}, LCM-$EAG1_{comp}$ and LCM-$EAG3_{comp}$ this relationship is positive, however LCM-11 exhibits a negative relationship with the probability of occurrence of species comprising EAG2 (LCM-$EAG2_{comp}$). All 11 variables within this latter model negatively influence the probability of occurrence of EAG2 species. The probabilities of occurrence of all species and EAG2 species were significantly affected by increasing area of water (inland) (LCM-131), with a positive relationship for all butterfly species and a negative relationship for EAG2 species. Arable land covers (arable horticulture LCM–42; arable cereals LCM-41; and arable non-rotational LCM-43) significantly negatively affected the probability of occurrence of all butterfly species (LCM-ALL_{comp}), and EAG2 species (LCM-42 and LCM-41 only).

Of the six variables in the EAG1 model, four of these significantly affected the probability of occurrence of EAG2 species as well as EAG1 species. The land covers coniferous woodland (LCM-21) and inland bare ground (LCM-161) are unique to the EAG1 model, with a negative relationship observed between EAG1 species occurrence and coniferous woodland and a positive relationship with inland bare ground. All three variables within the LCM-$EAG3_{comp}$ model are also associated with the occurrence of EAG2 species, with negative relationships observed for the variables improved grassland (LCM-51) and neutral grassland (LCM-61) in both cases.

All four compositional models significantly predicted the presence-absence of the corresponding butterfly species groups (Table 4.5a). Parameters across all four models have relatively small standard errors in proportion to the estimates and acceptable Variance Inflation Factors (VIF) values close to 1, indicating lack of co-linearity between the landscape compositional variables (Table 4.4). Goodness of fit as determined by the H-L statistic was acceptable for all four models, with non-significance indicating that there is no evidence for difference between the model predictions and the observed values (Table 4.5a). Model specificity was greatest for the LCM-EAG1$_{comp}$ and LCM-EAG3$_{comp}$ models (71.8 % and 74.5 % respectively); however sensitivity of these models was low in comparison (51.9 % and 53.8 % respectively) (Table 4.5b). Sensitivity was greatest for LCM-EAG2$_{comp}$ (61.6 %) and LCM-ALL$_{comp}$ (61.2 %) (Table 4.5b). Discrimination of the models as determined by the AUC$_{comp}$ was significant for all four models and ranged from 0.667 – 0.719. 'Fair' discrimination as defined by Araujo *et al.*, (2005) was obtained for the model LCM-EAG3$_{comp}$ (AUC$_{comp}$ = 0.719, $p<0.001$) (Table 4.5b).

Model/ Variable	Estimate	*P*	SE	VIF	Model/ Variable	Estimate	*P*	SE	VIF
LCM-ALL$_{comp}$					**LCM-EAG2$_{comp}$**				
Intercept	-3.318	<0.001	0.537	-	Intercept	3.347	<0.001	0.668	-
LCM-11	0.024	<0.001	0.006	1.285	LCM-172	-0.088	<0.001	0.022	1.547
LCM-131	0.092	<0.001	0.025	1.074	LCM-11	-0.038	<0.001	0.011	1.020
LCM-42	-0.011	0.012	0.004	1.470	LCM-52	-0.102	<0.001	0.025	1.244
LSIDI	1.712	0.015	0.704	1.804	LCM-51	-0.046	<0.001	0.011	2.090
LCM-41	-0.009	0.038	0.004	1.428	LCM-171	-0.041	<0.001	0.010	2.276
NLAND	0.090	0.054	0.047	1.551	LCM-42	-0.037	<0.001	0.010	1.961
LCM-52	0.021	0.056	0.011	1.161	LCM-131	-0.089	0.003	0.030	1.220
LCM-171	0.080	0.070	0.004	1.966	LCM-41	-0.029	0.003	0.010	1.776
LCM-43	-0.457	0.316	0.456	1.005					
LCM-EAG1$_{comp}$					**LCM-EAG3$_{comp}$**				
Intercept	0.373	0.094	0.222	-	Intercept	-3.646	<0.001	0.577	-
LCM-11	0.046	0.002	0.015	1.014	LCM-11	0.048	0.009	0.018	1.002
LCM-172	0.099	0.010	0.038	1.357	LCM-51	-0.034	0.149	0.024	1.012
LCM-52	0.059	0.042	0.029	1.011	LCM-61	-0.370	0.257	0.326	1.011
LCM-131	0.047	0.202	0.037	1.003					
LCM-21	-0.039	0.127	0.026	1.010					
LCM-161	0.154	0.235	0.130	1.351					

Table 4.4: The contribution of landscape compositional parameters for determining the presence-absence of all butterfly species (LCM-ALL) and butterfly species within each ecological attribute group (LCM-EAG1; LCM-EAG2 and LCM-EAG3). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

(a)

$Model_{comp}$	Model deviance			Goodness of fit (H-L)		
	X^2	*df*	*p*	X^2	*df*	*p*
LCM-ALL	162	9	<0.001	0.853	8	0.999
LCM-EAG1	48.6	6	<0.001	2.389	8	0.967
LCM-EAG2	68.1	8	<0.001	11.670	8	0.167
LCM-EAG3	11.9	3	0.008	7.949	8	0.438

(b)

$Model_{comp}$	Discrimination		Confusion matrix		
	AUC	*p*	*T*	SPC (%)	TPR (%)
LCM-ALL	0.667	0.000	0.2	62.6	61.2
LCM-EAG1	0.692	0.000	0.8	71.8	51.9
LCM-EAG2	0.698	0.000	0.4	67.4	61.6
LCM-EAG3	0.719	0.007	0.03	74.5	53.8

Table 4.5: Performance and accuracy of the compositional models for predicting presence-absence of all butterfly species ($LCM\text{-}ALL_{comp}$), and species comprising the three ecological attribute groups ($LCM\text{-}EAG1_{comp}$; $LCM\text{-}EAG2_{comp}$ and $LCM\text{-}EAG3_{comp}$). (a) Model fit determined by the model deviance (X^2) and the goodness of fit assessed by the H-L test statistic (X^2). Significance level (*p*) and degrees of freedom (*df*) are provided. (b) The discrimination of the models is provided in terms of area under the ROC curve (AUC) and the specificity (SPC) and sensitivity (True Positive Rate; TPR) as determined by threshold (*T*) equal to the prevalence rate for each model.

4.2.3 Landscape connectivity models: LCM 2000

The connectivity metrics included during the modelling process measured the connectivity of the 12 key land cover variables which were identified from the four compositional models in section 4.2.2 (see also section 2.2.4 for further methodological details). The connectivity metrics were considered in addition to the compositional variables during the connectivity modelling procedure, providing complementary assessments of composition. Only connectivity metrics which did not exhibit a strong correlation with the area of corresponding land cover class were considered during the modelling procedure (Table 4.6; see section 2.2.4). The forwards selection/ backwards elimination modelling procedure was therefore conducted on a total of 12 IIC metrics, which measured connectivity within each grid square, and a total of six varIIC metrics, which measured the connectivity across the whole of Warwickshire (Table 4.6). Of these connectivity metrics, seven metrics were not selected across the four models during the variable selection procedure with in ratios/ out ratios ranging from 1.00-3.86 (Table 4.6). The final four connectivity models comprised a total of five IIC metrics and one varIIC metric (Table 4.6).

Considering the connectivity metrics included within the final four connectivity models, the average connectivity of woodland habitat (L11_IIC = 0.430) is greater than that of the other three key habitats for the occupied 1 km grid squares, closely followed by the connectivity of set aside grassland (L52_IIC = 0.420) (Table 4.7). The connectivity of all four key land cover classes as measured by the IIC index varies greatly across the 1 km grid squares, with no connectivity occurring within grid squares (IIC = 0.0) ranging to maximum connectivity between patches with IIC values of >0.9 obtained (Table 4.7). The average importance of patches for maintaining connectivity of set–aside grassland (L52_varIICconn) across the whole of Warwickshire is low, and the variation between grid squares in this value is equally low (Table 4.7).

Metric	Average	SEM	Min	Max
All butterflies (*n* = 2427)				
L11_IIC	0.435	0.003	0.000	0.991
L52_varIICconn	0.031	0.004	0.000	3.990
EAGs (*n* = 515)				
L11_IIC	0.430	0.007	0.000	0.988
L21_IIC	0.224	0.013	0.000	0.957
L52_IIC	0.420	0.009	0.000	0.993
L161_IIC	0.130	0.011	0.000	0.991

Table 4.7: The mean and range in connectivity metrics included within the landscape compositional connectivity models for all butterfly species and species comprising each EAG.

Metric	LCM-ALL	LCM-EAG1	LCM-EAG2	LCM-EAG3
L11_IIC	○○○	○	○○○	○○○
L11_varIIC	×	○	○○	×
L11_varIICconn	○○	×	×	×
L131_IIC	○	○	○	
L131_varIICflux	○	×	×	
L161_IIC		○○○		
L171_IIC	○		○	
L172_IIC		○	○	
L21_IIC		○○○		
L41_IIC	○○○		○○	
L41_varIICconn	○○		○	
L42_IIC	○		○	
L42_varIICconn	○○		○	
L43_IIC	○○			
L51_IIC			○	○
L52_IIC	○○	○○○	○○	
L52_varIICconn	○○○	○	○○	
L61_IIC			○	○

Table 4.6: Connectivity metrics for the 12 key LCM 2000 land covers included within the connectivity modelling. × refers to metrics which strongly correlated with area of corresponding land cover classes and were not included during model development; ○ refers to metrics selected from the correlation analysis but not selected during the modelling procedure; ○○ refers to metrics selected during the modelling procedure but dropped from the final model; ○○○ refers to metrics retained within the final connectivity model; and no symbol indicates that the key habitat class for the metric did not occur in the corresponding compositional model.

Model parameterisation

The final four connectivity models (referred to a conn hereafter) comprised a combination of compositional and connectivity metrics, however, only a small number of connectivity metrics were included in comparison (Table 4.8). The LCM-ALL_{conn} model comprised metrics which measured the connectivity of the habitats arable cereals (L41_IIC) and broad-leaved/ mixed woodland (L11_IIC) in addition to the area of these habitats (LCM-41 and LCM-11). The connectivity of broad-leaved woodland (L11_IIC) was also found to be important in addition to the area of this habitat in the LCM-$EAG2_{conn}$ and LCM-$EAG3_{conn}$ models (Table 4.8). The connectivity of broad-leaved woodland positively influenced butterfly presence for all species (LCM-ALL_{conn}) and EAG3 species (LCM-$EAG3_{conn}$), however this relationship was negative for EAG2 species (LCM-$EAG2_{conn}$). For the LCM-EAG2 model, all eight compositional metrics were retained in addition to the woodland connectivity metric (L11_IIC) (Table 4.4 and Table 4.7).

A number of connectivity metrics were found to be more important than the area of corresponding habitat in predicting butterfly presence-absence (Table 4.4 and Table 4.7). In particular, the connectivity of set aside grassland across Warwickshire (L52_varIICconn) was found to positively influence presence of all butterfly species, but the area of this habitat was not selected during the modelling procedure. Furthermore, the LCM-$EAG1_{conn}$ model included metrics which measured the connectivity of the habitats set aside grassland (L51_IIC), coniferous woodland (L21_IIC) and inland bare ground (L161_IIC). The area of coniferous woodland (LCM-21) and inland bare ground (LCM-161) were not selected during the modelling procedure with a selection threshold of 1, and the area of set aside grassland (LCM-52) was dropped from the model at a threshold of 2 without significant increase in residual deviance ($\Delta D_1 = 1.8$, $p = 0.175$).

All four connectivity models significantly predicted the presence-absence of butterfly species of the corresponding species groups as a function of landscape connectivity and compositional variables (Table 4.9a). The parameters of the four connectivity models fit the data at an acceptable level with non-significant H-L test statistics obtained, indicating there is no evidence for differences between the observed and model-predicated values (Table 4.9a).

When considering the performance of the connectivity models there is little change in the discriminatory power of the LCM-ALL_{conn} model compared to the corresponding LCM-ALL_{comp} model (AUC_{conn} = 0.670, p<0.001; AUC_{comp} = 0.667, p<0.001) (Table 4.9b; Table 4.5b). Additionally, there is no change in the discriminatory power of the LCM-$EAG2_{conn}$ model in both cases ($AUC_{conn/\ comp}$ = 0.698, p<0.001). For the LCM-$EAG1_{conn}$ and LCM-$EAG3_{conn}$ models, discriminatory power has improved in comparison to the compositional models. For the LCM-$EAG1_{conn}$ model discriminatory power improved from 'poor' discrimination associated with the compositional model (AUC_{comp} = 0.692, p<0.001), as defined by Araujo *et al.*, (2005) to 'fair' discrimination by the connectivity model (AUC_{conn} = 0.728, p<0.001). For the LCM-$EAG3_{conn}$ model discriminatory power improved from 'fair' discrimination (AUC_{comp} = 0.719, p <0.001) to 'good' discriminatory power (AUC_{conn} = 0.816, p<0.001) (Table 4.9b; Table 4.5b). The specificity and sensitivity associated with LCM-$EAG1_{conn}$ and LCM-$EAG3_{conn}$ has also increased in comparison to the compositional models (Table 4.9b and Table 4.5b).

Model/ Variable	Estimate	*P*	SE	VIF	Model/ Variable	Estimate	*P*	SE	VIF
LCM-ALL$_{conn}$					**LCM-EAG2$_{conn}$**				
Intercept	-3.792	<.001	0.563	-	Intercept	3.189	<.001	0.944	-
LCM-131	0.092	<.001	0.025	1.073	LCM-172	-0.086	<.001	0.023	1.342
LCM-41	-0.001	0.003	0.005	1.801	LCM-11	-0.038	<.001	0.011	1.021
LCM-171	0.008	0.080	0.004	1.997	LCM-52	-0.100	<.001	0.026	1.091
NLAND	0.090	0.053	0.047	1.540	LCM-51	-0.049	<.001	0.012	1.739
LCM-42	-0.011	0.010	0.004	1.462	LCM-171	-0.040	<.001	0.010	1.827
LCM-11	0.014	0.075	0.008	1.702	LCM-42	-0.046	<.001	0.012	2.082
L41_IIC	0.433	0.089	0.254	1.527	LCM-131	-0.088	0.004	0.030	1.088
L11_IIC	0.844	0.041	0.413	1.373	LCM-41	-0.028	0.008	0.010	1.173
LSIDI	1.899	0.007	0.698	1.769	L11_IIC	-0.168	0.813	0.710	1.342
LCM-43	-0.492	0.293	0.468	1.006					
L52varIICconn	0.093	0.672	0.219	1.020					
LCM-EAG1$_{conn}$					**LCM-EAG3$_{conn}$**				
Intercept	0.173	0.530	0.326	-	Intercept	-6.090	<.001	1.150	-
LCM-172	0.092	0.018	0.039	1.201	L11_IIC	4.490	0.005	1.590	1.022
LCM-11	0.046	0.002	0.014	1.017	LCM-11	0.051	0.007	0.019	1.003
L21_IIC	-1.033	0.006	0.374	1.026	LCM-51	-0.026	0.280	0.024	1.033
L52_IIC	1.528	0.009	0.587	1.018	LCM-61	-0.325	0.286	0.305	1.012
L161_IIC	1.878	0.015	0.775	1.195					

Table 4.8: The contribution of landscape compositional and connectivity parameters for determining the presence-absence of all butterfly species (LCM-ALL$_{conn}$) and butterfly species within each ecological attribute group (LCM-EAG1$_{conn}$; LCM-EAG2$_{conn}$ and LCM-EAG3$_{conn}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

(a)

$Model_{conn}$	Model deviance			Goodness of fit (H-L)		
	X^2	*df*	*p*	X^2	*df*	*p*
LCM-ALL	172.0	11	<0.001	6.964	8	0.541
LCM-EAG1	59.1	5	<0.001	6.398	8	0.603
LCM-EAG2	68.2	9	<0.001	5.984	8	0.649
LCM-EAG3	19.6	4	<0.001	5.817	8	0.668

(b)

$Model_{conn}$	Discrimination		Confusion matrix		
	AUC	*p*	*T*	SPC (%)	TPR (%)
LCM-ALL	0.670	0.000	0.2	63.2	61.9
LCM-EAG1	0.728	0.000	0.8	71.8	62.6
LCM-EAG2	0.698	0.000	0.4	67.1	62.1
LCM-EAG3	0.816	0.000	0.03	78.7	61.5

Table 4.9: Performance and accuracy of the connectivity models for predicting presence-absence of all butterfly species ($LCM\text{-}ALL_{conn}$), and species comprising the three ecological attribute groups ($LCM\text{-}EAG1_{conn}$; $LCM\text{-}EAG2_{conn}$ and $LCM\text{-}EAG3_{conn}$). (a) Model fit determined by the model deviance (X^2) and the goodness of fit assessed by the H-L test statistic (X^2). Significance level (*p*) and degrees of freedom (*df*) are provided. (b) The discrimination of the models is provided in terms of area under the ROC curve (AUC) and the specificity (SPC) and sensitivity (True Positive Rate; TPR) as determined by threshold (*T*) equal to the prevalence rate for each model.

4.2.4 Landscape structural models: LCM 2000

A selection of 41 landscape structure metrics were considered during the development of the four landscape structural models (referred to by the notation 'struc' hereafter). Following the same procedure of forwards selection and backwards elimination outlined in section 2.2.5, a total of 17 landscape structure metrics were included within the final landscape structure models (Table 4.10). Of these 17 metrics, eight were associated with the landscape aspect of aggregation, three with patch shape, three with contrast and two with patch area (Table 4.10; Appendix A4 and A8). For the metrics PROX_MN, SIMI_MN, and ECON_AM, a number of summary statistics were selected which measured the variability and range associated with these metrics (Table 4.10, Table 4.11). For several of the metrics, there is large variability in metric value across the grid squares (n = 2427), particularly for the range and median Similarity Index (SIMI_RA; SIMI_MD), indicating that the grid square landscapes vary considerably in the degree of aggregation of similar patch types (Table 4.11). Within the occupied squares (n = 515) the range in average and median Proximity Index between patches of the same class (PROX_MN; PROX_MD) is small, with a minimum average proximity index of 0.22 and a maximum average proximity index of 40.39 (Table 4.11).

Metric	Aspect	LCM-ALL	LCM-EAG1	LCM-EAG2	LCM-EAG3
AREA_AM	Area	○	○○○	○	○
AREA_MD	Area	○	○	○	○
AREA_MN	Area	○○	○	○	○
CIRCLE_AM	Shape	○	○○○	○	○
CIRCLE_MN	Shape	○○	○	○	○
CIRCLE_RA	Shape	○	○	○	○
CIRCLE_SD	Shape	○	○	○	○
CONNECT	Aggregation	○	○	○	○
CONTIG_MD	Shape	○	×	×	×
CONTIG_MN	Shape	○	○	○	○
CONTIG_RA	Shape	○	○	○	○
CONTIG_SD	Shape	○	○○○	○	○
ECON_AM	Contrast	○○○	○○○	○○	○
ECON_MN	Contrast	○	○	○	○
ECON_RA	Contrast	○○	○○○	○	○
ECON_SD	Contrast	○	○	○○○	○
ENN_AM	Aggregation	○○	○	○	○
ENN_CV	Aggregation	×	○○	○	○
ENN_MD	Aggregation	○	○	○	○
ENN_MN	Aggregation	○	○○	○	○
ENN_SD	Aggregation	○	○○○	○○○	○
FRAC_RA	Shape	×	○○	○	○
GYRATE_CV	Area	×	○○○	○○	○
GYRATE_MD	Area	○	×	×	×
GYRATE_MN	Area	○	○	○	○
IJI	Aggregation	○○	○○	○○	○
PAFRAC	Shape	○	×	×	×
PRD	Diversity	○○○	○	○	○○
PROX_AM	Aggregation	○	○	○○	○
PROX_CV	Aggregation	○	○○○	○	○○○
PROX_MD	Aggregation	○○○	○	○○	○○○
PROX_MN	Aggregation	○	○○○	○○	○
SHAPE_AM	Shape	○	○	○	○○○
SHAPE_MD	Shape	○	○	○	○
SHAPE_MN	Shape	○○	○○	○	○
SHAPE_SD	Shape	○	○○	○	○
SIMI_AM	Aggregation	○○○	×	×	×
SIMI_CV	Aggregation	○	○	○	○
SIMI_MD	Aggregation	○○○	○	○	○○
SIMI_MN	Aggregation	○○○	○○	○○	○
SIMI_RA	Aggregation	○○○	×	×	×

Table 4.10: The 41 landscape structure metrics included within the forwards selection/ backwards elimination modelling procedure. × refers to metrics which strongly correlated with other metrics and were not included during model development; ○ refers to metrics selected from the correlation analysis but not selected during the modelling procedure; ○○ refers to metrics selected during the modelling procedure but dropped from the final model; ○○○ refers to metrics retained within the final connectivity model.

Metric	Average	SEM	Min	Max
All butterflies (*n* = 2427)				
ECON_AM (%)	60.43	0.22	15.04	86.95
PRD (n/ha)	7.68	0.03	3.00	12.00
PROX_MD	1.64	0.03	0.00	29.18
SIMI_AM	172.57	2.96	1.34	740.01
SIMI_MD	39.87	1.41	1.01	1330.17
SIMI_MN	124.20	2.10	12.31	989.49
SIMI_RA	528.91	5.62	29.51	1550.50
EAGs (*n* = 515)				
AREA_AM (ha)	16.84	0.47	5.09	85.46
CIRCLE_AM	0.61	0.00	0.39	0.72
CONTIG_SD	0.21	0.00	0.09	0.33
ECON_RA (%)	78.71	0.49	41.88	100.00
ECON_SD (%)	21.89	0.15	11.75	38.63
ENN_SD	164.54	2.46	25.97	380.22
GYRATE_CV (%)	79.56	0.63	46.17	137.27
PROX_CV (%)	186.24	2.49	81.99	448.04
PROX_MD	1.56	0.05	0.05	7.24
PROX_MN	7.02	0.20	0.22	40.39
SHAPE_AM	1.80	0.01	1.19	2.85

Table 4.11: Average and range in the 18 structural metrics selected during the landscape structure modelling procedure. Note: one metric (PROX_MD) is common to the all butterfly and EAG models. Metrics are unit less unless otherwise stated.

Model parameterisation

The final four landscape structure models differed in terms of the model parameters, performance, and accuracy (Table 4.12). The models for EAG2 and EAG3 species comprised very few variables in comparison to the models for all species and EAG1 species. The standard deviation of the Euclidean distance between patches (ENN_SD) positively influenced butterfly occurrence for EAG1 species but negatively influenced EAG2 species (Table 4.12). The edge contrast between neighbouring patches weighted by area (ECON_AM) positively influenced the occurrence of EAG1 species and significantly positively influenced ‘all butterfly’ species. For both ‘all butterflies’ and EAG3 species the median size and proximity between neighbouring patches (PROX_MD) influenced butterfly occurrence; this relationship was negative when considering ‘all butterfly’ species and positive when considering EAG3 species (Table 4.12).

The probability of butterfly presence for all species (LCM-ALL_{struc}) and for species associated with the three ecological attribute groups was significantly modelled as function of the structure of the grid square landscapes (Table 4.13a). The four models could be considered to fit the data well with non-significant H-L test statistics obtained (Table 4.13a). Model fit for LCM-$EAG1_{struc}$, however, was particularly poor (X^2_8 = 14.881, p = 0.061), with the significance value of this test statistic close to p = 0.05. The discriminatory power of the LCM-$EAG1_{struc}$ model was also considered to be poor (AUC_{struc} = 0.656, p = 0.789), however, the AUC_{struc} for this model was high in comparison to those for the LCM-ALL_{struc} and LCM-$EAG2_{struc}$ models (Table 4.13b). The AUC_{struc} values obtained for the models ranged from 0.599 for the LCM-$EAG2_{struc}$ model to 0.789 for the LCM-$EAG3_{struc}$ model. The highest AUC_{struc} value was obtained for the LCM-$EAG3_{struc}$ model (AUC_{struc} = 0.789, p = 0.000), which can be considered to provide 'good' discrimination between presence-absence grid squares. A higher proportion of grid squares were accurately classified as absent for the LCM-$EAG3_{struc}$ model (78.1 %) and LCM-$EAG1_{struc}$ model (64.1 %) in comparison to the other models. Model sensitivity was highest for the LCM-ALL_{struc} model (63.3 %) closely followed by the LCM-$EAG1_{struc}$ (61.9 %) and LCM-$EAG3_{struc}$ (61.5 %) models (Table 4.13b).

Model/ Variable	Estimate	*P*	SE	VIF
LCM-ALL$_{struc}$				
Intercept	-2.796	<.001	0.516	-
ECON_AM	0.016	0.017	0.007	1.929
SIMI_RA	-0.001	0.022	0.000	3.272
SIMI_AM	-0.001	0.058	0.001	2.384
SIMI_MD	-0.004	0.089	0.002	2.026
PROX_MD	-0.063	0.189	0.048	1.208
SIMI_MN	0.002	0.198	0.001	6.247
PRD	0.143	0.002	0.045	1.596
LCM-EAG1$_{struc}$				
Intercept	-5.920	0.024	2.630	-
ECON_RA	0.034	0.002	0.011	1.065
GYRATE_CV	-0.053	0.002	0.017	4.213
AREA_AM	0.068	0.005	0.024	4.796
PROX_MN	-0.068	0.008	0.026	1.232
PROX_CV	0.006	0.026	0.003	1.378
CIRCLE_AM	7.170	0.045	3.570	1.760
ENN_SD	0.003	0.118	0.002	1.109
CONTIG_SD	5.120	0.120	3.290	1.513
ECON_AM	0.019	0.124	0.012	1.247
LCM-EAG2$_{struc}$				
Intercept	2.070	0.004	0.718	-
ECON_SD	-0.091	0.002	0.029	1.005
ENN_SD	-0.004	0.025	0.002	1.005
LCM-EAG3$_{struc}$				
Intercept	-2.270	0.424	2.830	-
PROX_CV	0.018	<.001	0.005	1.309
PROX_MD	0.593	0.025	0.264	1.205
SHAPE_AM	-3.440	0.037	1.650	1.104

Table 4.12: The contribution of landscape structural parameters for determining the presence-absence of all butterfly species (LCM-ALL$_{struc}$) and butterfly species within each ecological attribute group (LCM-EAG1$_{struc}$; LCM-EAG2$_{struc}$ and LCM-EAG3$_{struc}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

(a)

$Model_{struc}$	Model deviance			Goodness of fit (H-L)		
	X^2	*df*	*p*	X^2	*df*	*p*
LCM-ALL	100.0	7	<0.001	6.560	8	0.585
LCM-EAG1	29.0	9	<0.001	14.881	8	0.061
LCM-EAG2	14.4	2	<0.001	7.231	8	0.512
LCM-EAG3	16.2	3	0.001	7.819	8	0.451

(b)

$Model_{struc}$	Discrimination		Confusion matrix		
	AUC	*p*	*T*	SPC (%)	TPR (%)
LCM-ALL	0.632	0.000	0.2	54.2	63.3
LCM-EAG1	0.656	0.000	0.8	64.1	61.9
LCM-EAG2	0.599	0.000	0.4	54.8	56.3
LCM-EAG3	0.789	0.000	0.03	78.1	61.5

Table 4.13: Performance and accuracy of the structural models for predicting presence-absence of all butterfly species ($LCM\text{-}ALL_{struc}$), and species comprising the three ecological attribute groups ($LCM\text{-}EAG1_{struc}$; $LCM\text{-}EAG2_{struc}$ and $LCM\text{-}EAG3_{struc}$). (a) Model fit determined by the model deviance (X^2) and the goodness of fit assessed by the H-L test statistic (X^2). Significance level (*p*) and degrees of freedom (*df*) are provided. (b) The discrimination of the models is provided in terms of area under the ROC curve (AUC) and the specificity (SPC) and sensitivity (True Positive Rate; TPR) as determined by threshold (*T*) equal to the prevalence rate for each model.

4.2.5 Combined models: LCM 2000

The landscape connectivity models incorporated both compositional and connectivity variables, and these were considered together with the key landscape structure variables from the four structural models developed in section 4.2.4 for the development of combined models of butterfly presence-absence. A total of 35 landscape variables were considered during the forwards selection/ backwards elimination procedure for the development of the combined models (referred to by the notation 'comb' hereafter).

Model parameterisation

The four combined models differed in terms of the model parameters, performance, and accuracy (Table 4.14a,b). Measures of landscape composition and connectivity occurred within most models. In particular, the area of broad-leaved / mixed woodland (LCM-11) influenced the occurrence of all four species groups (all butterflies and EAG1-3), and this relationship was significant for the EAG models (Figure 4.1a-d). Increases in the area of broad-leaved/ mixed woodland (LCM-11) positively influenced the occurrence of 'all butterfly species', EAG1 and EAG3 species (Figure 4.1a,b,d) and negatively influenced the occurrence of EAG2 species (Figure 4.1c).

The LCM-ALL_{comb} model comprised 12 metrics, with all parameters except the connectivity of arable cereals (L41_IIC) included from the LCM-ALL_{conn} model (Table 4.14). The occurrence of 'all butterfly species' was significantly positively influenced by increases in the area of standing water (LCM-131) (Figure 4.2a), suburban-rural developed land (LCM-171) and woodland connectivity (L11_IIC) (Table 4.14; Table 4.8). The arable land covers arable cereals (LCM-41) and arable horticulture (LCM-42) significantly negatively influenced butterfly occurrence, as did arable non-rotational (LCM-43) but this relationship was not significant. Only two structural metrics, the median similarity index (SIMI_MD) and average area-weighted edge contrast (ECON_AM) were retained from the six parameters in the structural model, and these two metrics significantly negatively influenced butterfly occurrence (Table 4.12; Table 4.14).

The LCM-EAG1$_{comb}$ model comprised all the parameters from the LCM-EAG1$_{conn}$ model, which included two compositional variables and three connectivity metrics (Table 4.14). The LCM-EAG1$_{comb}$ model comprised several structural metrics with six of the nine variables from the LCM-EAG1$_{struc}$ model retained (Table 4.14 and Table 4.12). Most notably, the positive relationship with patch elongation (CIRCLE_AM) was strongly significant (Figure 4.2b), with increases in patch elongation positively influencing the probability of occurrence of EAG1 species, whilst the variability in patch extent (GYRATE_RA) negatively influenced EAG1 species occurrence (Table 4.14). Increases in the area-weighted patch size (AREA_AM), range in edge contrast (ECON_RA), and variability in Euclidean nearest neighbour (ENN_SD) positively influenced EAG1 species occurrence. The mean size and proximity of patches (PROX_MN) of the same type was negatively associated with butterfly occurrence.

The LCM-EAG2$_{comb}$ model comprised compositional variables and connectivity metrics only (Table 4.14b). All variables negatively influenced EAG2 species occurrence expect for the connectivity of broad-leaved woodland, but this relationship was non-significant (Table 4.14b). Significant negative associations occurred with the area of continuous urban (LCM-172), standing water (LCM-131), suburban/ rural developed land (LCM-171), improved grassland (LCM-51), arable cereals (LCM-41) and arable horticulture (LCM-42) (Table 4.14b). In particular, the probability of occurrence of EAG2 species was strongly negatively influenced by increases in the area of arable horticulture (LCM-42) (Figure 4.2c).

The LCM-EAG3$_{comb}$ model comprised four variables; one compositional variable, one connectivity metric and two structural metrics (Table 4.14b). The occurrence of EAG3 species is positively influenced by increases in the area of broadleaved woodland (LCM-11) and woodland connectivity (L11_IIC). Several variables from the LCM-EAG3$_{conn}$ and LCM-EAG3$_{struc}$ models were not selected in the final combined model (Table 4.8; Table 4.12). These variables were the area of improved grassland (LCM-51), neutral grassland (LCM-61) and the median proximity index (PROX_MD). Structural metrics which were retained in the final combined model included the variation in average patch aggregation (PROX_CV), which significantly positively influenced EAG3 species occurrence and the area-weighted shape index

(SHAPE_AM). Increases in patch shape complexity, in particular for larger patches, significantly negatively decreased the probability of occurrence of EAG3 species (Figure 4.2d).

All four combined models were considered to fit the data well with non-significant H-L test statistics obtained, indicating there is no evidence for differences between the model predicted and observed values (Table 4.15a). The predictions from the LCM combined models can be seen to match the observed data at an acceptable level when comparing the probabilities of occurrence derived from the combined models and the observations of presence-absence within the original data (Figure 4.3a-h). In particular, for the LCM-EAG1$_{comb}$ and LCM-EAG3$_{comb}$ models the low probabilities of butterfly presence within grid squares can be seen to match the observed absence squares (70.9 % and 78.7 % respectively) and the high probabilities of butterfly presence match the presence squares (64.8 % and 76.9 % respectively) (Figure 4.3c,d; Table 4.15b). Model specificity and sensitivity was lowest for the LCM-ALL$_{comb}$ model (61.4 % and 61.2 % respectively), however the range in model predictions, can be seen to match the observed data, with presence squares corresponding to those squares with a higher occurrence probability and absence squares corresponding to those with lower probability of occurrence (Figure 4.3a,b).

The discriminatory power of the models, as measured by the AUC$_{comb}$ ranged from 0.671 to 0.823, with highest discrimination obtained for the LCM-EAG3$_{comb}$ model and lowest discrimination obtained for the LCM-ALL$_{comb}$ model. When comparing the discriminatory power of the four models to that obtained from the connectivity models slight improvement is obtained (Table 4.15b; 4.9b). For example, AUC$_{conn}$ obtained for the LCM-EAG1$_{conn}$ model improved from 0.728 to 0.760 for the LCM-EAG1$_{comb}$ model. The AUC$_{comb}$ values are much improved in comparison to the AUC$_{struc}$ (Table 4.15b; Table 4.13b). In particular, the LCM-EAG1$_{comb}$, LCM-EAG2$_{comb}$ and LCM-EAG3$_{comb}$ models improved a category according to the classification of AUC values by Araujo *et al.*, (2005). For example, model discrimination of the LCM-EAG2$_{comb}$ model was considered to be 'poor' (AUC$_{comb}$ = 0.698, P<0.001), in comparison to the model discrimination obtained for the LCM-EAG2$_{struc}$ model which was considered to 'fail' in its discriminatory power (AUC$_{struc}$

= 0.599, P<0.001). In parallel the specificity and sensitivity of the four combined models are higher in comparison to the structural models (Table 4.15b; Table 4.13b).

Model/ Variable	Estimate	*P*	SE	VIF
LCM-ALL$_{comb}$				
Intercept	-3.065	<.001	0.616	-
LCM-131	0.094	<.001	0.025	1.082
LSIDI	2.097	0.007	0.782	2.139
LCM-171	0.012	0.020	0.005	2.275
LCM-41	-0.011	0.022	0.005	1.594
SIMI_MD	-0.005	0.026	0.002	1.258
L11_IIC	0.878	0.036	0.418	1.377
LCM-42	-0.009	0.041	0.005	1.519
LCM-11	0.018	0.052	0.009	2.343
NLAND	0.093	0.055	0.048	1.682
ECON_AM	-0.012	0.186	0.009	3.232
LCM-43	-0.494	0.294	0.471	1.006
L52_varIICconn	0.094	0.667	0.218	1.021
LCM-EAG1$_{comb}$				
Intercept	-6.12	0.013	2.49	-
L161_IIC	2.158	0.008	0.817	1.272
LCM-11	0.047	0.001	0.015	1.026
L21_IIC	-1.077	0.006	0.392	1.076
LCM-172	0.085	0.032	0.040	1.352
L52_IIC	1.531	0.011	0.606	1.026
ECON_RA	0.032	0.008	0.012	1.182
PROX_MN	-0.055	0.042	0.027	1.118
ENN_SD	0.004	0.072	0.002	1.110
CIRCLE_AM	8.650	0.018	3.660	1.662
GYRATE_CV	-0.033	0.017	0.014	2.782
AREA_AM	0.056	0.015	0.023	3.764

Table 4.14a: The contribution of landscape composition, connectivity and structural parameters for determining the presence-absence of all butterfly species (LCM-ALL$_{comb}$) and butterfly species within EAG1 (LCM-EAG1$_{comb}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

Model/ Variable	Estimate	*P*	SE	VIF
LCM-EAG2$_{comb}$				
Intercept	3.189	<.001	0.944	-
LCM-172	-0.087	<.001	0.023	1.342
LCM-11	-0.038	<.001	0.011	1.021
LCM-52	-0.101	<.001	0.026	1.091
LCM-51	-0.045	<.001	0.012	1.739
LCM-171	-0.040	<.001	0.010	1.827
LCM-42	-0.036	0.001	0.011	2.082
LCM-131	-0.088	0.004	0.030	1.088
LCM-41	-0.028	0.008	0.011	1.173
L11_IIC	0.168	0.813	0.710	1.342
LCM-EAG3$_{comb}$				
Intercept	-2.580	0.356	2.790	-
L11_IIC	4.550	0.005	1.640	1.011
LCM-11	0.055	0.005	0.020	1.006
PROX_CV	0.011	0.007	0.004	1.095
SHAPE_AM	-3.660	0.030	1.690	1.090

Table 4.14b: The contribution of landscape composition, connectivity and structural parameters for determining the presence-absence of butterfly species within EAG2 (LCM-EAG2$_{comb}$) and EAG3 (LCM-EAG3$_{comb}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

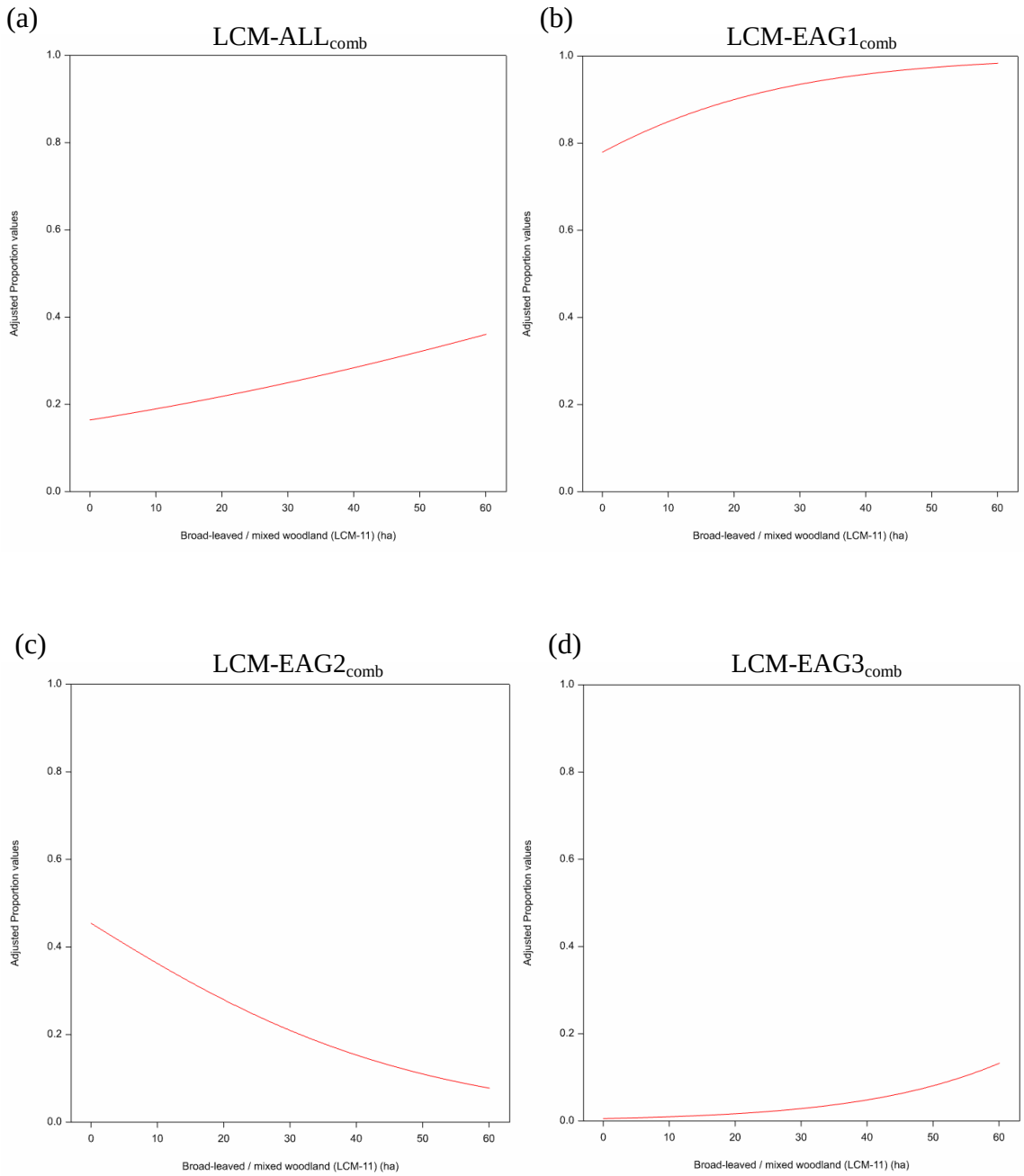


Figure 4.1a-d: Fitted relationship between landscape variables and the probability of butterfly occurrence adjusted for the influence of the other variables in each model. The relationship between the probability of occurrences and the area of broad-leaved woodland from the (a) LCM-ALL$_{comb}$ model, (b) LCM-EAG1$_{comb}$ model, (c) LCM-EAG2$_{comb}$ and (d) LCM-EAG3$_{comb}$ model.

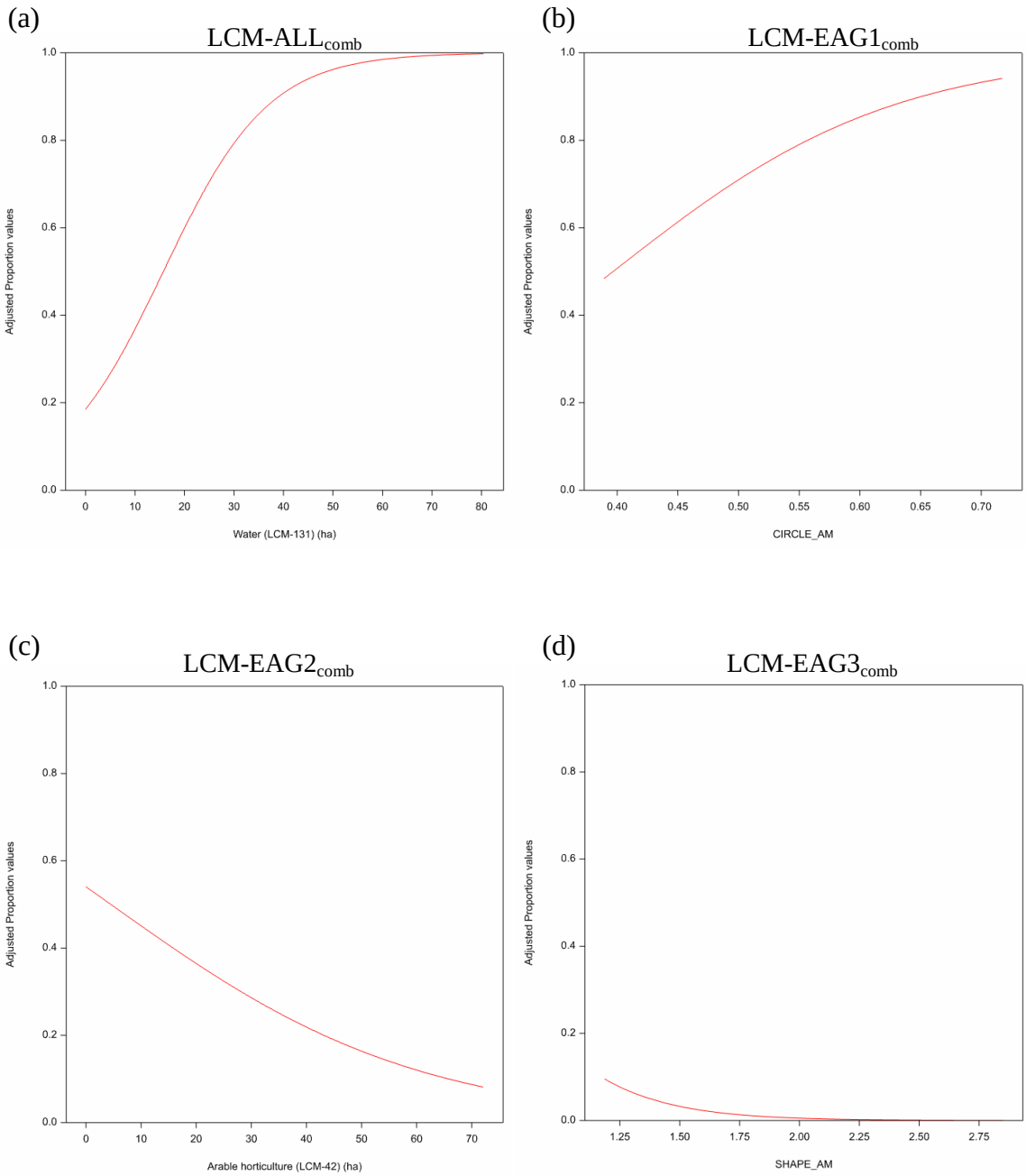


Figure 4.2a-d: Fitted relationship between landscape variables and the probability of butterfly occurrence adjusted for the influence of the other variables in each model. The relationship between the probability of occurrences and (a) the area of water from the LCM-ALL$_{comb}$ model, (b) the area-weighted average circle index (CIRCLE-AM) from the LCM-EAG1$_{comb}$ model, (c) the area of arable horticulture from the LCM-EAG2$_{comb}$ model and (d) the area-weighted shape index (SHAPE_AM) from the LCM-EAG3$_{comb}$ model.

(a)

Model	Model deviance			Goodness of fit (H-L)		
	X^2	*df*	*p*	X^2	*df*	*p*
LCM-ALL	177	12	<0.001	11.318	8	0.184
LCM-EAG1	76.2	11	<0.001	8.276	8	0.407
LCM-EAG2	68.2	9	<0.001	5.984	8	0.649
LCM-EAG3	25.54	4	<0.001	6.751	8	0.564

(b)

Model	Discrimination		Confusion matrix		
	AUC	*p*	*T*	SPC (%)	TPR (%)
LCM-ALL	0.671	0.000	0.2	61.4	61.2
LCM-EAG1	0.760	0.000	0.8	70.9	64.8
LCM-EAG2	0.698	0.000	0.4	63.7	64.7
LCM-EAG3	0.823	0.000	0.03	78.7	76.9

Table 4.15: Performance and accuracy of combined models for predicting presence-absence of all butterfly species ($LCM\text{-}ALL_{comb}$), and species comprising the three ecological attribute groups ($LCM\text{-}EAG1_{comb}$; $LCM\text{-}EAG2_{comb}$ and $LCM\text{-}EAG3_{comb}$). (a) Model fit determined by the model deviance (X^2) and the goodness of fit assessed by the H-L test statistic (X^2). Significance level (*p*) and degrees of freedom (*df*) are provided. (b) The discrimination of the models is provided in terms of area under the ROC curve (AUC) and the specificity (SPC) and sensitivity (True Positive Rate; TPR) as determined by threshold (*T*) equal to the prevalence rate for each model.

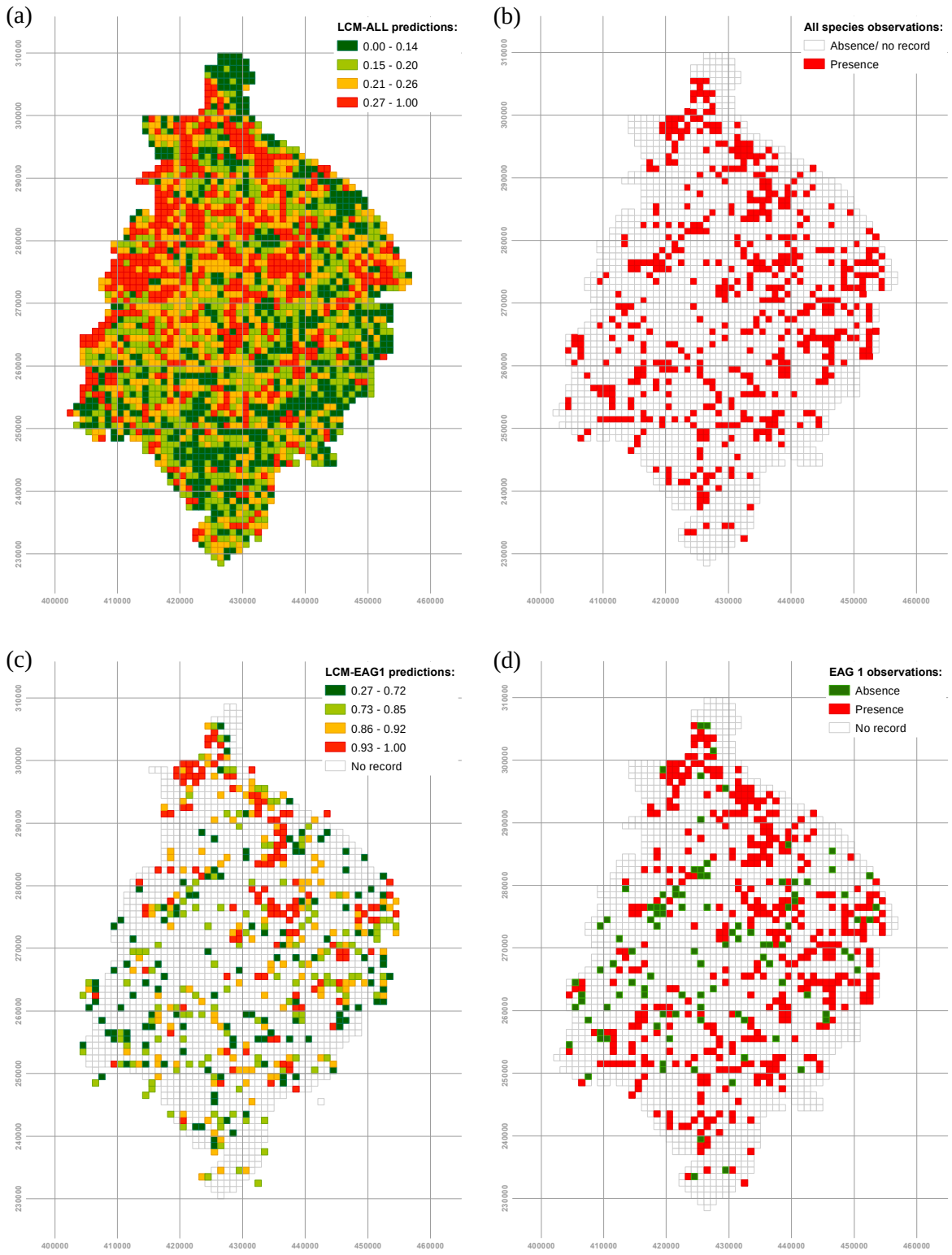


Figure 4.3 *(cont.)*

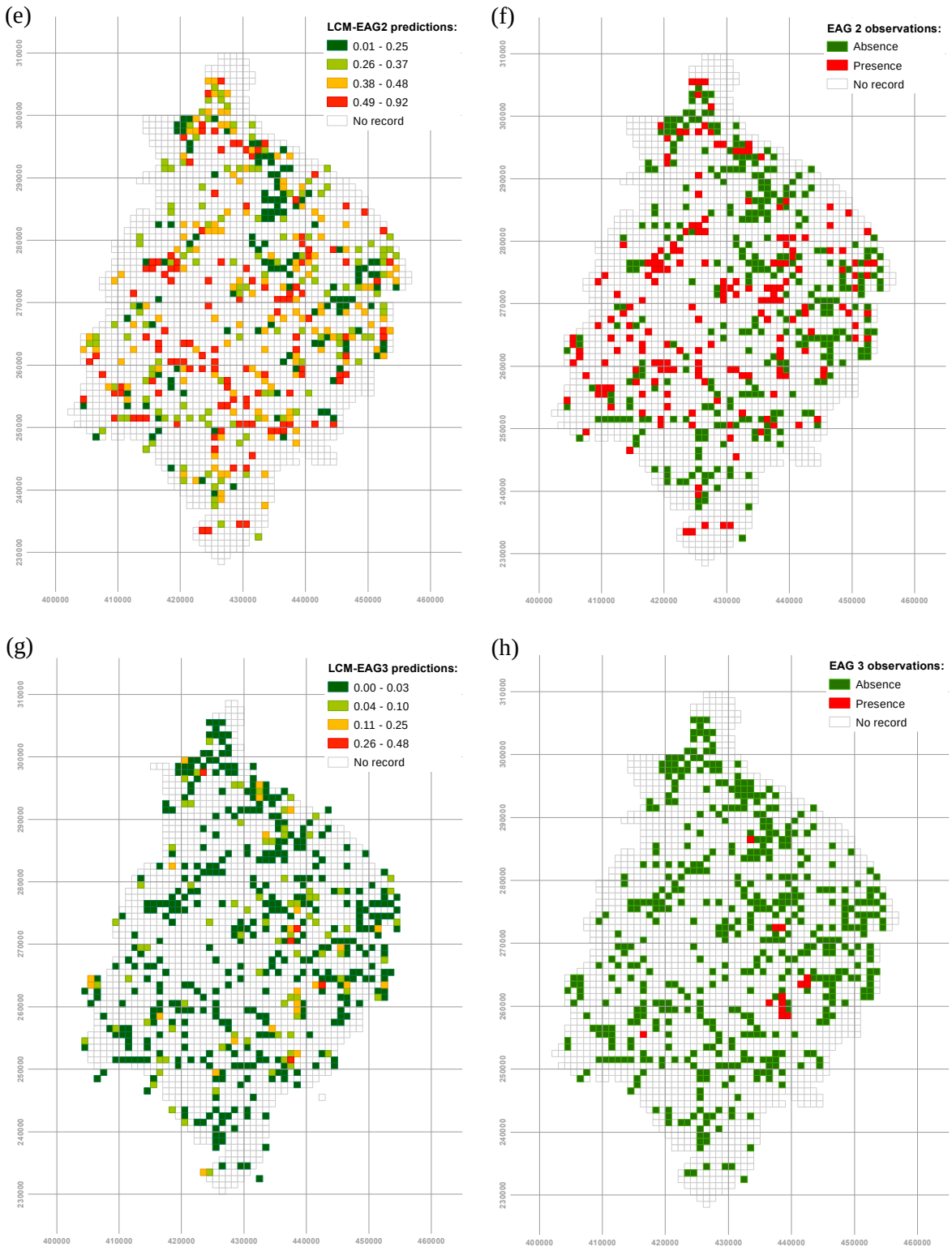


Figure 4.3: Comparison of the probability of butterfly occurrence derived from the four LCM combined landscape models to the observed butterfly presence-absence data for Warwickshire 1990-1999. The predicted values from (a) LCM-ALL$_{comb}$ (c) LCM-EAG1$_{comb}$ (e) LCM-EAG2$_{comb}$ and (g) LCM-EAG3$_{comb}$ models are compared to the observed presence-absence data for (b) all butterfly species (d) EAG1 species (f) EAG2 species and (h) EAG3 species. The quartile ranges for the model predicted values (a,c,e and g) are provided.

4.2.6 Landscape compositional models: PH1 2000

The PH1 2000 habitat map for Warwickshire classifies a total of 41 habitats (semi-natural and built), with arable land (PH-34) comprising the greatest average coverage per grid square (44.50 ha) (Table 2.5). Improved grassland (PH-19) and built up areas (PH-40) also dominate the landscape in comparison to the other habitats (25.34 ha/1 km grid square and 10.83 ha/ 1 km grid square respectively). The remaining 38 habitats have an average coverage per grid square of less than 5 ha across Warwickshire. Despite the low average coverage, ten habitats occur within more than half the grid squares, most notably semi-improved neutral grassland (PH-16), and broad-leaved semi-natural woodland (PH-1) occur within more than 70 % of the grid squares (85 % and 75 % respectively). Of the three habitats with greatest average coverage per grid square (PH-34; PH-19 and PH-40), improved grassland (PH-19) has the largest spatial distribution occurring within 96 % of grid squares.

When considering the 466 grid squares with butterfly records, there is little change in the average coverage of habitats per grid square, with arable, improved grassland and built up areas still dominating the landscape (35.06 ha, 24.91 ha and 13.08 ha respectively). The spatial distribution of habitats, however, has changed with an increased proportion of grid squares comprising semi-improved neutral grassland (PH-16; 92 %), dense/continuous scrub (PH-7; 73 %) and broad-leaved plantation woodland (PH-1; 70 %).

Model parameterisation

The probability of butterfly presence for all species (n = 2079) and for species associated with the three ecological attribute groups (n = 466) was significantly modelled as function of landscape composition derived from PH1 2000 habitat data (Table 4.16; Table 4.17a). Several PH-1 habitats were selected from the forward stepwise/ backward elimination procedure to be included within the final compositional models, with the PH1-EAG3$_{comp}$ model comprising the smallest number of variables (six in total).

Of the 12 variables associated with the occurrence of all butterfly species (PH1-ALL$_{comp}$), only three were unique to this model; inundation vegetation (PH-28), mixed semi-natural woodland (PH-5) and arable land (PH-34), with the last two

habitats exhibiting a negative influence on the occurrence of all butterfly species, and inundation vegetation exhibiting a positive influence (Table 4.16a). The remaining nine habitats within the PH1-ALL_{comp} model also influenced the occurrence of EAG species, but none of these occurred in all four models. Most notably the habitat broad-leaved semi-natural woodland (PH-1) positively influenced the occurrence of EAG2 and EAG3 species in addition to all butterfly species. Furthermore, increasing area of dense/continuous scrub (PH-7) positively influenced the occurrence of EAG1 and EAG3 species in addition to all butterfly species.

The three EAG models comprise different combinations of grassland habitats. A number of grassland habitats were included in the PH1-$EAG1_{comp}$ model, all of which had a positive influence on butterfly occurrence; unimproved acidic grassland (PH-13), semi-improved acidic grassland (PH-14), unimproved neutral grassland (PH-15), and semi-improved neutral grassland (PH-16). Unimproved acidic grassland (PH-13) and unimproved neutral grassland (PH-15) also positively influenced the occurrence of EAG2 species, in addition to improved grassland (PH-19). The PH1-$EAG2_{comp}$ model also comprised unimproved calcareous grassland (PH-17); however increases in area of this habitat negatively influenced the occurrence of EAG2 species. The model for EAG3 species comprised fewer grassland habitats than the other two EAG models, and these were semi-improved neutral grassland (PH-16), unimproved calcareous grassland (PH-17) and semi-improved calcareous grassland (PH-18), all of which positively influenced EAG3 butterfly occurrence.

A small number of habitats associated with the built environment were included in the models, including the land cover quarry (PH-31) which positively influenced the occurrence of EAG1 and all butterfly species (PH1-ALL_{comp}), and the land cover refuse tip (PH-33) which positively influenced with the occurrence of EAG2 species.

The parameters within all four models fit the data at an acceptable level, with non-significant H-L test statistic obtained for each model indicating there is no evidence for differences between the observed and model-predicted values (H-L $p > 0.5$) (Table 4.17a). 'Excellent' discrimination between presence-absence grid squares was obtained for the PH1-$EAG3_{comp}$ model ($AUC_{comp} = 0.947$, $p < 0.001$) (Table 4.17b). When considering the AUC_{comp} values obtained for the other three models, these

models classified the grid squares as present/absent significantly better than by chance ($p<0.001$), and performance of these models was considered to be fair with AUC_{comp} values obtained within the range of 0.7-0.8. The exceptional ability of the PH1-$EAG3_{comp}$ model to accurately predict presence-absence was also demonstrated by high model sensitivity (84.6 %) and specificity (90.5 %). The specificity of the PH1-ALL_{comp}, PH1-$EAG1_{comp}$ and PH1-$EAG2_{comp}$ models to accurately predict absences was also high (71.8 %; 70.1 % and 76.9 % respectively), however sensitivity of these models was poor in comparison (56.4 %; 68.8 % and 58.3 % respectively).

The variance inflation factors (VIF) for the four models were close to 1 indicating no strong correlations between independent variables (Table 4.16a,b). Within the PH1-$EAG1_{comp}$ model the standard errors and parameter estimates are particularly high for the variables unimproved acidic grassland (s.e. 168), semi-improved acidic grassland (s.e. 267), and introduced shrub (s.e. 428) (Table 4.16a). These variables are also insignificant in the model ($p>0.05$), suggesting that they may not be significantly contributing to the explanatory power of the model. However, the residual deviance of the model significantly increases with the individual removal of the variables introduced shrub ($\Delta D_1 = 4.1$, $p = 0.044$), semi-improved acidic grassland ($\Delta D_1 = 7.4$, $p = 0.007$), and unimproved acidic grassland ($\Delta D_1 = 6.7$, $p = 0.010$).

Model/ Variable	Estimate	*P*	SE	VIF
PH1-ALL$_{comp}$				
Intercept	-1.917	<.001	0.310	-
PH-7	0.290	<.001	0.050	1.195
PH-1	0.049	<.001	0.011	1.068
PH-34	-0.009	<.001	0.003	1.505
PH-29	0.060	0.002	0.020	1.036
PH-31	0.052	0.017	0.022	1.022
LSIDI	0.992	0.026	0.444	1.704
PH-16	0.017	0.037	0.008	1.170
PH-5	-0.876	0.111	0.550	1.006
PH-11	0.253	0.181	0.189	1.005
PH-13	1.330	0.192	1.020	1.010
PH-28	1.750	0.200	1.350	1.009
PH-17	0.549	0.231	0.458	1.109
PH1-EAG1$_{comp}$				
Intercept	1.193	0.017	0.499	-
PH-7	0.816	<.001	0.229	1.135
PH-16	0.094	0.002	0.031	1.122
PH-12	-11.030	0.007	4.110	1.003
PH-11	-0.862	0.029	0.396	1.009
PH-15	3.810	0.090	2.240	1.063
NLAND	-0.086	0.100	0.052	1.298
PH-31	0.272	0.238	0.231	1.037
PH-14	114.000	0.685	280.000	1.011
PH-13	66.000	0.691	167.000	1.023
PH-39	115.000	0.781	415.000	1.002

Table 4.16a: The contribution of landscape compositional variables for determining the presence-absence of all butterfly species (PH1-ALL$_{comp}$) and butterfly species within EAG1 (PH1-EAG1$_{comp}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

Model/ Variable	Estimate	*P*	SE	VIF
PH1-EAG2$_{comp}$				
Intercept	0.131	0.569	0.025	-
PH-1	0.181	<.001	0.066	1.097
PH-6	-0.784	0.006	0.374	1.033
LSIDI	0.013	0.019	0.005	1.106
PH-19	0.413	0.019	0.204	1.050
PH-15	5.910	0.036	7.610	1.078
PH-17	-2.370	0.043	3.250	1.010
PH-12	-0.047	0.116	0.038	1.005
PH-22	-3.390	0.154	3.240	1.163
PH-29	1.170	0.206	0.822	1.046
PH-13	0.131	0.295	0.025	1.096
PH-11	-1.671	0.437	0.712	1.041
PH-33	5.840	0.465	3.720	1.016
PH1-EAG3$_{comp}$				
Intercept	-5.726	<.001	0.735	-
PH-16	0.089	0.001	0.028	1.067
PH-18	0.841	0.002	0.273	1.094
PH-1	0.079	0.002	0.025	1.026
PH-17	1.370	0.022	0.596	1.226
PH-7	0.294	0.088	0.172	1.368
PH-27	-6.370	0.144	4.360	1.051

Table 4.16b: The contribution of landscape compositional variables for determining the presence-absence of butterfly species within EAG2 (PH1-EAG2$_{comp}$) and EAG3 (PH1-EAG3$_{comp}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

(a)

$Model_{comp}$	Model deviance			Goodness of fit (H-L)		
	X^2	*df*	*p*	X^2	*df*	*p*
PH1-ALL	226.0	12	<0.001	9.097	8	0.334
PH1-EAG1	94.0	10	<0.001	7.296	8	0.505
PH1-EAG2	108.7	12	<0.001	6.807	8	0.558
PH1-EAG3	50.3	6	<0.001	3.258	8	0.917

(b)

$Model_{comp}$	Discrimination		Confusion matrix		
	AUC	*p*	*T*	SPC (%)	TPR (%)
PH1-ALL	0.702	<0.001	0.2	71.8	56.4
PH1-EAG1	0.771	<0.001	0.7	70.1	68.8
PH1-EAG2	0.754	<0.001	0.4	76.9	58.3
PH1-EAG3	0.947	<0.001	0.1	90.5	84.6

Table 4.17: Performance and accuracy of the compositional models for predicting presence-absence of all butterfly species ($PH1\text{-}ALL_{comp}$), and species comprising the three ecological attribute groups ($PH1\text{-}EAG1_{comp}$; $PH1\text{-}EAG2_{comp}$ and $PH1\text{-}EAG3_{comp}$). (a) Model fit determined by the model deviance (X^2) and the goodness of fit assessed by the H-L test statistic (X^2). Significance level (*p*) and degrees of freedom (*df*) are provided. (b) The discrimination of the models is provided in terms of area under the ROC curve (AUC) and the specificity (SPC) and sensitivity (True Positive Rate; TPR) as determined by threshold (*T*) equal to the prevalence rate for each model.

.

4.2.7 Landscape connectivity models: PH1 2000

The connectivity metrics included during the modelling process measured the connectivity of the 21 key PH1 habitats which were identified from the four compositional models in section 4.2.6 (Table 4.18). The connectivity of woodland/ hedgerow network (P1H_IIC) and grassland/ hedgerow network (P1G_IIC) were also included during model development. As previously outlined in section 2.2.4 and section 4.2.3, the connectivity metrics were considered in addition to the compositional variables during the connectivity modelling procedure. Forwards selection/ backwards elimination procedure was conducted on a total of 23 IIC metrics, which measured connectivity within each grid square, and a total of two varIIC metrics, which measured the connectivity across the whole of Warwickshire (Table 4.18).

From a total of 25 connectivity metrics, capturing the connectivity of 21 key PH1 habitats, seven connectivity metrics are included within the final four connectivity models, and these were IIC metrics measuring the habitat connectivity within each grid square (Table 4.18; Table 4.19). Average connectivity of the woodland / hedgerow network (P1H_IIC) occurred within the connectivity models for all species model (n = 2079), and EAG species (n = 466). The connectivity of this woodland/ hedgerow network within all grid squares and occupied squares was higher than the other connectivity metrics included within the models (P1H_IIC = 0.504; P1H_IIC = 0.451 respectively) (Table 4.19). The average connectivity of the combined grassland / hedgerow network (PGH_IIC) within the occupied squares (n = 466) is also high (PGH_IIC = 0.406) in comparison to the connectivity of semi-improved acidic grassland (P14_IIC = 0.019). The connectivity of all six (unique) habitats ranged greatly across the grid square landscapes, from no connectivity for each habitat (IIC = 0) to well-connected patches for each habitat occurring within grid squares with IIC values of > 0.778 obtained across all six metrics.

Metric	PH1-ALL	PH1-EAG1	PH1-EAG2	PH1-EAG3
P1_IIC	○		○	○
P5_IIC	○			
P6_IIC			○	
P7_IIC	○○	○		○
P11_IIC	○○○	○	○	
P12_IIC		○	○	
P13_IIC	○○○	○	○	
P14_IIC		○○○		
P15_IIC		○	○	
P16_IIC	○	○○		○
P16_varIICconn	○	○○		○
P17_IIC	○		○	○
P18_IIC				○
P19_IIC			○	
P22_IIC			○	
P27_IIC				○
P28_IIC	○○○			
P29_IIC	○		○	
P31_IIC	○	○○		
P33_IIC			○	
P34_IIC	○			
P34_varIIC	○○			
P39_IIC		○		
P1H_IIC	○○○		○○○	
PGH_IIC	○	○○○		○

Table 4.18: Connectivity metrics for the 21 key PH1 2000 habitats and two hedgerow network variables included within the connectivity modelling. × refers to metrics which strongly correlated with area of corresponding land cover classes and were not included during model development; ○ refers to metrics selected from the correlation analysis but not selected during the modelling procedure; ○○ refers to metrics selected during the modelling procedure but dropped from the final model; ○○○ refers to metrics retained within the final connectivity model; and no symbol indicates that the key habitat class for the metric did not occur in the corresponding compositional model.

Metric	Average	SEM	Min	Max
All butterflies (*n* = 2427)				
P1H_IIC	0.504	0.003	0.000	0.960
P11_IIC	0.003	0.001	0.000	0.859
P13_IIC	0.005	0.001	0.000	0.952
P29_IIC	0.320	0.006	0.000	0.999
EAGs (*n* = 515)				
P14_IIC	0.019	0.005	0.000	0.837
PGH_IIC	0.406	0.004	0.000	0.778
P1H_IIC	0.451	0.006	0.000	0.906

Table 4.19: The mean and range in seven connectivity metrics included within the landscape connectivity models for all butterfly species and species comprising each EAG.

Model parameterisation

Across the four connectivity models, seven connectivity metrics are included; however, no connectivity metrics are selected within the PH-EAG3$_{conn}$ model (Table 4.20b). The connectivity of woodlands/ hedgerows (P1H_IIC) occurs in both the PH1-ALL$_{conn}$ and PH1-EAG2$_{conn}$ models (Table 4.20a,b). This variable significantly influences butterfly occurrence in both cases; however increased woodland/ hedgerow connectivity negatively influences the occurrence of 'all butterfly' species, but positively influences the occurrence of EAG2 species. With the addition of the woodland/ hedgerow connectivity metric in the PH1-EAG2$_{conn}$ model, two variables from the PH1-EAG2$_{comp}$ model (Table 4.16b) were no longer included; the area of recently felled woodland (PH-11) was not selected and the area of improved grassland (PH-19) was dropped, without significant increase in residual deviance ($\Delta D_1 = 3.6$, $p = 0.058$).

When comparing the PH1-ALL$_{conn}$ model to the compositional model, the area of the habitat unimproved acidic grassland (PH-13) has been dropped from the model without significant increase in residual deviance ($\Delta D_1 = 1$, $p = 0.270$), whilst the connectivity of this habitat (P13_IIC) has been retained (Table 4.20a). Within the PH1-ALL$_{conn}$ model, the IIC metrics measuring the connectivity of the habitats inundation vegetation (P28_IIC) and recently felled woodland (P11_IIC) have been selected within the model in addition to the area of those habitats (Table 4.20a).

However, with the addition of these connectivity metrics the area of unimproved calcareous grassland has also been dropped from the model ($\Delta D_1 = 3$, $p = 0.107$).

The PH1-EAG1$_{conn}$ model comprises two IIC metrics which measure the connectivity of the grassland/ hedgerow network (PGH_IIC) and the habitat semi-improved acidic grassland (P14_IIC), in addition to the area of several grassland variables (PH-13, PH-14, PH-15 and PH-16). Increasing connectivity of the grassland/ hedgerow network negatively influenced occurrence of EAG1 species, whilst increasing area of grassland habitats positively influenced EAG1 species occurrence (Table 4.20).

The occurrence of 'all butterfly' species and EAG1 and EAG2 species was significantly modelled as function of landscape connectivity and composition (Table 4.21a). No connectivity metrics were included, however, within the PH1-EAG3 compositional model (Table 4.20b). For all three connectivity models, the model estimates fit the data at an acceptable level, with non-significant H-L test statistics indicating there is no evidence for differences between the observed and model-predicted values (Table 4.21a). Discriminatory power of the three models, as measured by the AUC$_{conn}$ ranges from 0.712 obtained for the PH1-ALL$_{conn}$ model to 0.777 obtained for the PH1-EAG1$_{conn}$ model (4.21b). Model discrimination is considered 'fair' for these three models with AUC$_{conn}$ values obtained within the range of 0.7><0.8. In comparison to the compositional models, small increases in discriminatory power were observed for the PH1-ALL$_{conn}$ model (AUC$_{conn}$ = 0.712; AUC$_{comp}$ = 0.702) and PH1-EAG1$_{conn}$ model (AUC$_{conn}$ = 0.777; AUC$_{comp}$ = 0.771). Model specificity is high for PH1-ALL$_{conn}$ (70.5 %) and PH1-EAG2$_{conn}$ (74.5 %) and sensitivity is highest for the PH1-EAG1$_{conn}$ model (71.3 %). Model sensitivity of the PH1-ALL$_{conn}$ model is poor in comparison however (58.2 %). In comparison to the compositional models, slight improvements are observed for all three models in their sensitivity, but these improvements are coupled with slight decreases in model specificity (Table 4.21b; Table 4.17b).

Within the PH1-EAG1$_{conn}$ model high standard errors are obtained for the explanatory variables unimproved acidic grassland (PH-13) (s.e 171), semi-improved acidic grassland (PH-14) (s.e 241), and introduced shrub (PH-39) (s.e 402). Acceptable Variance Inflation Factors (VIFs) are obtained for the explanatory

variables across all three connectivity models, indicating no evidence of co-linearity between variables (Table 4.21a,b).

Model/ Variable	Estimate	*P*	SE	VIF
PH1-ALL$_{conn}$				
Intercept	-1.263	<.001	0.388	-
PH-7	0.272	<.001	0.050	1.101
PH-29	0.074	<.001	0.022	1.089
PH-1	0.059	<.001	0.012	1.129
P1H_IIC	-1.488	0.005	0.525	1.248
PH-16	0.021	0.011	0.008	1.178
PH-34	-0.006	0.017	0.003	1.657
P13_IIC	1.915	0.021	0.827	1.017
PH-31	0.046	0.034	0.022	1.025
LSIDI	0.859	0.054	0.446	1.713
P11_IIC	-4.330	0.064	2.340	1.086
P28_IIC	-2.180	0.079	1.240	1.090
PH-11	0.782	0.082	0.449	1.077
PH-5	-0.946	0.092	0.561	1.008
PH-28	2.270	0.195	1.750	1.031
PH1-EAG1$_{conn}$				
Intercept	2.309	0.011	0.903	-
PH-7	0.842	<.001	0.233	1.145
PH-16	0.101	0.001	0.031	1.182
PH-12	-11.160	0.007	4.150	1.005
PH-11	-0.875	0.027	0.396	1.012
NLAND	-0.109	0.050	0.056	1.407
P14_IIC	-2.810	0.087	1.650	1.099
PH-15	3.810	0.089	2.240	1.069
PGH_IIC	-2.210	0.123	1.430	1.117
PH-31	0.254	0.259	0.225	1.039
PH-14	137.000	0.570	241.000	1.083
PH-13	68.000	0.692	171.000	1.027
PH-39	102.000	0.799	402.000	1.003

Table 4.20a: The contribution of landscape connectivity and compositional variables for determining the presence-absence of all butterfly species (PH1-ALL$_{conn}$) and butterfly species within EAG1 (PH1-EAG1$_{conn}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

Model/ Variable	Estimate	*P*	SE	VIF
PH1-EAG2$_{conn}$				
Intercept	-1.286	0.055	0.670	-
PH-1	0.107	<.001	0.025	1.218
PH-6	0.169	0.008	0.064	1.026
P1H_IIC	2.444	0.011	0.956	1.186
PH-15	-0.613	0.046	0.308	1.076
PH-22	1.677	0.046	0.839	1.131
PH-17	0.393	0.058	0.208	1.009
PH-13	-5.310	0.106	3.280	1.090
PH-12	5.730	0.117	3.660	1.005
LSIDI	-1.128	0.120	0.725	1.156
PH-29	-0.043	0.208	0.034	1.046
PH-33	-2.420	0.387	2.790	1.013
PH1-EAG3$_{conn}$				
Intercept	-5.726	<.001	0.735	-
PH-16	0.089	0.001	0.028	1.067
PH-18	0.841	0.002	0.273	1.094
PH-1	0.079	0.002	0.025	1.026
PH-17	1.370	0.022	0.596	1.226
PH-7	0.294	0.088	0.172	1.368
PH-27	-6.370	0.144	4.360	1.051

Table 4.20b: The contribution of landscape connectivity and compositional variables for determining the presence-absence of butterfly species within EAG2 (PH1-EAG2$_{conn}$) and EAG3 (PH1-EAG3$_{conn}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

(a)

$Model_{conn}$	Model deviance			Goodness of fit (H-L)		
	X^2	*df*	*p*	X^2	*df*	*p*
PH1-ALL	242	14	<0.001	11.988	8	0.152
PH1-EAG1	98.7	12	<0.001	3.971	8	0.860
PH1-EAG2	103.6	11	<0.001	3.266	8	0.917
PH1-EAG3*	50.3	6	<0.001	3.258	8	0.917

(b)

$Model_{conn}$	Discrimination		Confusion matrix		
	AUC	*p*	*T*	SPC (%)	TPR (%)
PH1-ALL	0.712	<0.001	0.2	70.5	58.2
PH1-EAG1	0.777	<0.001	0.7	67.0	71.3
PH1-EAG2	0.743	<0.001	0.4	74.5	60.0
PH1-EAG3*	0.947	<0.001	0.1	90.5	84.6

Table 4.21: Performance and accuracy of the connectivity models for predicting presence-absence of all butterfly species ($PH1\text{-}ALL_{conn}$), and species comprising the three ecological attribute groups ($PH1\text{-}EAG1_{conn}$; $PH1\text{-}EAG2_{conn}$ and $PH1\text{-}EAG3_{conn}$). (a) Model fit determined by the model deviance (X^2) and the goodness of fit assessed by the H-L test statistic (X^2). Significance level (*p*) and degrees of freedom (*df*) are provided. (b) The discrimination of the models is provided in terms of area under the ROC curve (AUC) and the specificity (SPC) and sensitivity (True Positive Rate; TPR) as determined by threshold (*T*) equal to the prevalence rate for each model.* Model is the same as the compositional model.

4.2.8 Landscape structural models: PH1 2000

A total of 39 landscape structure metrics were included within the forwards selection/ backwards elimination modelling procedure, and from this a total of 24 metrics were selected across the four landscape structure models (Table 4.22; Table 4.23). Of these 24 metrics, three metrics (AREA_MN, GYRATE_CV and CONNECT) were common to the $PH1\text{-}ALL_{struc}$ model and the EAG models (Table 4.22). Considering the 21 unique metrics across all four models, three are associated with the landscape aspect area, six measure patch shape, ten are associated with patch aggregation, one with edge contrast and one with landscape diversity (Table 4.22; Appendix A4 and A8). Several summary statistics were selected which

measure the variability and range associated with the average values obtained for the metrics GYRATE, CONTIG, SHAPE, ENN, PROX, and SIMI.

Across the grid square landscapes, the 24 landscape structure metric values vary considerably with a large range in values obtained, most notably for the aggregation metrics (SIMI, ENN, and PROX) in addition to the metrics which measure patch size (GYRATE and AREA). In particular, patch size (AREA_MN) ranges across the grid square landscapes from a minimum of 0.709 ha to a maximum of 100 ha, with one single patch therefore occupying the whole landscape extent (Table 4.23).

Metric	Aspect	PH1-ALL	PH1-EAG1	PH1-EAG2	PH1-EAG3
AREA_AM	Area	○	○	○	○
AREA_CV	Area	○	○	○○	○
AREA_MD	Area	○	×	×	×
AREA_MN	Area	○○○	○	○	○○○
CIRCLE_AM	Shape	○	○	○	○
CIRCLE_MD	Shape	○○	○	○	○
CIRCLE_MN	Shape	○○	○	○	○
CIRCLE_RA	Shape	○	○	○	○
CIRCLE_SD	Shape	○○	○	○	○
CONNECT	Aggregation	○○○	○	○	○○○
CONTIG_MN	Shape	×	○○○	○	○
CONTIG_RA	Shape	○	○	○	○○○
CONTIG_SD	Shape	○○	○○○	○○○	○○○
ECON_AM	Contrast	○	○	○	○
ECON_MN	Contrast	○	○○○	○	○
ECON_RA	Contrast	○	○	○○	○
ECON_SD	Contrast	○	○	○	○
ENN_AM	Aggregation	○	○○○	○	○
ENN_CV	Aggregation	○	○	○	○
ENN_MD	Aggregation	○	○	○	○
ENN_MN	Aggregation	○	○	○○○	○○○
ENN_SD	Aggregation	○○○	○	○	○
GYRATE_CV	Area	○○○	○○○	○	○
GYRATE_MN	Area	×	○	○○	○
GYRATE_RA	Area	○○○	○	○	○
IJI	Aggregation	○○	○	○	○
PRD	Contrast	×	○	○	○○○
PROX_AM	Aggregation	○○○	○	○	○○○
PROX_CV	Aggregation	○○	○○○	○○	○
PROX_MN	Aggregation	○○	○	○	○
SHAPE_AM	Shape	○	○	○	○○○
SHAPE_MD	Shape	○○	○	○	○
SHAPE_MN	Shape	○○○	○	○	○○○
SHAPE_SD	Shape	○	○○○	○○	○
SIMI_AM	Aggregation	○○○	○	○○	○
SIMI_CV	Aggregation	○○	○	○○○	○
SIMI_MD	Aggregation	×	○	○	○
SIMI_MN	Aggregation	○○○	○	○○	○
SIMI_RA	Aggregation	○○○	×	×	×

Table 4.22: The 39 landscape structure metrics included within the forwards selection/ backwards elimination modelling procedure. × refers to metrics which strongly correlated with other metrics and were not included during model development; ○ refers to metrics selected from the correlation analysis but not selected during the modelling procedure; ○○ refers to metrics selected during the modelling procedure but dropped from the final model; ○○○ refers to metrics retained within the final structural model.

Metric	Average	SEM	Min	Max
All butterflies (*n* = 2427)				
AREA_MN (ha)	4.906	0.087	0.709	100.000
GYRATE_RA (m)	306.240	0.974	0.000	547.308
GYRATE_CV (%)	121.713	0.456	0.000	209.337
SHAPE_MN	1.343	0.002	1.000	1.843
PROX_AM	20.977	0.616	0.000	161.364
SIMI_MN	128.521	1.683	0.000	738.598
SIMI_AM	159.789	1.783	0.000	557.511
SIMI_RA	472.342	4.375	0.000	1389.361
ENN_SD (m)	164.893	1.622	0.000	636.451
CONNECT	46.581	0.326	0.000	100.000
EAGs (*n* = 515)				
AREA_MN (ha)	3.786	0.096	0.709	100.000
GYRATE_CV (%)	119.485	0.898	72.625	189.754
SHAPE_AM	1.845	0.012	1.127	2.925
SHAPE_SD	0.465	0.006	0.166	0.927
CONTIG_MN	0.408	0.004	0.201	0.716
CONTIG_RA	0.906	0.002	0.355	0.963
CONTIG_SD	0.292	0.001	0.134	0.381
PROX_CV	237.724	3.509	104.432	621.408
SIMI_CV (%)	131.772	1.463	66.116	251.798
ENN_MN	182.940	2.976	67.249	520.437
ENN_AM	107.324	2.318	0.920	385.323
ECON_MN	75.262	0.327	42.870	93.326
CONNECT	47.000	0.552	22.222	100.000
PRD	10.610	0.133	3.855	20.000

Table 4.23: The mean and range in 24 PH1 landscape structure variables included within the landscape structure models for all butterfly species and species comprising each EAG.

Model parameterisation

The presence-absence of butterfly species was significantly modelled as a function of landscape structure (Table 4.24; Table 4.25a). More than six structural metrics were selected in the PH1-ALL_{struc}, PH1-$EAG1_{struc}$ and PH1-$EAG3_{struc}$ models, however only a few landscape structure metrics were selected within the PH1-$EAG2_{struc}$ model (Table 4.24). Of the ten landscape structure metrics selected in the PH1-ALL_{struc} model, three occurred within the EAG models (AREA_MN, GYRATE_CV and CONNECT). The occurrence of ‘all butterfly’ species and EAG3 species were negatively associated with average patch size (AREA_MN), and ‘all butterfly’

species and EAG1 species were negatively associated with variability in mean patch extent (GYRATE_CV). The functional connectivity of patches within the landscape (CONNECT), significantly positively influenced the occurrence of 'all butterfly' species and EAG3 species. The standard deviation in patch boundary configuration (CONTIG_SD) was selected in all the EAG models. Increases in the variability of this shape metric positively influenced the occurrence of EAG2 and EAG3 species but negatively influenced the occurrence of EAG1 species.

All four models fit the data well, with non-significant H-L test statistic values obtained, indicating no evidence for differences between observed and model predicted values (Table 4.25a). Model discrimination, as measured by the AUC_{struc}, ranged from 0.621 – 0.922, with 'excellent' discrimination achieved for the PH1-$EAG3_{struc}$ model (Table 4.25b). Model specificity and sensitivity was also high for this model (82.3 % and 76.9 % respectively). The discrimination of the PH1-ALL_{struc}, PH1-$EAG1_{struc}$ and PH1-$EAG2_{struc}$ models were considered to be 'poor', with AUC_{struc} values within the range of 0.621 - 0.676 obtained. Model specificity was lowest for the PH1-$EAG1_{struc}$ model (37.1 %) however; the sensitivity of this model was more accurate than the other three models (85.6 %).

The standard errors of the variables across the four models were low with exception of CONTIG_RA in the PH1-$EAG3_{struc}$ model (s.e. 23.6), which reflects the high coefficient obtained for this variable (57.4) (Table 4.24). The variance inflation factors (VIF) were all sufficiently low, indicating no strong correlations between the landscape structure explanatory variables.

Model/ Variable	Estimate	*P*	SE	VIF
PH1-ALL$_{struc}$				
Intercept	-1.330	0.300	1.290	-
AREA_MN	-0.167	<.001	0.033	1.492
PROX_AM	-0.010	<.001	0.003	1.257
SIMI_AM	-0.003	0.003	0.001	1.734
SIMI_RA	0.001	0.005	0.000	2.725
GYRATE_RA	-0.004	0.020	0.002	2.177
CONNECT	0.011	0.022	0.005	1.106
SHAPE_MN	1.682	0.034	0.792	2.036
SIMI_MN	-0.003	0.036	0.001	2.627
ENN_SD	0.002	0.057	0.001	1.098
GYRATE_CV	-0.004	0.414	0.005	3.082
PH1-EAG1$_{struc}$				
Intercept	9.970	0.002	3.160	-
CONTIG_SD	-11.590	0.006	4.250	1.016
CONTIG_MN	-5.750	0.009	2.210	2.152
SHAPE_SD	2.810	0.010	1.080	1.271
ECON_MN	-0.036	0.099	0.020	1.410
ENN_AM	0.004	0.118	0.003	1.125
PROX_CV	-0.003	0.125	0.002	1.125
GYRATE_CV	-0.012	0.172	0.009	2.111
PH1-EAG2$_{struc}$				
Intercept	-6.270	<.001	1.260	-
CONTIG_SD	14.480	<.001	3.620	1.011
SIMI_CV	0.008	0.012	0.003	1.011
ENN_MN	0.003	0.067	0.002	1.000
PH1-EAG3$_{struc}$				
Intercept	-75.500	0.004	26.000	-
SHAPE_MN	13.910	0.003	4.710	1.278
AREA_MN	-1.648	0.006	0.605	3.633
CONNECT	0.104	0.006	0.038	1.439
ENN_MN	0.019	0.008	0.007	1.456
SHAPE_AM	-6.090	0.014	2.490	1.222
CONTIG_RA	57.400	0.015	23.600	1.354
PROX_AM	-0.049	0.112	0.031	1.120
CONTIG_SD	21.300	0.158	15.100	1.357
PRD	0.159	0.391	0.185	2.667

Table 4.24: The contribution of landscape structural variables for determining the presence-absence of all butterfly species (PH1-ALL$_{struc}$) and butterfly species within each ecological attribute group (PH1-EAG1$_{struc}$; PH1-EAG2$_{struc}$ and PH1-EAG3$_{struc}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

(a)

$Model_{struc}$	Model deviance			Goodness of fit (H-L)		
	X^2	*df*	*p*	X^2	*df*	*p*
PH1-ALL	152	10	<0.001	10.945	8	0.205
PH1-EAG1	30.3	7	<0.001	9.288	8	0.319
PH1-EAG2	24	3	<0.001	11.161	8	0.193
PH1-EAG3	43.17	9	<0.001	8.644	8	0.373

(b)

$Model_{struc}$	Discrimination		Confusion matrix		
	AUC	*p*	*T*	SPC (%)	TPR (%)
PH1-ALL	0.676	<0.001	0.2	67.4	59.6
PH1-EAG1	0.669	<0.001	0.7	37.1	85.6
PH1-EAG2	0.621	<0.001	0.4	57.7	61.1
PH1-EAG3	0.922	0.000	0.03	82.3	76.9

Table 4.25: Performance and accuracy of the landscape structure models for predicting presence-absence of all butterfly species (PH1-ALL_{struc}), and species comprising the three ecological attribute groups (PH1-$EAG1_{struc}$; PH1-$EAG2_{struc}$ and PH1-$EAG3_{struc}$). (a) Model fit determined by the model deviance (X^2) and the goodness of fit assessed by the H-L test statistic (X^2). Significance level (*p*) and degrees of freedom (*df*) are provided. (b) The discrimination of the models is provided in terms of area under the ROC curve (AUC) and the specificity (SPC) and sensitivity (True Positive Rate; TPR) as determined by threshold (*T*) equal to the prevalence rate for each model.

4.2.9 Combined models: PH1 2000

The development of the PH1 combined landscape models, incorporated the compositional and connectivity variables from the connectivity models developed in section 4.2.7, in addition to the landscape structure variables from the four structural models developed in section 4.2.8. A total of 49 landscape variables (22 compositional variables; six connectivity metrics; and 21 structural variables) were considered during the forwards selection/ backwards elimination procedure for the development of the combined models.

Model parameterisation

A number of landscape explanatory variables occur in more than one combined model of butterfly occurrence (Table 4.26). In particular, increasing area of broad-leaved semi-natural woodland is important in determining the presence of 'all butterfly' species combined and species comprising EAG2 and EAG3 (Figure 4.4a-c). This relationship with butterfly occurrence is significantly positive in all three models; however, the relationship is weaker with EAG3 species, when considering the influence of other parameters in the PH1-EAG3$_{comb}$ model (Table 4.26b; Figure 4.4c). Additionally, increasing area of the habitat dense/ continuous scrub (PH-7) is significantly related to the occurrence of all butterfly species combined and EAG1 and EAG3 species.

The PH1-ALL$_{comb}$ combined model comprised 20 variables, which measure the composition, connectivity and structure of the landscape (Table 4.26a). All 14 variables from the connectivity model were retained in the final combined model (Table 4.26a; Table 4.20a). The occurrence of all butterfly species was significantly positively influenced by several variables, including the area of broad-leaved semi-natural woodland (PH-1), semi-improved neutral grassland (PH-16), quarry (PH-31), dense/ continuous scrub (PH-7), and standing water (PH-29). In particular, the probability of 'all species' was significantly strongly associated with increasing area of standing water (PH-29) (Figure 4.4d). Increasing area of arable land (PH-34) significantly negatively influenced 'all species' butterfly occurrence (Table 4.26a).

The combined model also comprised four connectivity metrics, with increasing connectivity of unimproved acidic grassland (P13_IIC) significantly positively influencing butterfly occurrence. Increasing connectivity of the woodland/ hedgerow

network (P1H_IIC), recently felled woodland (P11_IIC) and inundated vegetation (P28_IIC) negatively influenced butterfly occurrence. The combined models also included six structural variables; two area metrics, one shape metric and three aggregation metrics. The relationship with butterfly occurrence was significant for five of these metrics, with positive relationships observed with functional connectivity of patches (CONNECT), and the standard deviation of the Euclidean distance between patches of the same patch type (ENN_SD). Negative relationships were observed between the occurrence of butterfly species and mean patch size (AREA_MN), the range in patch extent (GYRATE_RA) and the size and proximity between patches of the same patch type (PROX_MN).

The PH1-EAG1$_{comb}$ model comprised 14 landscape variables (Table 4.26a). Of the 12 variables from the landscape connectivity model ten are included in the final combined model, however the two grassland connectivity metrics (PGH_IIC and P14_IIC) were not selected during the forwards stepwise and backwards elimination procedure (Table 4.26a; Table 4.20a). Of the compositional variables, five were significant in the PH1-EAG1$_{comb}$ model. The variables unimproved acidic grassland (PH-13), semi-improved acidic grassland (PH-14) and introduced shrub (PH-39) which had high standard errors in the PH1-EAG1$_{comp}$ and PH1-EAG1$_{conn}$ models were recoded as binary variables representing the presence-absence of those habitats. Incorporating these grassland variables as binary variables reduced the standard errors of these parameters. The increasing area of dense scrub (PH-7), semi-improved neutral grassland (PH-15), and unimproved neutral grassland (PH-16) were significantly positively associated with the occurrence of EAG1 species. The habitat ‘quarry’ (PH-31) also positively influenced the occurrence of EAG1 species, however this relationship was non-significant. Increasing area in the habitats orchard (PH-12) and recently felled woodland (PH-11) significantly negatively influenced the occurrence of EAG1 species (Figure 4.4e). EAG1 occurrence was also negatively associated with a high number of land cover classes (NLAND); however, this relationship was not significant. The PH1-EAG1 combined model also included four of the seven structural variables from the structural model, which included two shape metrics, one aggregation metric and one contrast metric. The increase in standard deviation in average patch shape (SHAPE_SD) was significantly positively associated with occurrence of EAG1 species. The variability in patch boundary

configuration (CONTIG_SD) negatively influenced EAG1 species occurrence, but this relationship was non-significant. Increases in the variability in average size and proximity between patches of the same type (PROX_CV) and average edge contrast (ECON_MN) also negatively influenced the occurrence of EAG1 species.

The PH1-EAG2$_{comb}$ combined model comprised ten landscape variables (Table 4.26b). Most notably eight of these were compositional variables. The probability of occurrence of EAG2 species was significantly positively influenced by the area of semi-natural broad-leaved woodland (PH-1), and mixed plantation woodland (PH-6). Furthermore, increasing connectivity of woodlands/ hedgerow network (P1H_IIC) also significantly positively influenced EAG2 species occurrence (Figure 4.4f). A number of non-significant positive relationships with EAG2 species occurrence were observed which were with the area of orchard (PH-12), continuous bracken (PH-22), and unimproved calcareous grassland (PH-17). Non-significant negative relationships were also observed with the increasing areas of unimproved acidic grassland (PH-13), unimproved neutral grassland (PH-15) and standing water (PH-29). The combined model also included one structural metric, with the occurrence of EAG2 species significantly associated with increases in the variability in the average size and proximity between similar patch types (SIMI_CV).

The PH1-EAG3 combined model comprised seven parameters, all of which were significantly associated with the occurrence of EAG3 species (Table 4.26b). Six of the landscape variables were measures of landscape composition, with increasing areas of semi-natural woodland (PH-1) and dense scrub (PH-7) positively influencing the occurrence of EAG3 species. A number of grassland variables also positively influenced the occurrence of EAG3 species; semi-improved neutral grassland (PH-16), unimproved calcareous grassland (PH-17) (Figure 4.4g) and semi-improved calcareous grassland (PH-18). An increase in the area of swamp habitat (PH-27) was negatively associated with EAG3 occurrence. The combined model also comprised one structural variable, the area-weighted average patch shape (SHAPE_AM), which negatively influenced EAG3 species occurrence.

The parameters of the four models fit the data well with non-significant H-L test statistics obtained indicating no evidence for differences between observed and model predicted values (Table 4.27a). The predictions from the PH1 combined

models can been seen to match the observed data at an acceptable level when comparing the probabilities of occurrence derived from the model and the observations of presence-absence within the original data (Figure 4.5a-h). High probabilities of butterfly occurrence match the classification of presence squares, this match is highest for the PH1-EAG3 model (TPR = 92.3 %) and PH1-EAG1 model (77.8 %) (Table 4.27b; Figure 4.5). Classification of absence squares was most accurate for the PH1-EAG2 model (73.1 %) and PH1-EAG3 model (89.0 %). The model predicted values for the PH1-EAG3 model can be seen to match the observed values; however, the majority of these observations are absence squares (Figure 4.5g,h).

This good discrimination between presence-absence for the four models is also reflected in the AUC_{comb} values obtained for the modes (Table 4.27b). The AUC_{comb} values of the four models ranged from 0.720 for the $PH1\text{-}ALL_{comb}$ model to 0.963 for the $PH1\text{-}EAG3_{comb}$ model. The discrimination of the $PH1\text{-}ALL_{comb}$ and $PH1\text{-}EAG2_{comb}$ models were considered 'fair' (AUC_{comb} = 0.720; AUC_{comb} = 0.752 respectively), and the $PH1\text{-}EAG1_{comb}$ model considered 'good' (AUC_{comb} = 0.806). 'Excellent' discrimination was obtained for the $PH1\text{-}EAG3_{comb}$ model (AUC_{comb} = 0.963). For all four combined models the discriminatory power is higher than that obtained when modelling butterfly presence-absence as a function of separate landscape components. This improvement is largest when comparing to the landscape structure models, most notably for the PH1-EAG1 model (AUC_{comb} = 0.806, $p<0.001$; AUC_{stru} = 0.669, $p <0.001$).

Model/ Variable	Estimate	*P*	SE	VIF
PH1-ALL$_{comb}$				
Intercept	-1.260	0.276	1.160	-
PH-7	0.276	<.001	0.050	1.112
PH-29	0.070	<.001	0.022	1.096
PH-1	0.059	<.001	0.012	1.141
PROX_AM	-0.009	0.001	0.003	1.584
AREA_MN	-0.110	0.003	0.037	1.884
PH-34	-0.008	0.004	0.003	1.736
CONNECT	0.014	0.004	0.005	1.126
L_16_AREA	0.020	0.021	0.008	1.190
GYRATE_RA	-0.004	0.022	0.002	1.881
P1H_IIC	-1.248	0.025	0.557	1.322
PH-31	0.050	0.026	0.023	1.029
ENN_SD	0.002	0.039	0.001	1.115
P13_IIC	1.685	0.041	0.826	1.020
P11_IIC	-4.470	0.065	2.420	1.088
P28_IIC	-2.280	0.070	1.260	1.093
PH-5	-0.981	0.091	0.581	1.009
LSIDI	-1.088	0.105	0.671	3.643
PH-11	0.707	0.116	0.449	1.079
PH-28	2.360	0.198	1.830	1.034
SHAPE_MN	1.502	0.689	0.029	1.313
PH1-EAG1$_{comb}$				
Intercept	5.540	0.010	2.160	-
PH-7	0.814	<.001	0.228	1.190
PH-16	0.100	0.002	0.032	1.164
PH-12	-11.020	0.007	4.100	1.010
PH-11	-0.899	0.023	0.396	1.018
SHAPE_SD	2.350	0.025	1.050	1.152
PH-15	4.340	0.057	2.280	1.076
CONTIG_SD	-8.730	0.067	4.760	1.125
Noland	-0.101	0.084	0.058	1.549
PROX_CV	-0.003	0.145	0.002	1.125
ECON_MN	-0.028	0.162	0.020	1.161
PH-31	0.256	0.253	0.224	1.041
PH-13*	11.800	0.360	12.900	1.136
PH-14*	8.100	0.624	16.500	1.058
PH-39*	7.300	0.633	15.200	1.018

Table 4.26a: The contribution of landscape composition, connectivity and structural variables for determining the presence-absence of all butterfly species (PH1-ALL$_{comb}$) and butterfly species within EAG1 (PH1-EAG1$_{comb}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided. * recoded as binary variables.

Model/ Variable	Estimate	*P*	SE	VIF
PH1-EAG2$_{comb}$				
Intercept	-3.491	<.001	0.687	-
PH-1	0.091	<.001	0.022	1.050
P1H_IIC	3.068	0.001	0.952	1.088
SIMI_CV	0.010	0.005	0.003	1.044
PH-6	0.158	0.008	0.060	1.003
PH-22	1.530	0.059	0.809	1.010
PH-15	-0.570	0.062	0.305	1.016
PH-17	0.349	0.100	0.212	1.050
PH-12	5.820	0.105	3.590	1.005
PH-29	-0.063	0.115	0.040	1.130
PH-13	-4.770	0.133	3.180	1.044
PH1-EAG3$_{comb}$				
Intercept	7.480	0.070	4.120	-
PH-16	0.141	<.001	0.040	1.068
PH-1	0.093	0.002	0.030	1.040
SHAPE_AM	-8.430	0.003	2.880	1.033
PH-7	0.506	0.005	0.182	1.369
PH-17	2.410	0.005	0.854	1.227
PH-18	0.622	0.012	0.248	1.108
PH-27	-13.380	0.019	5.700	1.055

Table 4.26b: The contribution of landscape composition, connectivity and structural variables for determining the presence-absence of EAG2 and EAG3 species (PH1-EAG2$_{comb}$; PH1-EAG3$_{comb}$). Variables were identified from forward selection and backwards elimination and are listed in order of significance for each model. The Variance Inflation Factor (VIF) of each parameter is provided.

(a)

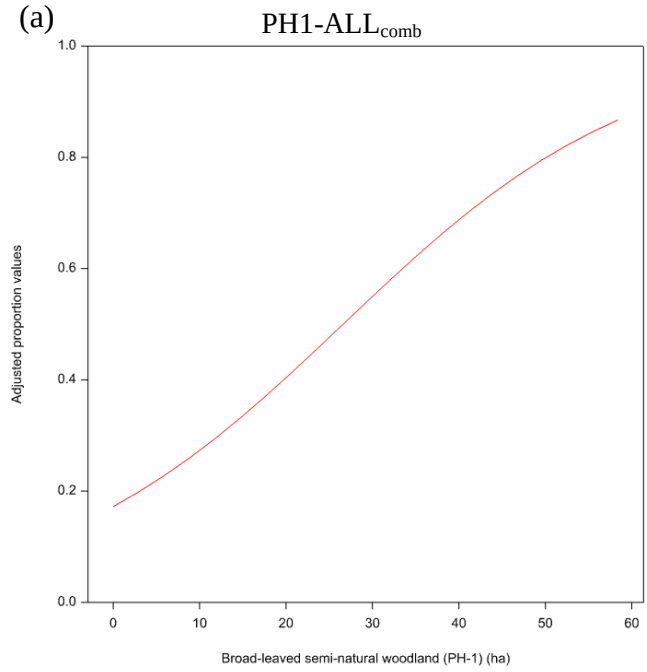


Figure 4.4 *(cont.)*

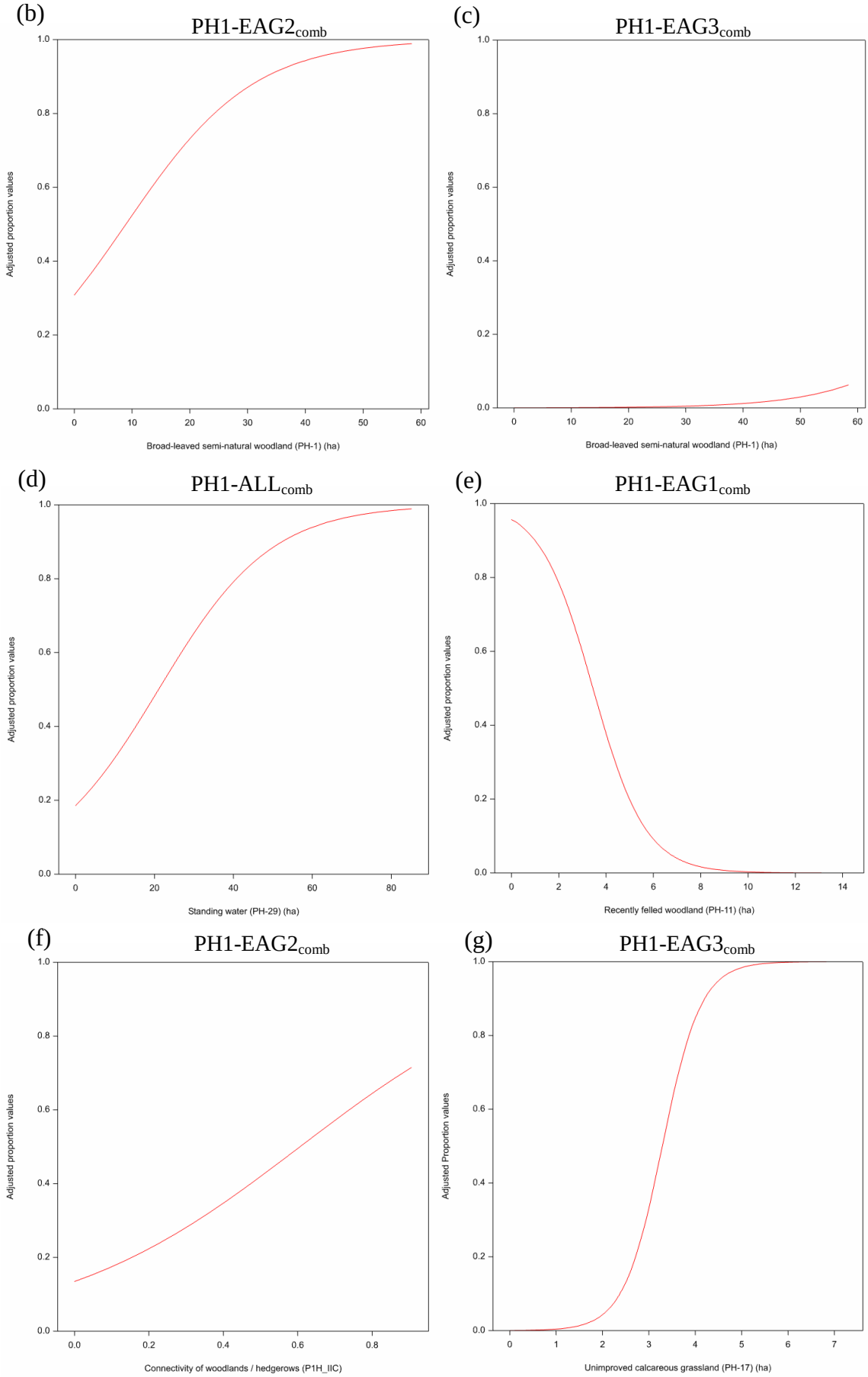


Figure 4.4a-g: Fitted relationship between landscape variables and the probability of butterfly occurrence. The relationship between the probability of occurrences and the area of broad-leaved woodland from the (a) PH1-ALL$_{comb}$, (b) PH1-EAG2$_{comb}$ and (c) PH1-EAG3$_{comb}$ models. The relationship between the probability of occurrences and (d) the area of water from PH1-ALL$_{comb}$, (e) area of recently felled woodland from PH1-EAG1$_{comb}$ (f) connectivity of woodland/hedgerow from PH1-EAG2$_{comb}$ and (g) area of unimproved calcareous grassland from PH1-EAG3$_{comb}$.

(a)

$Model_{comb}$	Model deviance			Goodness of fit (H-L)		
	X^2	*df*	*p*	X^2	*df*	*p*
PH1-ALL	269.0	20	<0.001	7.107	8	0.525
PH1-EAG1	106.9	14	<0.001	4.801	8	0.779
PH1-EAG2	104.7	10	<0.001	8.422	8	0.393
PH1-EAG3	63.02	7	<0.001	4.338	8	0.825

(b)

$Model_{comb}$	Discrimination		Confusion matrix		
	AUC	*p*	*T*	SPC (%)	TPR (%)
PH1-ALL	0.720	<0.001	0.2	68.4	60.9
PH1-EAG1	0.806	<0.001	0.7	61.9	77.8
PH1-EAG2	0.752	<0.001	0.4	73.1	61.1
PH1-EAG3	0.963	<0.001	0.03	89.0	92.3

Table 4.27: Performance and accuracy of the combined models for predicting presence-absence of all butterfly species ($PH1\text{-}ALL_{comb}$), and species comprising the three ecological attribute groups ($PH1\text{-}EAG1_{comb}$; $PH1\text{-}EAG2_{comb}$ and $PH1\text{-}EAG3_{comb}$). (a) Model fit determined by the model deviance (X^2) and the goodness of fit assessed by the H-L test statistic (X^2). Significance level (*p*) and degrees of freedom (*df*) are provided. (b) The discrimination of the models is provided in terms of area under the ROC curve (AUC) and the specificity (SPC) and sensitivity (True Positive Rate; TPR) as determined by threshold (*T*) equal to the prevalence rate for each model.

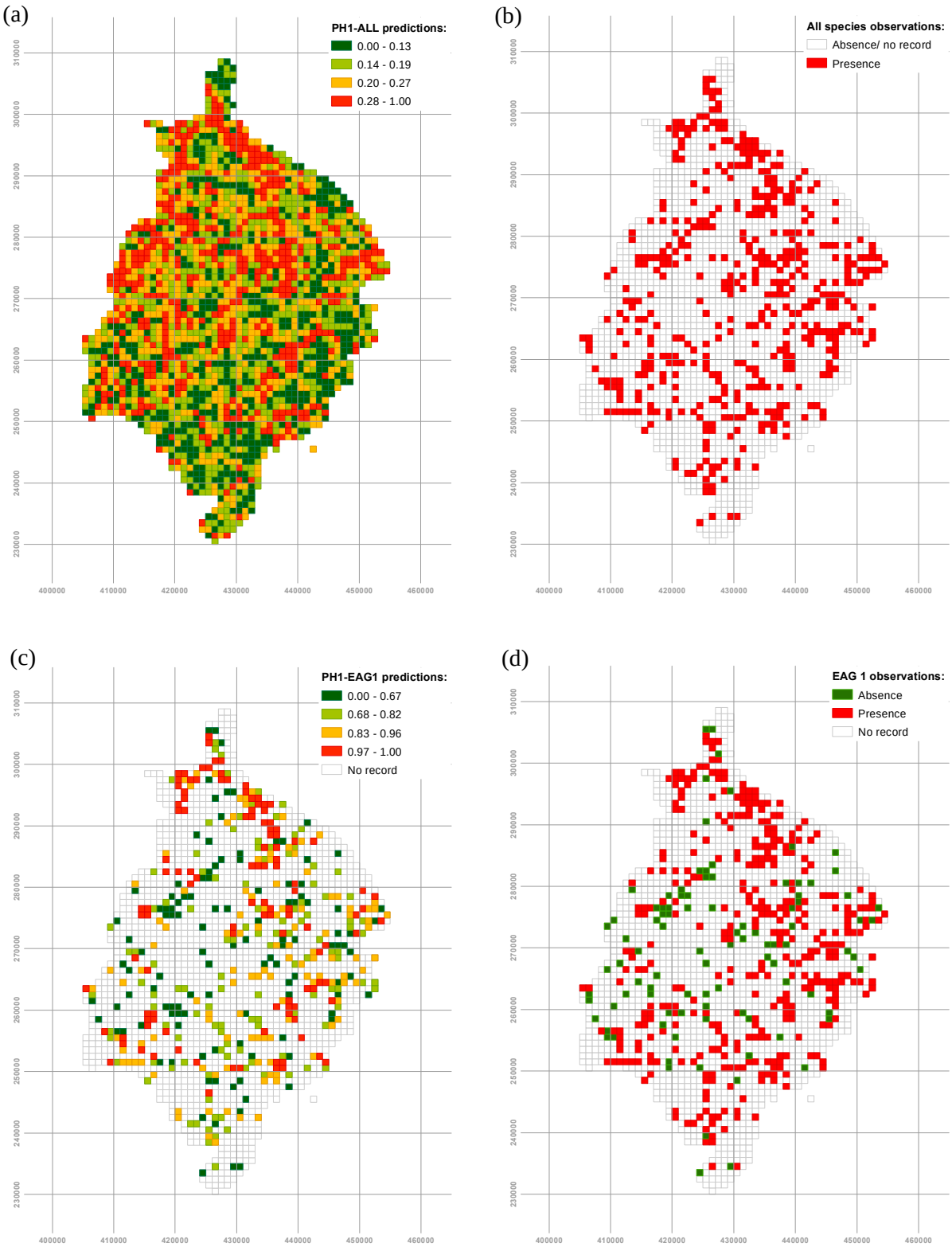


Figure 4.5 *(cont.)*

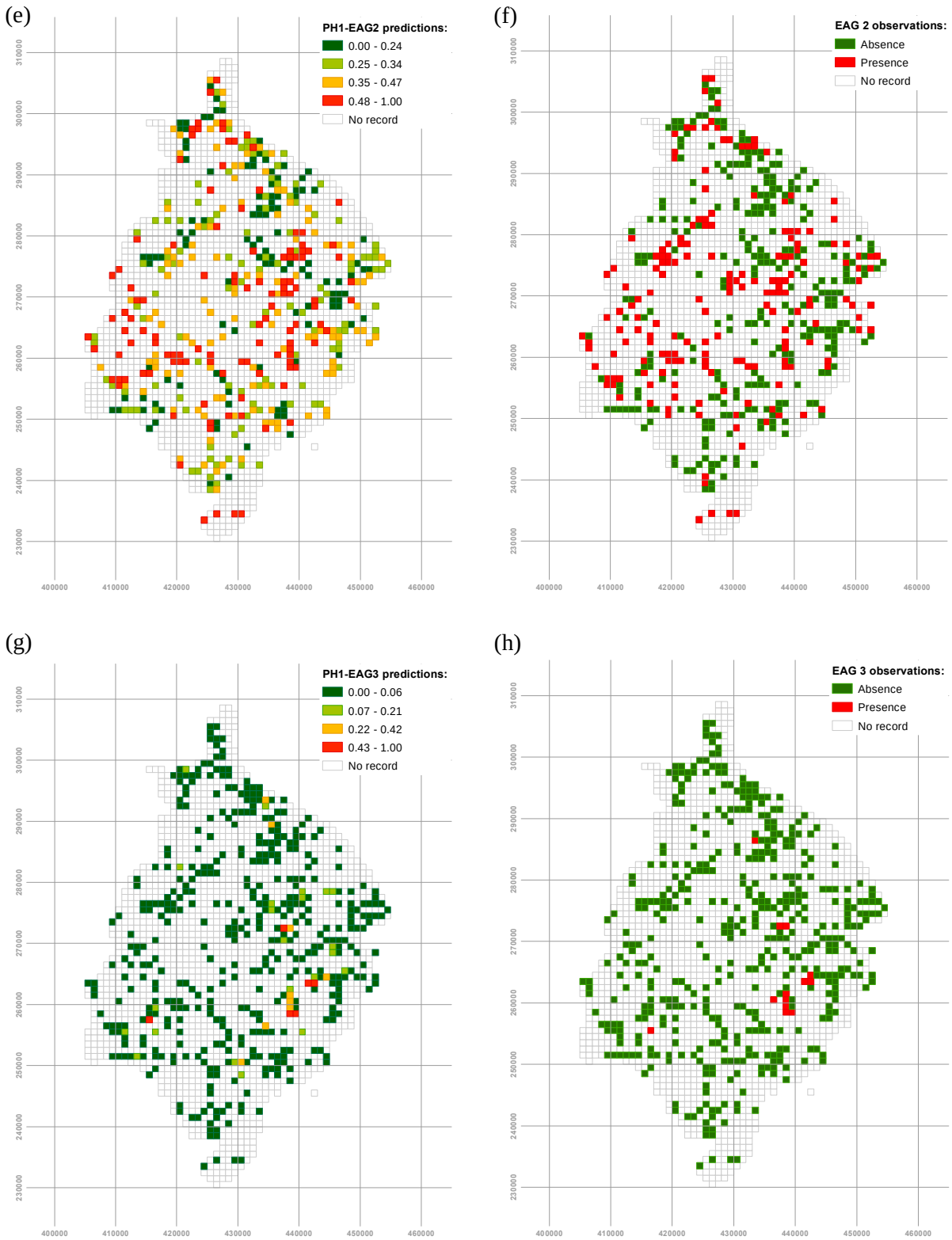


Figure 4.5a-h: Comparison of the probability of butterfly occurrence derived from the four PH1 combined landscape models to the observed butterfly presence-absence data for Warwickshire 1990-1999. The predicted values from (a) PH1-ALL_{comb} (c) PH1-$EAG1_{comb}$ (e) PH1-$EAG2_{comb}$ and (g) PH1-$EAG3_{comb}$ models are compared to the observed presence-absence data for (b) all butterfly species (d) EAG1 species (f) EAG2 species and (h) EAG3 species. The quartile ranges for the model predicted values (a,c,e) are provided.

4.2.10 Comparison between PH1 2000 and LCM 2000 combined models

Similarities can be drawn when comparing the PH1 and LCM combined models, developed in sections 4.2.5 and 4.2.9, in terms of the landscape variables which are required to parameterise the models (Appendix A9). Correlations were calculated to assess the relationship between the compositional, structural and connectivity metrics derived from the PH1 habitats and the equivalent LCM broad habitats included within each model. Only five of the variables which were common to the PH1 and LCM models were significantly correlated, and most correspondence occurred between the two data types for the variables within the 'all butterfly species' model.

The area of semi-natural broad-leaved woodland (PH-1) as classified by the PH1 habitat map positively correlated with the LCM broad habitat equivalent, the area of broad-leaved mixed woodland (LCM-11) (S = 0.504, p<0.001). Woodland area was an important variable in the models for 'all butterfly' species and EAG3 species derived from both the PH1 2000 and the LCM 2000 data sets. Both classifications of woodland habitat (LCM-11 and PH-1) positively influenced the occurrence of 'all butterfly' and EAG3 species (Figure 4.4a,c; Figure 4.1a,d). Woodland area was also important in the PH1-EAG2$_{comb}$ and LCM-EAG2$_{comb}$ models; however the influence on butterfly occurrence contrasted between the two data sources, with an increase in the area of woodland positively influencing EAG2 species occurrence in the PH1-EAG2$_{comb}$ model, but negatively influencing occurrence in the LCM-EAG2$_{comb}$ model (Table 4.14b; Table 4.26b). Despite the contrasting influence of Woodland area, the connectivity of woodland as measured by the IIC metric within the two data sets (L11_IIC and P1H_IIC) both positively influenced EAG2 species occurrence, however, these two variables are not significantly correlated (S = -0.008, p = 0.708). The woodland connectivity metric for the PH1 model incorporates the interaction with hedgerows (P1H_IIC), which are not considered in the LCM metric equivalent (L11_IIC).

In particular, the LCM-ALL$_{comb}$ model includes three arable land cover types (LCM-41; LCM-42 and LCM-43), and the PH1-ALL$_{comb}$ model comprises the habitat 'arable land' (PH-34) which incorporates all types of arable land cover (Table 4.14a; Table 4.26a). In both models increases in the area of these arable habitats have a negative influence on butterfly occurrence. The broad habitats 'arable cereals' (LCM-41) and 'arable horticulture' (LCM-42), are strongly significantly correlated

with the phase 1 arable habitat (PH-34) (S = 0.750, p<0.001; S = 0.670, p<0.001 respectively). The LCM-EAG2$_{comb}$ model also comprises arable land covers (LCM-41 and LCM-42), however, arable land is not included in the PH1-EAG2$_{comb}$ model (Table 4.14b; Table 4.26b).

There is direct correspondence between the area of standing water from the two data sets, which occurs within the PH1-ALL$_{comb}$ and LCM-ALL$_{comb}$ models (Figure 4.4d; Figure 4.2a), as well as the PH1-EAG2$_{comb}$ and LCM-EAG2$_{comb}$ models (Table 4.14; Table 4.26). Increasing area of standing water positively influences the occurrence of all butterflies according to both model types (PH1 and LCM) but negatively influences EAG2 species occurrence in both data sets. The area of standing water as classified by the two data sets is weakly correlated (S = 0.309, p<0.001). The diversity of land cover types was important in both the LCM-ALL$_{comb}$ and PH1-ALL$_{comb}$ model; however the relationship with butterfly occurrence differed between the two data sources. Increasing landscape diversity negatively influences butterfly occurrence according to the PH1-ALL$_{comb}$ model but positively influences occurrence in the LCM-ALL$_{comb}$ model. Despite the contrasting influence on butterfly occurrence between the two models, the diversity of phase 1 habitats and the diversity of LCM land covers is strongly correlated (S = 0.636, p<0.001).

There is little correspondence in terms of the model parameters when comparing the LCM-EAG1$_{comb}$ and PH1-EAG1$_{comb}$ models (Table 4.14a; Table 4.26a). Most notably, the PH1-EAG1$_{comb}$ model includes four grassland habitats, however, the LCM-EAG1$_{comb}$ model only comprises one grassland variable, the connectivity of set aside grassland (L52_IIC). No significant correlations occur between the PH1 grassland habitats and the LCM variable L52_IIC (>0.05). Woodland habitats are important in the LCM-EAG1$_{comb}$ model, with the area of broad-leaved woodland (LCM-11) and the connectivity of coniferous woodland (L21_IIC), influencing EAG1 species occurrence. In contrast, the PH1-EAG1$_{comb}$ model does not comprise any woodland habitat.

Little correspondence occurred between the data types in terms of the structural metrics within each model. Only one structural metric was important for a particular species model according to both data sources. The metric, SHAPE_AM was important for EAG3 species within the LCM-EAG3$_{comb}$ and the PH1-EAG3$_{comb}$

models, with a negative relationship with butterfly occurrence in both instances (Figure 4.2d; Table 4.26b). This metric was not significantly correlated between the two data sources however (S = 0.101, p<0.001). The EAG1 models comprised several structural metrics, with the metric PROX and ECON occurring in both models but different summary statistics associated with these measures. For example, the PH1-EAG1$_{comb}$ includes the metric average edge contrast (ECON_MN) and the LCM-EAG1$_{comb}$ includes the metric range in edge contrast (ECON_RA). No significant correlations were observed, however, between the structural metrics included in the two EAG1 models (p>0.05).

Comparison of model accuracy

When accounting for the degrees of freedom in each model, the PH1 models for the butterfly species EAGs have higher model deviance than the LCM EAG models, in particular, the model deviance for the PH1-EAG2$_{comb}$ model (X^2_{10} = 104.7, p<0.001) is much higher than that obtained for the LCM-EAG2$_{comb}$ model (X^2_9 = 68.2, p<0.001), despite the extra degree of freedom (Table 4.27a; Table 4.15a). Model deviance for the LCM-ALL$_{comb}$ model (X^2_{12} = 177, p<0.001) is higher than that obtained for the PH1-ALL$_{comb}$ model (X^2_{20} = 269, p<0.001), when considering the difference in the degrees of freedom.

When considering the specificity and sensitivity of the combined models, model accuracy for the PH1-EAG3$_{comb}$ model is higher in comparison to the LCM-EAG3$_{comb}$ (Table 4.27b; Table 4.15b). The specificity of the PH1-ALL$_{comb}$ and PH1-EAG2$_{comb}$ models are higher than the LCM equivalents, whereas the sensitivity of the LCM-ALL$_{comb}$ and LCM-EAG2$_{comb}$ models are higher in comparison. For the EAG1 combined models, specificity is highest for the LCM model, whereas sensitivity is highest for the PH1 model. Model discrimination as measured by the AUC$_{comb}$ is much higher for each PH1 model including the PH1-ALL$_{comb}$ model, with the discrimination of each PH1 model classified within a higher category than the LCM equivalent in terms of model performance (Table 4.27b; Table 4.15b). For example, the discrimination of the PH1-EAG1$_{comb}$ is considered to be ‘good’ (AUC$_{comb}$ = 0.806), and in comparison the discrimination of the LCM-EAG1$_{comb}$ model is ‘fair’ (AUC$_{coomb}$ = 0.760). All PH1 models have discriminatory power of ‘fair’ or higher, whereas two of the LCM models are considered to be ‘poor’.

4.3 Discussion

This chapter demonstrates that the presence-absence of butterfly species can be reliably predicted as a function of landscape based components. Models were developed considering the relationship between landscape components and all species combined and by species grouped by their ecological attributes. Species-landscape associations differed, however, between these four species groups. Additionally, model accuracy and performance differed depending upon the type of landscape explanatory variables (composition, connectivity, and structural) used and the thematic resolution of the landscape data from which explanatory variables were derived. In particular, models of butterfly occurrence based on landscape structure variables performed poorly when considering the discriminating ability of the models as assessed by the AUC statistic. Model performance was improved when the landscape components were considered together in a combined model for both data types. These findings are in agreement with the general consensus within the literature that a diverse array of independent landscape measures are required to capture the complexity of the landscape and develop relationships with biodiversity (Robinson, *et al.*, 2014; Schindler, *et al.*, 2008).

Although the combination of landscape parameters differed amongst the models, a small number of individual landscape parameters influenced butterfly occurrence across a number of species groups and the two data types. In particular the areas of the phase 1 habitat broad-leaved semi-natural woodland (PH-1) and the LCM equivalent broad-leaved/ mixed woodland (LCM-11) had positive influence on the occurrence of most butterfly groups. The only species group where LCM woodland was not a positive influence on (EAG2) was instead positively influenced by the connectivity of LCM woodland (L11_IIC). Positive association between several butterfly species and woodland habitat are widely documented, in particular for the species comprising EAG2 (Shreeve, *et al.*, 2001; Tudor, *et al.*, 2004). The distribution of EAG2 species within Warwickshire, including the purple hairstreak *(Neozephyrus quercus)* and the white letter hairstreak *(Satyrium w-album)* are dependent on woodland and scrub habitat throughout their life cycle, utilising trees (Elm *sp*. and Oak *sp*. respectively) and shrubs for oviposition and larval development as well as daily adult activities such as feeding and basking (Shreeve, *et al.*, 2001). As well as documented associations between EAG2 species and woodland, there are

many examples of the presence of other EAG species being associated with woodland habitats. For example, a number of EAG1 species will utilise woodlands, in particular the speckled wood *(Pararge aegeria),* small pearl-bordered fritillary (*Boloria selene*) and the pearl-bordered fritillary (*Boloria euphrosyne)* are often classified as woodland species outside of their ecological attribute group (open grassland). The ecological requirements of this group are for open grassland habitats, which often occur within woodland habitats, such as open glades and rides (Shreeve, *et al.*, 2001).

The dependency of butterflies on larval food plants and previous reporting of the presence of several EAGs in woodlands is the most likely reason for the strong association between butterfly occurrence and woodland as classified. In particular the PH1-EAG2 combined model includes the area of different types of woodland habitats in addition to broad-leaved woodland, including mixed plantation woodland (PH-6), and orchards (PH-12), as well as the connectivity of woodland/ hedgerows (P1H_IIC). The classification system of the LCM is associated with broad level habitats and as such is coarse in comparison (Cherrill, *et al.*, 1995), a potential reason why the LCM-EAG2 comprised a surprisingly negative association with woodland area. Although the two woodland classifications were significantly correlated this relationship was weak, and as such there are evident differences in the classification of this habitat. Furthermore, the negative relationship observed with woodland area in the LCM model, may be a surrogate for landscape diversity, and with increases in woodland area the diversity of the landscape decreases. As well as area of woodland it is clear from these models that connectivity is important, specifically in the LCM-EAG2 model when woodland area had a negative influence. This indicates that a network of small well-connected woodland patches is perhaps as beneficial for EAG2 species as the overall woodland area.

In addition to woodland area and connectivity, a variety of grassland habitat types were important in determining the distribution of the four species groups according to the PH1 models in contrast to the LCM models. The classical grassland EAGs, EAG1 and EAG3 were both positively correlated with the area of several different PH1 grassland habitats. The ecological requirements of the EAG1 butterfly species are strongly related to open grassland habitats, with a variety of grassy species, such as meadow grasses (*Poa spp*.), providing feeding, roosting and basking sites and tall

herbs providing sites for mate location (Shreeve, *et al.*, 2001). As several different grass species can be utilised as host plants, EAG1 species can be found within a variety of grassland habitats, in particular, the meadow brown (*Maniola jurtina)* and the small skipper (*Thymelicus sylvestris*) often occupy field margins and roadside verges in addition to grasslands (Warmington and Vickery, 2003). As such, the EAG1 species within Warwickshire were mostly widespread and common, occurring within the most grid squares in comparison to the other two EAGs, and in turn several grassland habitats were included within the PH1-EAG1$_{comb}$ model. Furthermore, half of the Warwickshire EAG1 species are monitored by the UKBMS as indicators of the wider countryside for grassland and/ or woodland habitats (Defra, 2012; Fox, *et al.*, 2011). For the PH1-EAG1$_{comb}$ model high standard errors were obtained for the grassland habitats unimproved and semi improved acidic grassland (PH-13 and PH-14), and this is likely due to the small distribution and variation in area of these habitats across the grid squares. When modelled as binary variables these grassland habitats were still selected within the final combined model, although this relationship was not significant, it suggests that the presence of this habitat is just as important as area.

In contrast, the LCM-EAG1$_{comb}$ model only comprised the grassland variable connectivity of set aside grassland (L52_IIC) which was positively associated with EAG1 species occurrence. Furthermore, only one additional grassland variable (improved grassland LCM-51) was important within the LCM models. It is evident that grassland habitats are of importance for numerous butterfly species, particularly those associated with EAG1 and EAG2. The limited importance of grassland habitats within the LCM models may be due to inaccuracies and misclassification of land cover, which is further supported by the non-significant correlations between several PH1 grassland habitats and grasslands classified by the LCM 2000. Different species-landscape associations, however, can be drawn from the LCM models which are also ecologically meaningful. In particular, for the LCM-EAG1$_{comb}$ model, the connectivity of inland bare ground (L161_IIC) and area of continuous urban (LCM-172) were positively associated with EAG1 occurrence. Positive associations with these habitats have been documented for several species, with bare ground providing opportunities for basking, and pockets of grassland and gardens within an urban

matrix providing nectar sources for the more generalist EAG1 species (Luoto, *et al.*, 2006; Warmington and Vickery, 2003).

Within the PH1-EAG3$_{comb}$ model positive associations with the grassland habitats semi-improved neutral grassland (PH-16), unimproved and semi-improved calcareous grassland (PH17, PH-18) occurred in combination with dense scrub (PH-7) and woodland (PH-1). The EAG3 species are most notably associated with open grassland comprising species-rich short turf, with larvae feeding and adult roosting occurring within the ground layer and basking occurring on bare earth (Shreeve, *et al.*, 2001). The species of EAG3, such as the brown argus (*Aricia agestis)* and the small blue (*Cupido minimus),* are most typically associated with calcareous grassland (Shreeve, *et al.*, 2001). The majority of the EAG3 species which occur in Warwickshire, however, are not limited to calcareous grassland and will utilise a range of open homogenous grassland habitats (Warmington and Vickery, 2003). Considering the primary association of EAG3 species with grassland, it is surprising that the LCM-EAG3$_{comb}$ model did not comprise any grassland cover. This could be explained by both the coarser classification of grasslands in the LCM 2000 in comparison to the PH1 2000, as well as by the strong correlation of EAG3 species with woodland in the LCM-EAG3 model. Across Warwickshire the three grassland habitats incorporated within the PH1-EAG3 model occur in combination with woodland in butterfly occupied sites.

Although woodland and grassland types were the main predictors of butterfly occurrence, there were other predictors such as standing water (LCM-131; PH-29), arable land (LCM-41-43; PH-34), and inland rock/ quarry (LCM-161; PH-31) which correlated with the occurrence of butterflies in the PH1-ALL$_{comb}$ and LCM-ALL$_{comb}$ models. A large continuous surface of water is considered to provide no resources for resting, with exception to the large white (*Pieris brassicae*) which has been identified to land and take off from water (Dennis & Hardy, 2007). Water availability, however, has been found to be positively associated with butterfly population density (Robinson, *et al.*, 2014). Additionally, these associations are likely to have arisen due to the surrounding bankside vegetation, which can provide ideal habitat for food plants adapted to boggy conditions. For example, the cuckoo flower (*Cardamine pratensis*) is the primary larval foodplant for the Orange tip (*Anthocharis cardamines*) butterfly, and adult nectar source for several species and Hemp-

agrimony (*Eupatorium cannabinum*) is also a vital adult nectar source (BC, 2012). The importance of abandoned quarries across Warwickshire is recognised by the designation of this habitat as a local biodiversity action plan habitat (LBAP). Furthermore, several abandoned limestone and sandstone quarries across Warwickshire have been designated as SSSIs and LNRs due to their importance for supporting diverse flora and fauna, invertebrate species in particular (Falk, 2003; Warmington and Vickery, 2003).

There were many predictors which were unique to models, which is not surprising considering the differences between butterfly requirements in each different EAG. For example, the PH1-EAG1$_{comb}$ model included introduced shrub habitat, which can provide important nectar sources for several EAG1 butterflies, with shrub beds particularly important for the Gatekeeper (*Pyronia tithonus*). The PH1-EAG2$_{comb}$ model included bracken habitat which was a positive influence on the occurrence of EAG2 species and bracken habitats have been identified to provide important microclimatic conditions for the growth of the larval food plant, the common dog violet (*Viola riviniana*), for several fritillary species, including the silver washed fritillary (*Argynnis paphia*) (Clarke, *et al.*, 2011). Within the LCM-EAG1$_{comb}$ model a few unique parameters were included, including the connectivity of in land bare rock associated with EAG1 species as previously discussed.

A similar number of structural metrics are important across the four LCM models and four PH1 models, with aggregation metrics the most common structural metric occurring within the models. Most notably, the PH1-ALL model comprised three aggregation metrics and the positive relationship with functional connectivity (CONNECT) and the negative relationships with patch extent (GYRATE_RA), and patch size (AREA_MN; PROX_MN) suggests that a network of well-connected small patches are important. This is not surprising considering the vast number of landscape compositional variables associated with the occurrence of all species in this model.

The structural aspects of the landscapes associated with butterfly occurrence also differed between species groups and between the different data types. Different structural metrics were important in the models when comparing the two data sources and no significant correlations occurred among the structural metrics

between these two landscape data sets. This is to be expected considering the differences in the level of precision of the classification of habitats/ land cover in the LCM and PH1 data sets (Bailey, *et al.*, 2007; Turner, *et al.*, 2001). Structural metrics, such as patch density and mean patch size have been shown to differ depending on thematic resolution, with an increase in patch density and a decrease in mean patch size as the number of land cover classes classified within the landscape increases (Castilla, *et al.*, 2009). The range in edge contrast (ECON_RA) was positively associated with EAG1 species in the LCM-EAG1 model, however average edge contrast (ECON_MN) was negatively associated with EAG1 species in the PH1-EAG1 model. The degree of edge contrast between neighbouring patches determines the permeability of the landscape for species movement (Ewers and Didham, 2006), and this impact is dependent upon species perception and the land covers considered. Within the LCM-EAG1$_{comb}$ model positive correlation with range in edge contrast (ECON_RA) is not surprising when considering the land cover variables in this model which are positively associated with EAG1 species occurrence, as these all vary greatly in their composition. For example, this model includes the areas of continuous urban (LCM-172) and broad-leaved woodland (LCM-11) in addition to the connectivity between inland rock (L161_IIC), and set aside grassland (L52_IIC). As such a mosaic of contrasting land cover types is important for EAG1 species when considering broad habitat classifications. Average edge contrast (ECON_MN), however, was negatively associated with EAG1 occurrence in the PH1-EAG1$_{comb}$ model, and this could be a reflection of the habitats within this model, which were similar in their composition, for example, four varieties of grassland.

Response to patch shape differed between EAGs when considering both the LCM and PH1 models. Average patch shape irregularity (SHAPE_MN) and patch elongation (CIRCLE_MN) were positive influences on the occurrence of all butterfly species in the PH1-ALL$_{comb}$ model and the occurrence of EAG1 species in the LCM-EAG1$_{comb}$ model. Increases in patch shape complexity are associated with increases in the amount of edge habitat due to a higher perimeter to area ratio, resulting in a higher edge contrast (Ewers and Didham, 2007), and range in edge contrast was positively associated with EAG1 species in the LCM-EAG1$_{comb}$ model as previously discussed.

In contrast to the PH1-ALL$_{comb}$ and LCM-EAG1$_{comb}$ models, negative associations with average patch shape weighted by area (SHAPE_AM) were detected for EAG3 species in both the LCM-EAG3 and PH1-EAG3 models. Large patches are more likely to be of a complex shape (Ewers and Didham, 2007), and as shape complexity increases and patches become less circular, the probability of occurrence of EAG3 species decreases. This is in accordance with findings by Yamaura *et al.*, (2008), who found that patches which are more circular in shape, were associated with increased abundance and species richness of butterflies associated with open habitats.

The landscape based models developed in this chapter are based on aggregated species data from the UKBMS, and the Warwickshire biological records centre. Due to the collection of data by a network of volunteers over several years, it is possible that errors may have occurred in the detection of species, which varies between observers and between years (Asher, *et al.*, 2001; Kery and Schmid, 2004; Lele, *et al.*, 2012). This may have resulted in the under-recording of cryptic species such as the hair streaks within woodland habitat and over recording of more common widespread species, such as the meadow brown, which are associated with more open habitats (Liley, *et al.*, 2004). The aggregation of species data within grid squares, however, will limit this bias that is likely to occur in different habitats. Aggregation within grid squares also accounts for differences in the spatial accuracy of data recordings as well as allowing for the consideration of a landscape context for modelling butterfly distribution, rather than site or habitat based approach. Despite the potential limitations within the butterfly data set it is evident that reliable presence-absence models can be built using this data and that ecologically sound relationships can be inferred from the models.

4.3.1 Conclusion

The landscape based models demonstrate that different combinations of landscape parameters are important across the four species groups. The accuracy and performance of the models differed between the four groups, with the EAG3 models comprising the most accurate discrimination. However, this is likely to be reflection of high proportion of absence squares for EAG3 species and as such the model is particularly good at predicting absence. Although, this maybe the case ecologically valid species-landscape associations can be drawn from the PH1-EAG3 model, in particular, when considering the ecological requirements of these species. The PH1-

ALL_{comb} and LCM-ALL_{comb} combined models exhibited the lowest discriminatory ability as measured by the AUC in comparison to the EAG models, however this difference was negligible. Furthermore several parameters of the all butterfly species models also occurred in the EAG models. These findings suggest that the PH1-ALL and LCM-ALL models are adequate for predicting the occurrence of butterfly species despite the lack of absence data used to develop these models. This is a positive finding considering the sheer lack of absence data in most biological records centres. The transferability of these models spatially and temporally however needs to be assessed to determine whether these models will still perform adequately despite the limited data used for their development.

It is evident that the LCM and PH1 models captured different relationships with butterfly occurrence across the four species groups, which are ecologically valid. Differences in detected relationships are likely to have arisen due to the different thematic resolutions of the two data sources. Most notably, the PH1 models comprised 19 different habitats across the four models, and the LCM comprised nine different habitats. This reflects the differences in the level of precision associated with the two data types, with broader habitats classified within the LCM 2000 incorporating several different PH1 habitats. Furthermore, very few significant correlations were observed between the corresponding habitats within each species model. The coarse classification of LCM 2000 is reflected in some surprising trends (such as EAG2 occurrence being negatively associated with woodland area), and often one LCM class covers multiple PH1 classes (Appendix A9), a likely reason for the higher number of PH1 habitats included in the models in comparison to the LCM models. Despite these differences, species-landscape associations detected by the LCM models are ecologically plausible. Furthermore, although the discriminating ability of the PH1 models were more accurate than the LCM equivalents, it is important to take into account the labour intensive field work needed to produce PH1 maps and in turn the limited availability of these maps on a UK or regional scale in comparison to the LCM 2000. Thus despite the relative poor performance of the LCM models in comparison to the PH1 models, the LCM models will be more readily transferable to a countrywide scale in comparison to the PH1 models. As such it is important to assess the transferability of these models temporally and spatially. Model validation is essential also to determine the quality and usability of

the models. Overall, the results from this chapter demonstrate that the presence-absence of all butterfly species and species grouped by their ecological attributes can be reliably predicted as function of the combined landscape based components; landscape composition, connectivity and structure.

Chapter 5: Model validation

5.1 Introduction

5.1.1 The importance of model validation and common approaches

Predicting species occurrence or habitat suitability in response to changes in abiotic or biotic drivers is paramount for targeting effective conservation, particularly over landscape scales, and is central to a growing body of literature (Guisan and Zimmermann, 2000; Lawler, *et al.*; Lawler, *et al.*, 2011). Species distribution maps have been widely used for predicting the distribution of several taxonomic groups, including butterfly species and plants (Cowley, *et al.*, 2000). Such predictions are often derived from habitat-association models or other empirical predictive models, for example, by the probability of presence from logistic regression models (Guisan and Zimmermann, 2000). Evaluating and understanding the performance of predictive models underpinning species distribution maps is essential to ensure accurate use and transferability to new data sets (Vaughan and Ormerod, 2005).

Model evaluation involves model validation and model accuracy. Model accuracy determines the discriminating ability of the model via its sensitivity and specificity (see section 2.2.6 and Chapter 4) (Dormann, *et al.*, 2012; Luoto, *et al.*, 2006), and model validation assesses the ability of the model to correctly predict responses which were not used during model development or calibration (Dormann, *et al.*, 2012). Model validation allows conclusions to be drawn on the efficacy and uncertainty of the model under different conditions, and identify potential problems and areas for further research (Fleishman, *et al.*, 2003; Vaughan and Ormerod, 2005).

A key element of model validation involves the use of ‘testing data’, which is data that has not been used during model development (‘training data’). Despite this the most common type of validation of species distribution or occurrence models involves cross–validation, where the observed data is partitioned into subsets comprising ‘training data’ and ‘testing data’ (Schröder, *et al.*, 2009). The testing data is omitted during model development and using just the training data the ability of the model to accurately predict the testing data is assessed (Luoto, *et al.*, 2006). Cross-validation has been applied in several studies aiming to predict species distribution, for example by Robinson *et al.*, (2014) for the assessment of Butterfly distribution models. There are limitations with this approach, for example the data used for testing is not completely independent of the data used for construction

(Luoto, *et al.*, 2006), and as such the performance of the model when transferred spatially or temporally is uncertain (Dormann, *et al.*, 2012; Vaughan and Ormerod, 2005).

5.1.2 Collection and quality of testing data

The quality of testing data influences the assessment of model efficacy and transferability (Vaughan and Ormerod, 2005) and thus testing data should ideally be temporally and spatially independent of the training data (Araujo, *et al.*, 2005; Dormann, *et al.*, 2012; Vaughan and Ormerod, 2005). Cross-validation is common as obtaining independent data is difficult due to time and resource constraints (Hirzel, *et al.*, 2006). To date there are very few studies which have obtained spatially and temporally independent data sets. An alternative approach is to satisfy at least one criterion, for example using temporally independent data but within the same region such as with Dorman *et al.*, (2012), though this approach is still infrequent. Assessment of the temporal transferability of the model in the same region from which the model is based is first required before the model can be successfully applied to different regions (Fleishman, *et al.*, 2003). In addition to independence, testing data should be representative of the conditions of the region for which the model is to be extrapolated (Vaughan and Ormerod, 2005).

5.1.3 Predictive modelling assumptions and limitations

A major assumption of empirical models predicting species distributions is that habitat classes, often mapped at a coarse scale, reflect habitat quality at local scales, such as vegetation composition or structure or the presence of nectar and larval host plants (Guisan and Zimmermann, 2000). Butterfly species require different but complementary resources at each stage of their life cycle, and as such occupy several different vegetation types (Dennis, 2010). Habitat quality is therefore an important factor in determining the distribution and persistence of butterfly species with strong preferences for local habitat characteristics, particularly for breeding, foraging and roosting (Dennis, 2010; Flick, *et al.*, 2012; Ouin, *et al.*, 2004; van Halder, *et al.*, 2008). Furthermore, the compositional and configurational heterogeneity of the landscape was found to positively affect butterfly species diversity in agricultural landscapes through the provision of a variety of habitat types and associated resources required during the life cycle of several different species (Dennis, 2010; Flick, *et al.*, 2012).

Land cover or habitat classes derived at a coarse landscape scale may not necessarily capture species specific habitat and resource requirements (Dennis, 2010; Schweiger, *et al.*, 2006). As a consequence several studies advocate that predictive models should incorporate both landscape scale and local characteristics in order to accurately predict species distribution (van Halder, *et al.*, 2008). For example Schweiger *et al.*, (2006) found that both local habitat quality and regional characteristics were important in predicting the distribution of the Speckled Wood butterfly (*Pararge aegeria*). In particular, local habitat quality was of major importance when the number of woodland patches in the landscape was low. In contrast, Cowley *et al.*, (2000) found that with the exception of widely distributed species such as the European Peacock (*Inachis io*), habitat-association models were better at predicting species distribution than models based on host-plant distributions.

It is evident from these contrasting studies that determining the importance of habitat quality over habitat composition based models is dependent on the local and regional characteristics measured as well as the species modelled. Relating local habitat quality and characteristics to coarse scale habitat and land cover data is therefore a vital stage of validating predictive models to establish if habitat-association models are capturing local habitat characteristics (Schweiger, *et al.*, 2006). If relationships between observed butterfly communities and habitat quality can be validated, than butterfly presence can be used as an indication of the floral composition and vegetation structure within particular habitats.

5.1.4 Aims

The aims of the work reported in this chapter are to: -

1. Validate the predictive performance and accuracy of the combined landscape based models developed in Chapter 4 using temporally independent landscape data.
2. Assess the ability of the landscape based models developed in Chapter 4 to predict butterfly community measurements, including the abundance, richness, diversity and species composition of all butterfly species and species grouped by their ecological attributes, in addition to local habitat characteristics (hence biodiversity).

These aims will address the hypothesis 'Landscape based models developed in Chapter 4 can be used to predict the presence-absence and community assemblage of butterfly species from a temporally independent landscape data set'.

5.2 Results: Warwickshire 2000-2009

The PH1 combined models developed in Chapter 4 (section 4.2.9) based on landscape metrics derived from the PH1 2000 were validated using temporally independent PH1 habitat data (PH1 2010). The LCM combined models (section 4.2.5) derived from the LCM 2000 were validated considering the temporally different butterfly data (2000-2009) only.

The 38 variables from the PH1 combined models were significantly correlated between the two time periods, most notably strong correlations (S >0.75) occurred between 20 of the 21 compositional variables. The weakest correlation was observed for the area of mixed semi-natural woodland (PH-5) (S = 0.663, p<0.001). All four connectivity variables were significantly correlated between the two time periods, with the weakest correlation occurring between the connectivity of inundation vegetation (P28_IIC) (S = 0.690, p<0.001). The 12 structural metrics were also significantly correlated with weakest correlation occurring between five metrics; CONNECT (S = 0.618, p<0.001); ENN_SD (S = 0.647, p<0.001); CONTIG_SD (S = 0.568, p<0.001); ECON_MN (S = 0.669, p<0.001); and PROX_CV (S = 0.676, p<0.001).

5.2.1 Phase 1 predictive model: Observed results 2000-2009 Warwickshire dataset compared to predicted results

Comparing the 2000-2009 Warwickshire butterfly dataset to the predicted values obtained from the Phase 1 habitat map 2010 using the Phase 1 combined models demonstrates the specificity of the Phase 1 all butterfly species model (PH1-ALL_{comb}) to predict absence of all butterflies to be 63.2 % (Table 5.1a). The sensitivity for prediction of presence of all butterflies was 67.5 % (Table 5.1a), and as such grid squares with high predictions adequately match the occurrence of 2000-2009 butterfly species (Figure 5.1a,b). Furthermore, the ROC area under the curve (AUC) of the model indicates it to be a ‘fair’ performing model, for predicting the observed data irrespective of a threshold for defining presence (Table 5.1a), as classified by Araujo, *et al.*, (2005).

For specific Ecological Attribute Group (EAG) models the specificities and sensitivities varied (Table 5.1a; Figure 5.1c-f), in particular for the EAG3 species model (PH1-$EAG3_{comb}$) the specificity was 88.7 %, however the sensitivity (28.6 %)

was markedly lower than for PH1-ALL_{comb} or any other EAG model. The PH1-$EAG1_{comb}$ and PH1-$EAG2_{comb}$ models were found to adequately match the presence and absence of the observed data (Figures 5.1c-f) with high specificity and sensitivity (Table 5.1a). Furthermore, the AUC values obtained for the two models indicate the model performances to be 'fair' (PH1-$EAG1_{comb}$ AUC = 0.766, p<0.001; PH1-$EAG2_{comb}$ AUC =0.758, p <0.001) (Table 5.1a).

The AUC, specificity and sensitivity values for the models applied to the PH1 2010 landscape data set were similar to those derived from the PH1 2000 landscape data set for the PH1-ALL_{comb}, PH1-$EAG1_{comb}$ and PH1-$EAG2_{comb}$ models (Table 5.1a). However, for the PH1-$EAG3_{comb}$ model sensitivity was 63.7 % lower than that derived for the PH1 2000 data set. Furthermore, the PH1-$EAG3_{comb}$ model is considered to 'fail' at predicting species occurrence with a non-significant AUC (AUC = 0.584, p = 0.175), which is considerably lower than that obtained from the PH1 2000 data set (Table 5.1a).

The predictions obtained using the PH1 2010 landscape data are highly correlated with those obtained from the PH1 2000 landscape data, with significant positive correlations between the predictions from the two data sets for PH1-ALL_{comb} (r = 0.897, p<0.001), PH1-$EAG1_{comb}$ (r = 0.822, p <0.001) and PH1-$EAG2_{comb}$ (r = 0.918, p<0.001).

5.2.2 LCM predictive model: Observed results 2000-2009 Warwickshire dataset compared to predicted results

Comparing the 2000-2009 Warwickshire butterfly dataset to the predicted values obtained from the LCM 2000 combined model demonstrates the specificity of the LCM all butterfly species model (LCM-ALL_{comb}) to accurately predict absence of all butterflies (93.7 %), however the sensitivity for prediction of presence was much lower with only 16.6 % correct predictions (Table 5.1b). In contrast, the threshold independent measure of discrimination (AUC) indicates a slight increase in the ability of the model to predict the observed data (Table 5.1b).

For specific Ecological Attribute Groups (EAGs) the specificities and sensitivities were similar to each other, for example for EAG2 the specificity was 65.6 % and the sensitivity was 55.7 %. The sensitivity of LCM-$EAG1_{comb}$ and LCM-$EAG2_{comb}$ models were much higher in comparison to the LCM-ALL_{comb} model (Table 5.1b).

However, for LCM-EAG3$_{comb}$, the specificity was 84.9 %, and the sensitivity was 19.5 %, a similar pattern to the LCM-ALL$_{comb}$ model. With exception to the EAG3 model specificity and sensitivity values for 2010 were similar to those derived from the 2000 data set (Table 5.1b). Furthermore there was little difference in the AUC between the two time periods for both the EAG1 and EAG2 models (Table 5.1b). The sensitivity of the EAG3 model had reduced by 57.4 % when comparing predictions for the year 2010 to 2000. Additionally the AUC for this model had reduced considerably and is no longer significantly different to that obtained by chance (Table 5.1b).

(a)	PH1 Combined Model					
	1990-1999			2000-2009		
Model type	Specificity % (Absence)	Sensitivity % (Presence)	AUC	Specificity % (Absence)	Sensitivity % (Presence)	AUC
PH1-ALL	68.4	60.9	0.720	63.2	67.5	0.704
PH1-EAG1	63.9	75.6	0.806	71.4	66.5	0.766
PH1-EAG2	73.8	59.4	0.752	71.9	60.4	0.758
PH1-EAG3	89.0	92.3	0.963	88.7	28.6	0.584

(b)	LCM Combined Model					
	1990-1999			2000-2009		
Model type	Specificity % (Absence)	Sensitivity % (Presence)	AUC	Specificity % (Absence)	Sensitivity % (Presence)	AUC
LCM-ALL	61.4	61.2	0.671	93.7	16.6	0.692
LCM-EAG1	70.9	64.8	0.760	54.4	75.8	0.720
LCM-EAG2	63.7	64.7	0.698	65.6	55.7	0.683
LCM-EAG3	78.7	76.9	0.823	84.9	19.5	0.566

Table 5.1: Model accuracy of the PH1 combined models (a) and the LCM combined models (b), as determined by the ROC area under the curve (AUC) and the percentage correct predictions (confusion matrix) of the observed butterfly presence-absence data for 2000-2009 compared to that obtained for predicting the observed data for 1990-1999. Model predictions for the 2000-2009 butterfly data set from the PH1 combined models are derived from the PH1 2010 habitat map. LCM combined models are derived from LCM 2000 in both instances. The AUC, specificity and sensitivity is provided for the 'all butterflies' (ALL), EAG1, EAG2 and EAG3 models. Significance values: light grey <0.05, medium grey <0.05, and dark grey <0.001.

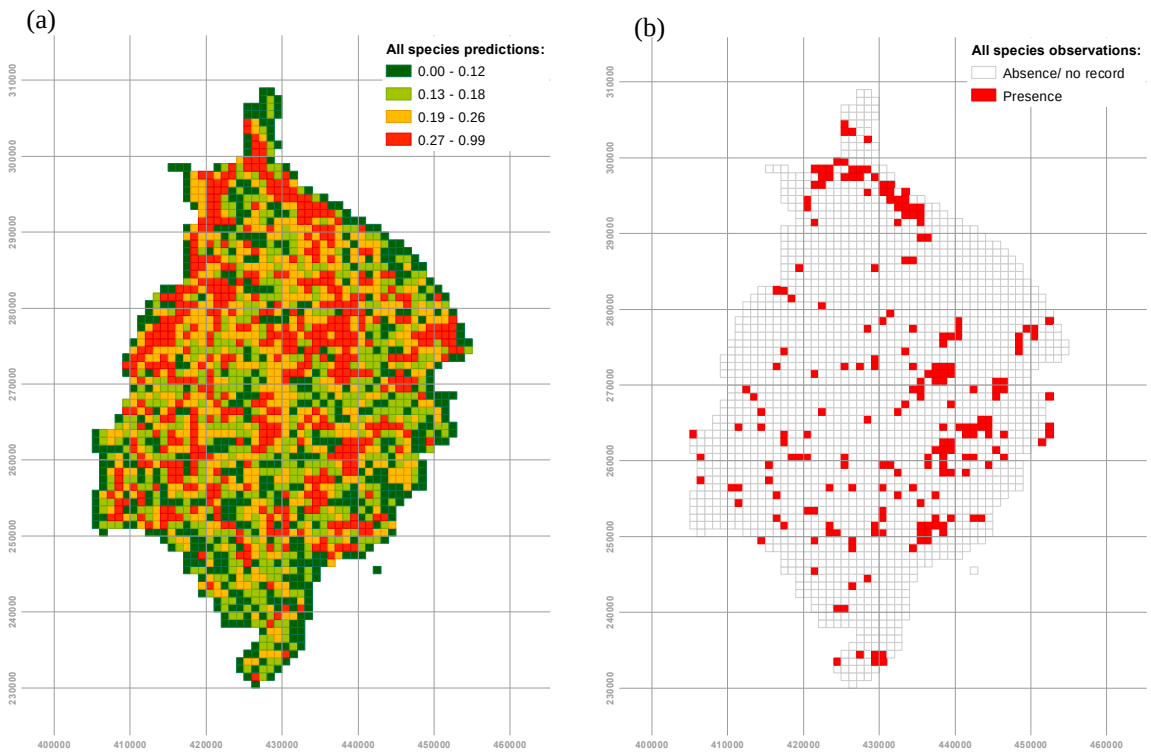


Figure 5.1: *(cont.)*

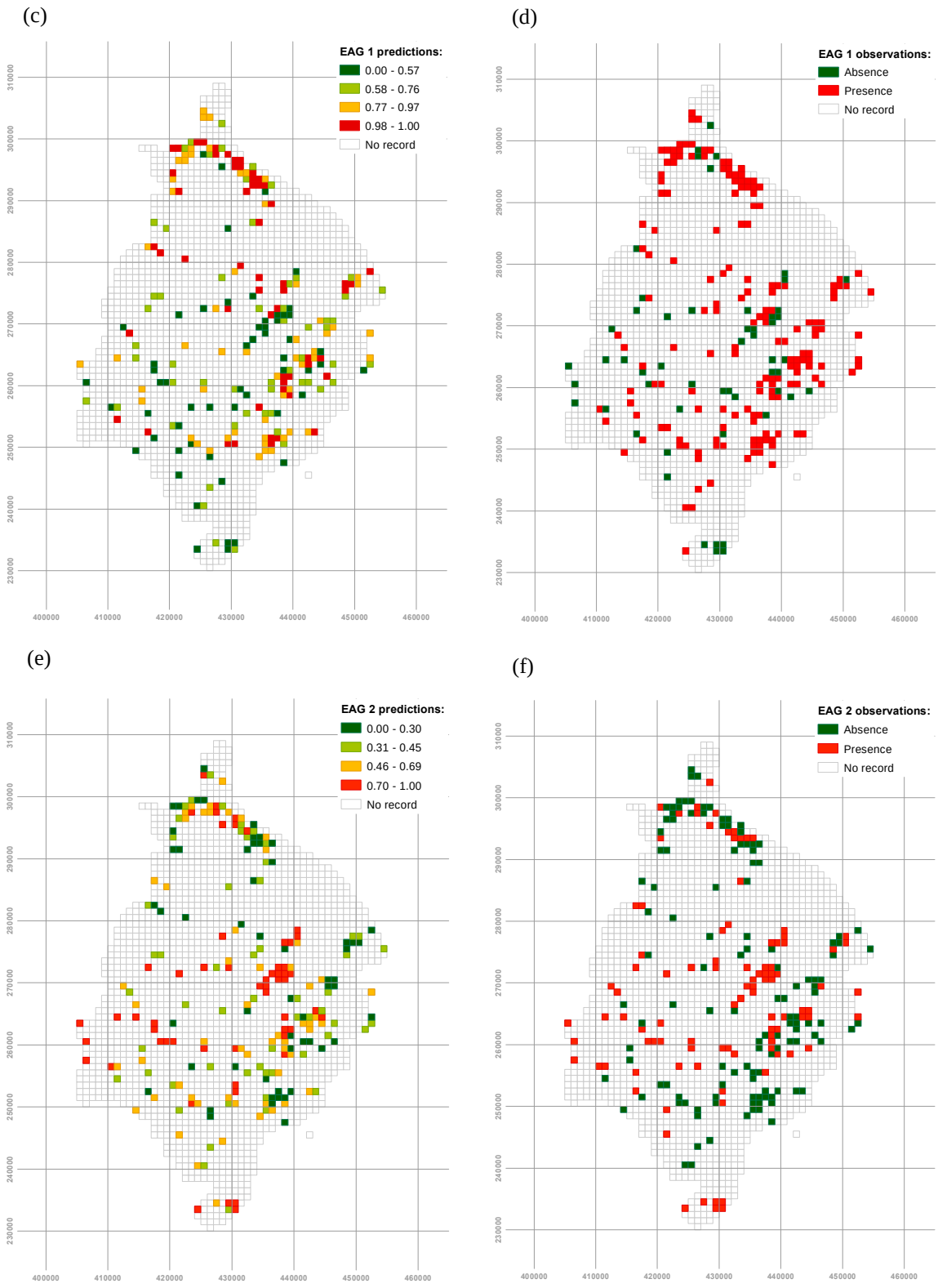


Figure 5.1: The predicted values from the PH1-ALL (a) PH1-EAG1 (c) and PH1-EAG2 (e) models based on the PH1 habitat map 2010 compared to the observed occurrence of Warwickshire butterflies 2000-2009 for all butterfly species (b) and the presence and inferred absence of EAG2 species (d) and EAG3 species (f). The quartile ranges for the model predicted values (a,c,e) are provided.

5.2.3 Phase 1 predictive model: relationships between combined models and butterfly community measurements

When comparing the predicted values from the PH1 2010 combined model for all butterfly species (PH1-ALL_{comb}) with the butterfly community measurements of standardised abundance, diversity (reciprocal and Simpsons) and species richness for the Warwickshire data set 2000 – 2009, significant positive correlations were observed, however these coefficients were weak (Table 5.2). The predicted values obtained from the PH1-$EAG1_{comb}$ and PH1-$EAG2_{comb}$ models significantly positively correlated with the standardised abundance, richness, and diversity of butterfly species within the corresponding EAG. For EAG1 strongest correlations were observed between predicted values and species richness (r = 0.355, p <0.001), and reciprocal diversity index (r = 0.355, p <0.001). For EAG2 strongest correlations were observed between predicted values and species richness (r = 0.385, p <0.001), and standardised abundance (r = 0.377, p <0.001) (Table 5.2).

Butterfly community measurements	PH1 2010 combined models		
	All*	EAG1	EAG2
Standardised abundance	0.215	0.296	0.377
Species richness	0.215	0.355	0.385
Simpsons Diversity Index	0.188	0.276	0.168
Reciprocal diversity	0.221	0.355	0.307

Table 5.2: The relationship between the predicted values from the PH1 2010 combined models and the butterfly community measurements of all butterfly species (All) and EAG1 and EAG2 species obtained for 2000-2009. Pearson Product Moment Correlation Coefficients for predicted values from the PH1 2010 combined models for EAG1 and EAG2 with the standardised abundance, species richness and diversity (Simpsons and Reciprocal) are shown. * Spearman's rank correlation coefficients are provided. Significance values: light grey <0.05, and medium grey <0.01.

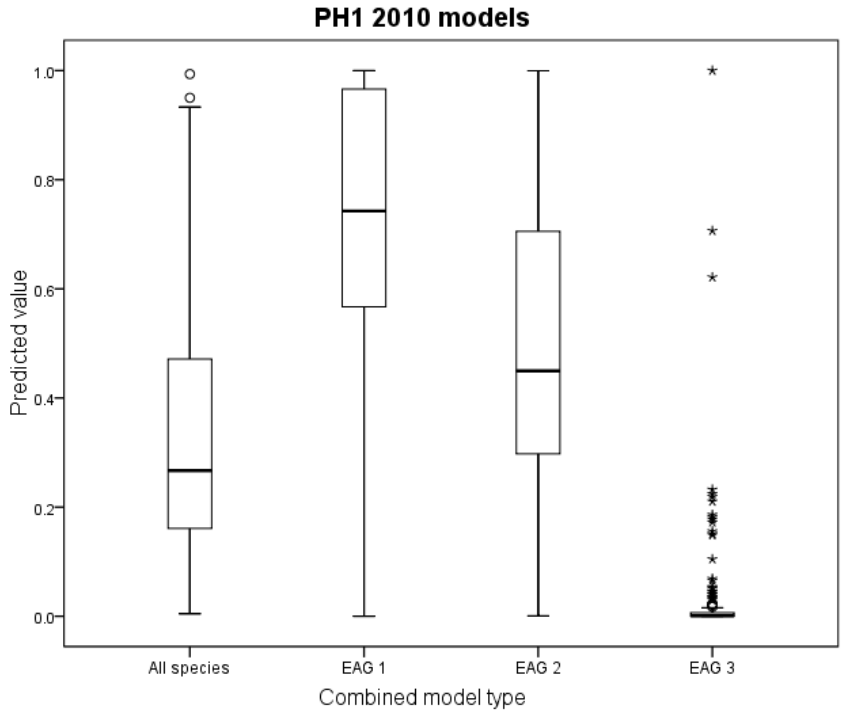


Figure 5.2: The distribution of predicted values using the PH1 2010 habitat map to predict the occurrence of all butterfly species (*n* = 2073), EAG 1 (*n* = 224), EAG 2 (*n* = 224) and EAG3 (*n* = 224) species across Warwickshire. For each model (with exception to EAG 3), four suitability groups were derived from the quartiles represented by the lower, middle and upper bars of the boxplots.

The predicted values obtained using the PH1 2010 habitat map for the all butterfly species model (PH1-ALL$_{comb}$), PH1-EAG1$_{comb}$, PH1-EAG2$_{comb}$ and PH1-EAG3$_{comb}$ models, provide an indication of the suitability of grid squares for supporting the corresponding butterfly community. Based on the quartiles of the predicted values for each model, four habitat suitability groups were identified, with group 1 corresponding with low suitability and group 4 corresponding with high suitability for supporting butterflies (Figure 5.2). When considering the differences in the butterfly community measurements between these four groups, the median standardised abundance across all butterfly species was found to differ significantly (H_3 = 110.296, p = 0.000), as did species richness (H_3 = 111.121, p = 0.000), species diversity (Simpsons) (H_3 = 81.886, p = 0.000) and the reciprocal diversity index (H_3 = 112.369, p = 0.000). Tukey HSD comparisons revealed that in all cases group 4

(high suitability) supported significantly higher abundance, species richness and diversity than all other groups ($p<0.05$) (Figure 5.3a,c,e).

The butterfly community measurements of EAG1 species was found to differ between the four suitability groups (Figure 5.4a,c,e), with significant differences in the average standardised abundance ($F_{3,220} = 7.670$, $p<0.001$), species diversity ($F_{3,220} = 4.937$, $p=0.002$), reciprocal species diversity ($F_{3,220} = 10.665$, $p<0.001$), and species richness ($F_{3,220} = 11.173$, $p<0.001$). Tukey HSD comparisons indicated that average standardised abundance, species richness, and diversity (Simpsons) was significantly higher in group 4 (high suitability) in comparison to group 1 and group 2 (low suitability) ($p<0.05$). Furthermore, average species richness and reciprocal diversity was significantly higher in group 3 (medium to high suitability) in comparison to group 1 ($p<0.005$).

The suitability groups derived from the distribution of EAG2 butterfly species significantly differed in terms of their average standardised abundance ($F_{3,220} = 16.182$, $p<0.001$), species richness ($F_{3,220} = 17.344$, $p<0.001$), and reciprocal diversity index ($F_{3,220} = 11.058$, $p<0.001$). In all cases group 4 (high suitability) significantly differed from all other groups ($p<0.05$), with higher standardised abundance, species richness, and diversity (reciprocal) (Figure5.5a,c,e). Significant difference was also observed between the average Simpsons Diversity Index of the four groups ($F_{3,220} = 3.314$, $p=0.021$) with significantly higher diversity in group 4 in comparison to group 2 ($p<0.05$).

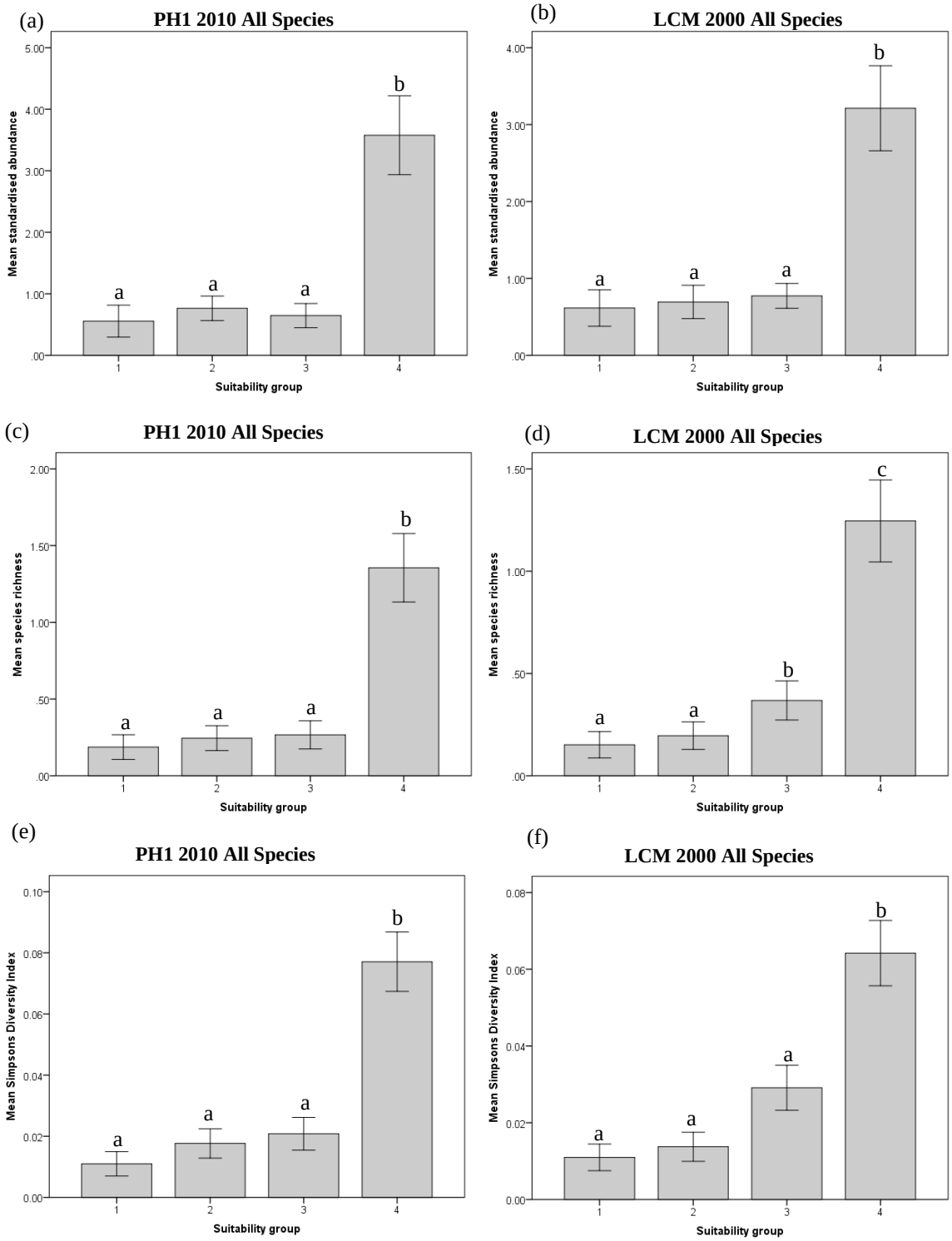


Figure 5.3: Butterfly community measurements for all butterfly species per habitat suitability group for the PH1 2010 habitat map (a, c, e) and the LCM 2000 map (b, d, f). Mean (±SE) standardised abundance for all butterfly species per suitability group for (a) PH1 2010 and (b) LCM 2000; species richness for (c) PH1 2010 and (d) LCM 2000; and Simpsons Diversity Index for (e) PH1 2010 and (f) LCM 2000. Bars that do not share a letter have significantly different means (Tukey test, $p<0.05$). Suitability groups are derived from the quartiles of the predicted values for each PH1 and LCM model (Figure 5.2; Figure 5.6).

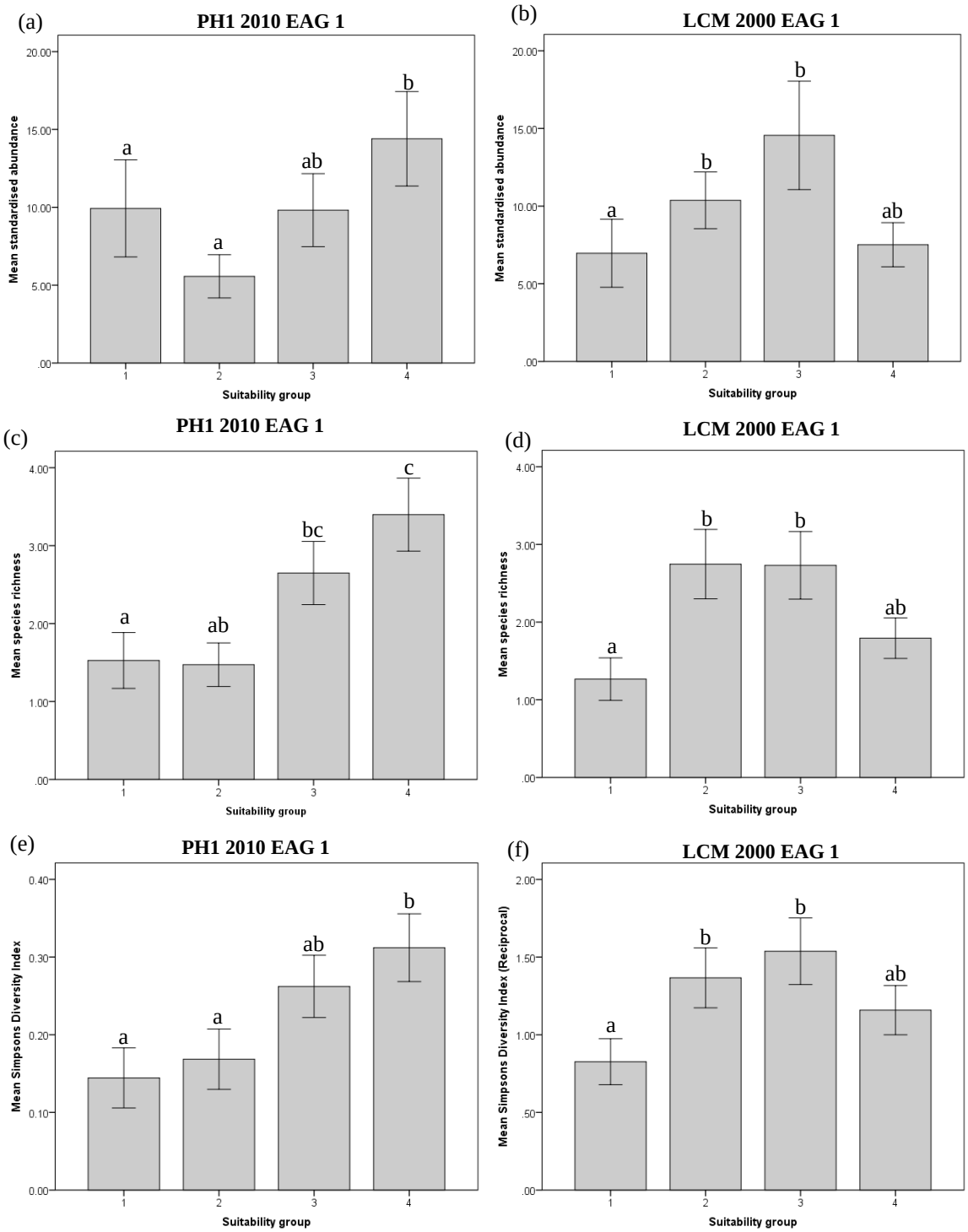


Figure 5.4: Butterfly community measurements for EAG1 species per habitat suitability group for the PH1 2010 habitat map (a, c, e) and the LCM 2000 map (b, d, f). Mean (±SE) standardised abundance for all butterfly species per suitability group for (a) PH1 2010 and (b) LCM 2000; species richness for (c) PH1 2010 and (d) LCM 2000; Simpsons Diversity Index for (e) PH1 2010; and Reciprocal Diversity Index for (f) LCM 2000. Bars that do not share a letter have significantly different means (Tukey test, $p<0.05$). Suitability groups are derived from the quartiles of the predicted values for each PH1 and LCM model (Figure 5.2; Figure 5.6).

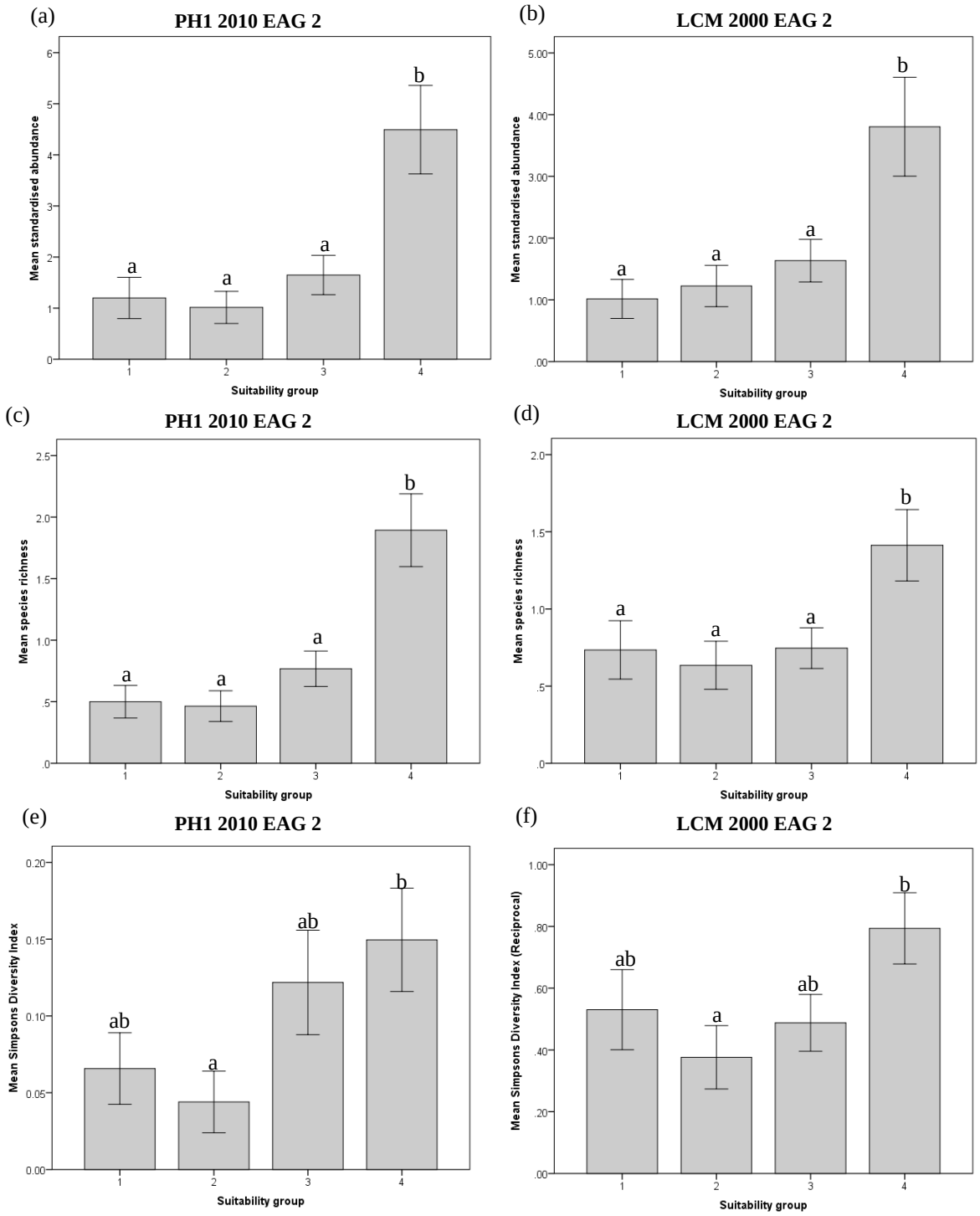


Figure 5.5: Butterfly community measurements for EAG2 species per habitat suitability group for the PH1 2010 habitat map (a, c, e) and the LCM 2000 map (b, d, f). Mean (±SE) standardised abundance for all butterfly species per suitability group for (a) PH1 2010 and (b) LCM 2000; species richness for (c) PH1 2010 and (d) LCM 2000; Simpsons Diversity Index for (e) PH1 2010; and Reciprocal Diversity Index for (f) LCM 2000. Bars that do not share a letter have significantly different means (Tukey test, $p<0.05$). Suitability groups are derived from the quartiles of the predicted values for each PH1 and LCM model (Figure 5.2; Figure 5.6).

5.2.4 LCM 2000 predictive model: relationships between combined models and butterfly community measurements

When comparing the predicted values from the LCM 2000 combined model for all butterfly species (LCM-ALL_{comb}) with the butterfly community measurements of standardised abundance, diversity and species richness for the Warwickshire data set 2000 – 2009, weak positive correlations were observed (Table 5.3). The predicted values obtained from the EAG models were also weakly positively correlated with the butterfly community measurements with some significant correlations observed for EAG1 and EAG2 (Table 5.3). In particular, the strongest correlation was obtained between the predicted values from the LCM-$EAG2_{comb}$ model and standardised abundance of EAG2 species (r = 0.336, p <0.001).

The predicted values obtained from the LCM 2000 map for the all butterfly species model (LCM-ALL_{comb}), LCM-$EAG1_{comb}$, LCM-$EAG2_{comb}$ and LCM-$EAG3_{comb}$ models provide an indication of the suitability of grid squares for supporting the corresponding butterfly community. Based on the quartiles of the predicted values for each model, four habitat suitability groups were identified, with group 1 corresponding with low suitability and group 4 corresponding with high suitability for supporting butterflies (Figure 5.6).

Butterfly community measurements for 'all butterfly' species were found to differ between these suitability groups with significant differences in median standardised abundance (H_3 = 99.060, p<0.001), species richness (H_3 = 99.122, p<0.001), Simpsons Diversity Index (H_3 = 56.362, p<0.001), and Reciprocal Diversity Index (H_3 = 87.242, p<0.001). Group 4 (high suitability) supported significantly higher abundance, species richness, and diversity (Simpsons, and reciprocal) in comparison to all other groups (p<0.05) (Figure 5.3b,d,f). Additionally, group 3 comprised significantly higher median species richness in comparison to group 1 (p<0.05) (Figure 5.3b,d).

Butterfly community measurements	LCM 2000		
	All*	EAG1	EAG2
Standardised abundance	0.202	0.123	0.336
Species richness	0.203	0.138	0.259
Simpsons Diversity Index	0.156	0.051	0.06
Reciprocal diversity	0.183	0.11	0.18

Table 5.3: The relationship between the predicted values from the LCM 2000 combined models and the butterfly community measurements of all butterfly species (All) and EAG1 and EAG2 species obtained for 2000-2009. Pearson Product Moment Correlation Coefficients for predicted values from the LCM 2000 combined models for EAG1 and EAG2 with the standardised abundance, species richness and diversity (Simpsons and Reciprocal) are shown. * Spearman's rank correlation coefficients are provided. Significance values: light grey <0.05, and medium grey <0.01.

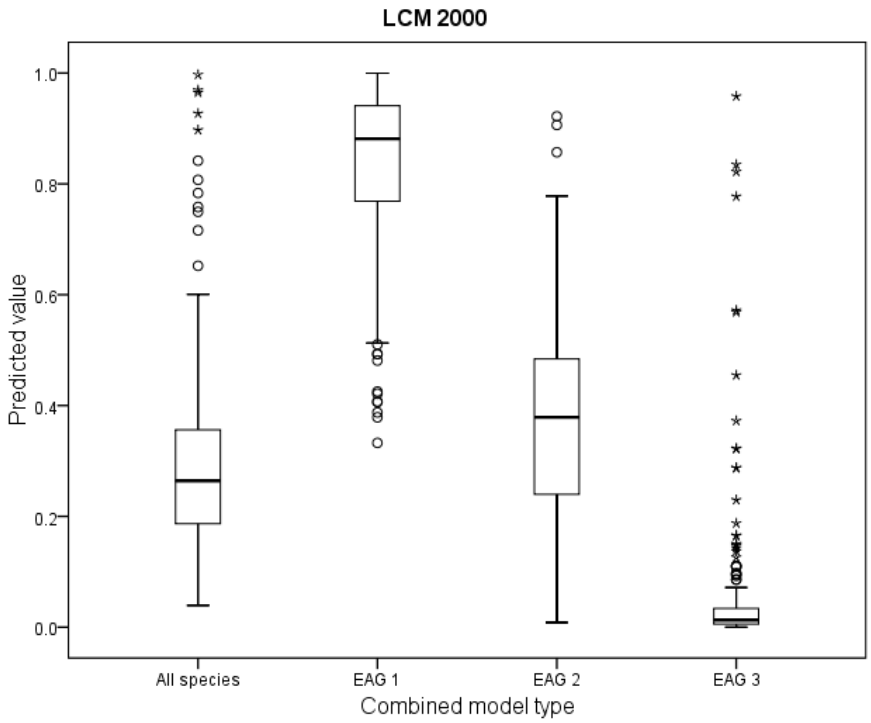


Figure 5.6: The distribution of predicted values using the LCM 2000 map to predict the occurrence of all butterfly species (*n* = 2427), EAG 1 (*n* = 253), EAG 2 (*n* = 253) and EAG3 (*n* = 253) species across Warwickshire. For each model (with exception to EAG 3), four suitability groups were derived from the quartiles represented by the lower, middle and upper bars of the boxplots.

The butterfly community of EAG1 differed significantly between the four habitat suitability groups in terms of average standardised abundance ($F_{3,249}$ = 3.602, p=0.014), species richness ($F_{3,249}$ = 4.393, p=0.005), and reciprocal diversity index ($F_{3,249}$ = 3.058, p=0.029). In contrast to the LCM-ALL_{comb} model, group 4 comprised a similar butterfly community to group 1 (p>0.05) (Figure 5.4b,d,f). However, group 1 differed significantly in comparison to groups 2 and 3, with lower standardised abundance and species richness (p<0.05) (Figure 5.4b,d). Significant differences in the average reciprocal diversity index were only observed between groups 1 and 3, with group 3 supporting a higher diversity (p<0.05) (Figure 5.4f). Simpsons Diversity Index was found to not significantly differ between groups ($F_{3,249}$ = 1.138, p=0.334).

Differences in the butterfly community of EAG2 species between suitability groups are similar to that observed for the LCM-ALL_{comb} model. Between these four groups significant differences were observed in the average standardised abundance ($F_{3,249}$ = 8.818, p<0.001), species richness ($F_{3,249}$ = 6.609, p<0.001), and reciprocal diversity index ($F_{3,249}$ = 4.200, p=0.006). Group 4 was found to support significantly higher standardised abundance, and species richness in comparison to all other groups (p<0.05) (Figure 5.5b,d). The reciprocal diversity Index only differed significantly between group 2 and 4 (p<0.05) (Figure 5.5f). The average Simpsons Diversity Index was found not to differ significantly between suitability groups ($F_{3,249}$ = 1.192, p=0.313).

5.2.5 Comparison of predictions between PH1 2010 and LCM 2000

The predictions obtained from the PH1 2010 and LCM 2000 models for all butterfly species (PH1-ALL_{comb} and LCM-ALL_{comb}) are both weakly correlated with the butterfly community measurements of the Warwickshire 2000-2009 data set. However, when grid squares are grouped by their habitat suitability obtained from these predictions, the PH1 2010 and LCM 2000 models perform similarly with significant differences observed between group 4 and group 1 in terms of median standardised abundance, species richness and diversity (Simpsons and reciprocal) for all butterfly species (Figure 5.3).

The most striking difference between predictions obtained from the PH1 and LCM models, is the predictions obtained for the EAG1 butterfly community. Correlations

between predicted values and butterfly community measurements are much weaker and non-significant for the LCM-EAG1$_{comb}$ predictions in comparison to the PH1-EAG1$_{comb}$ predictions (Table 5.3; Table 5.2). Furthermore, for the PH1-EAG1$_{comb}$ model, habitat suitability group 4 was found to support significantly higher standardised abundance, species richness and diversity of EAG1 species in comparison to the lower suitability groups (groups 1 and 2) (Figure 5.4a,c,e). In contrast the butterfly community measurements of EAG1 species associated with grid squares with high suitability (group 4) based on the LCM-EAG1$_{comb}$ model is similar to that observed for group 1 (low suitability). Groups 2 and 3 with medium suitability supported significantly higher standardised abundance, species richness and diversity than group1 (Figure 5.4b,d,f).

Correlations between the predictions obtained from the PH1-EAG2$_{comb}$ and LCM-EAG2$_{comb}$ models and the butterfly community measurements of EAG2 species are similarly low (Table 5.2; Table 5.3). Furthermore, predictions from both these models indicate significant differences between the high suitability group (group 4) and all other groups in their associated standardised abundance and species richness of EAG2 species (Figure 5.5a-d). Differences between the predictions from the two model data types can be seen when detecting differences in the diversity of EAG2 species. The correlations between the Simpsons Diversity Index and predictions from the PH1-EAG2$_{comb}$ model are significant (although very weak), and significant differences in diversity are observed between group 4 and all other groups (Table 5.2; Figure 5.5e). In contrast LCM-EAG1$_{comb}$ predictions do not significantly correlate with the Simpsons Diversity Index and as such no significant differences are observed between suitability groups in the average Simpsons diversity index of EAG2 species (Table 5.3).

5.3 Results: Sample sites

5.3.1 Sample site characteristics

A total of nineteen sites (1 km grid squares) were visited eight times over the spring and summer of 2013 (Figure 2.4; Section 2.3.1). The selected sites ranged in their composition, connectivity and structural characteristics across the 37 variables included within the PH1 combined models (derived from PH1 2010 data) and the 26 variables in the LCM combined models (see sections 4.2.5 and 4.2.9). When considering the PH1 2010 data, the sample sites comprised a total of 14 habitats with arable land (PH-34) the most dominant habitat per site (33 ha), occurring within 95 % of the sample sites. Semi-improved neutral grassland (PH-16) and broad-leaved semi-natural woodland (PH-1) were also dominant habitats (5.2 ha and 4.9 ha/ per site respectively), occurring within 84 % of the sites (Table 5.4). However, despite the widespread occurrence of these habitats their area ranged greatly, in particular the coverage of arable land ranged from 0 ha to 95 ha across the 19 sites. The sites also ranged considerably in terms of the connectivity of woodlands/ hedgerows (P1H_IIC) and recently felled woodland (P11_IIC) in addition to 11 structural metrics which were important across the four PH1 models. In particular, the sites varied in the range in patch extent (GYRATE_AM), and the variability in the aggregation of patch types (PROX_CV; SIMI_CV and ENN_SD) (Table 5.4).

5.3.2 Butterfly species abundance, richness and diversity

A total of 10,390 individuals of 22 species were observed across all the sites, with a maximum of 19 species observed within a single site and an average of 14.36 (SE ± 0.63) species per site. Accounting for differences in transect length, the total standardised abundance was 433.78 individuals per 100 m. Species comprising EAG4 were the most abundant across all sites comprising 54 % of the total abundance and occurring within every site (Table 5.5). This was closely followed by EAG1 species comprising 44 % of the total abundance and also occurring within every site. In particular, four species accounted for 75 % of the total abundance: Large white (*Pieris brassicae*) (22 %), Meadow brown (*Maniola jurtina*) (21 %), Small white (*Pieris rapae*) (20 %) and Ringlet (*Aphantopus hyperantus*) (12 %) (Table 5.4). Species comprising EAG2 and EAG3 were observed in low numbers comprising only 0.6 % and 1.3 % respectively of the total abundance. However,

EAG2 species occurred within 17 of the 19 sites, whereas EAG3 species occurred within only 11 sites. Migrant species comprised only 0.24 % of the total abundance.

Model parameter	Abbreviation	Model	Mean (ha)	±SE	Prop (%)	Min (ha)	Max (ha)
Landscape composition							
Broad-leaved semi-natural woodland (ha)	PH-1	ALL EAG2	4.891	1.296	84	0.000	20.188
Mixed semi-natural woodland (ha)	PH-5	ALL	0.082	0.072	11	0.000	1.375
Mixed plantation woodland (ha)	PH-6	EAG2	1.329	1.035	32	0.000	19.813
Dense/continuous scrub (ha)	PH-7	ALL	0.697	0.189	68	0.000	3.000
Recently felled woodland (ha)	PH-11	ALL EAG1	0.003	0.003	5	0.000	0.063
Orchard (commercial) (ha)	PH-12	EAG1	0.109	0.070	16	0.000	1.250
Unimproved acidic grassland (ha)	PH-13	EAG1	0.000	0.000	0	0.000	0.000
Semi-improved acidic grassland (ha)	PH-14	EAG1	0.003	0.003	5	0.000	0.063
Unimproved neutral grassland (ha)	PH-15	EAG1	0.003	0.003	5	0.000	0.063
Semi-improved neutral grassland (ha)	PH-16	ALL EAG1	5.197	1.799	84	0.000	34.688
Unimproved calcareous grassland (ha)	PH-17	EAG2	0.000	0.000	0	0.000	0.000
Continuous bracken (ha)	PH-22	EAG2	0.740	0.740	5	0.000	14.063
Inundation vegetation (ha)	PH-28	ALL	0.000	0.000	0	0.000	0.000
Standing water (ha)	PH-29	ALL EAG2	4.678	1.685	79	0.000	24.000
Quarry (ha)	PH-31	ALL EAG1	1.197	1.197	5	0.000	22.750
Arable (ha)	PH-34	ALL	33.000	6.289	95	0.000	94.625
Introduced shrub (ha)	PH-39	EAG1	0.046	0.043	11	0.000	0.813
Landscape Simpsons Diversity Index	LSIDI	ALL	0.636	0.048	-	0.103	0.842
Landscape heterogeneity	NLAND	EAG1	12.842	0.821	-	5.000	18.000

Table 5.4 *(cont.)*

Model parameter	Abbreviation	Model	Mean (ha)	±SE	Prop (%)	Min (ha)	Max (ha)
Connectivity metrics							
Woodland and hedgerow connectivity	P1H_IIC	ALL EAG2	0.485	0.033	100	0.278	0.889
Connectivity of recently felled woodland	P11_IIC	ALL	0.041	0.041	5	0.000	0.786
Connectivity of unimproved acidic grassland	P13_IIC	ALL	0.000	0.000	0	0.000	0.000
Connectivity of inundation vegetation	P28_IIC	ALL	0.000	0.000	0	0.000	0.000
Landscape structure metrics							
Mean patch area (ha)	AREA_MN	ALL	2.460	0.323	-	1.235	6.667
Range in patch extent (ha)	GYRATE_RA	ALL	294.210	9.841	-	191.775	368.903
Mean patch shape index	SHAPE_MN	ALL	1.274	0.016	-	1.126	1.422
Standard deviation of shape index	SHAPE_SD	EAG1	0.417	0.030	-	0.233	0.720
Area-weighted mean proximity index	PROX_AM	ALL	9.837	2.682	-	0.028	42.773
Coefficient of variation of proximity index (%)	PROX_CV	EAG1	284.026	16.720	-	141.889	424.241
Coefficient of variation of Similarity Index (%)	SIMI_CV	EAG2	173.265	4.319	-	146.212	210.707
Standard deviation of Euclidean nearest neighbour (m)	ENN_SD	ALL	147.645	13.038	-	90.535	309.155
Mean edge contrast (%)	ECON_MN	EAG1	83.450	1.011	-	77.004	89.929
Standard deviation of contiguity Index	CONTIG_SD	EAG1	0.292	0.006	-	0.251	0.350
Connect Index (%)	CONNECT	ALL	39.839	1.485	-	29.268	52.941

Table 5.4: Characteristics across the 19 sample sites derived from the PH1 2010 habitat map in terms of the parameters of the Phase 1 models for all species (PH1-ALL_{comb}), EAG1 species (PH1-$EAG1_{comb}$) and EAG2 species (PH1-$EAG2_{comb}$). The connectivity metrics refer to the Integral Index of Connectivity (IIC) (see Section 2.2.4). Metrics are unit less unless otherwise stated. See Appendix A4 and A8 for structural metric definitions.

Species	EAG	Total Abundance	Standardised abundance	Occurrence (sites)
Apatura iris	2	1	0.04	1
Pyrgus malvae	1	2	0.08	1
Neozephyrus quercus	2	2	0.08	2
Vanessa cardui	0	4	0.16	2
Colias croceus	0	6	0.26	2
Vanessa atalanta	0	15	0.62	10
Lycaena phlaeas	3	50	2.10	7
Melanargia galathea	1	51	2.12	8
Gonepteryx rhamni	2	63	2.63	17
Polyommatus icarus	3	80	3.39	9
Polygonia c-album	4	104	4.27	16
Anthocharis cardamines	4	156	6.51	17
Pieris napi	4	162	6.71	14
Thymelicus lineola/ sylvestris	1	163	6.86	17
Aglais urticae	4	339	14.18	19
Pararge aegeria	1	430	17.78	18
Aglais io	4	430	18.09	19
Pyronia tithonus	1	490	20.74	17
Aphantopus hyperantus	1	1259	52.66	19
Pieris rapae	4	2115	88.61	19
Maniola jurtina	1	2187	91.10	19
Pieris brassicae	4	2281	94.77	19

Table 5.5: The abundance and distribution of the 22 butterfly species observed across the 19 sample sites. The total abundance refers to the number of individuals counted, the standardised abundance accounts for transect length and occurrence refers to the number of sites in which the species was recorded. The corresponding Ecological Attribute Group (EAG) for each species is provided.

5.3.3 Comparisons between habitat suitability of sample sites: PH1 2010

The 19 sample sites ranged in their suitability for supporting all butterfly species and species comprising EAG1 and EAG2 according to the predicted values from the PH1 combined models derived from the PH1 2010 data set. The predicted values for the 19 sites obtained from the three PH1 models (PH1-ALL_{comb}, PH1-$EAG1_{comb}$, PH1-$EAG2_{comb}$) differ, with an insignificant positive correlation between PH1-ALL_{comb} and PH1-$EAG1_{comb}$ (r = 0.341, p = 0.152), and an insignificant negative correlation with PH1-$EAG2_{comb}$ (r = -0.358, p = 0.133). Predicted values from PH1-$EAG1_{comb}$ and PH1-$EAG2_{comb}$ did not correlate significantly with each other either (r = 0.411, p = 0.080). In several cases, sites with high predicted occurrence from the PH1-

ALL_{comb} model were characterised by low predicted occurrence from the PH1-$EAG1_{comb}$ and PH1-$EAG2_{comb}$ models.

Correlations between models: Bubbenhall site

Out of the 19 sample sites, Bubbenhall is the only site characterised by high predictions from all three PH1 models, and as such is consistently grouped within habitat suitability group 4. When considering the parameters of the models (Table 5.4), this site comprises high coverage of the habitat 'quarry' (PH-31; 22.75ha) (Figure 5.7a) in comparison to the other sites and this variable has a strong positive coefficient in the PH1-ALL_{comb} and PH1-$EAG1_{comb}$ models (Table 4.26). Furthermore, this site is characterised by high coverage of broad-leaved semi-natural woodland (PH-1; 20.19 ha), which has a strong positive coefficient in the PH1-ALL_{comb} model, and the PH1-$EAG2_{comb}$ model, and high coverage of mixed plantation woodland (PH-6; 0.88 ha) which has a strong positive coefficient in the PH1-$EAG2_{comb}$ model (Figure 5.7a; Table 4.26). Furthermore, the site comprises low coverage of arable (PH-34; 38.81 ha), which has a strong negative coefficient in the PH1-ALL_{comb} model.

In terms of the structure and connectivity of the landscape, the values for the metrics within the models are comparable to the average for the rest of the sites, particularly for those within the EAG1 model (Table 5.4). However, the site is characterised by high variability around the mean Euclidean distance between patches of the same type (ENN_SD = 309 m), in comparison to the average for the remaining sites (ENN_SD = 148 m) (Table 5.4), and this variable has a positive coefficient in the PH1-ALL_{comb} model (Table 4.26). This site also has high connectivity of woodlands/ hedgerow (P1H_IIC = 0.60) in comparison to the average for the 19 sites (P1H_IIC = 0.485) and this has a high positive coefficient in the PH1-$EAG2_{comb}$ model.

Correlations between models: Packington site

The habitat suitability of the Packington site is predicted to be very low (suitability group 1) for all butterfly species (PH1-ALL), contrasting to very high predictions from the PH1-$EAG1_{comb}$ and PH1-$EAG2_{comb}$ models, and as such it is characterised by high habitat suitability for these species (suitability group 4). A low prediction is observed for this site from the PH1-ALL_{comb} model because this site comprises a

very low proportion of semi-natural broad-leaved woodland (PH-1; 3.94 ha) (Figure 5.7b), which has a significant coefficient in the PH1-ALL_{comb} model but is not included in the PH1-$EAG1_{comb}$ model. Although semi-natural broad-leaved woodland is also a significant positive coefficient of the PH1-$EAG2_{comb}$ model, Packington comprises a very high coverage of continuous bracken (PH-22; 14.1 ha) (Figure 5.7b) in comparison to the other sites (Table 5.4) and this has a strong positive coefficient in the PH1-EAG2 model. Furthermore, the variability in the mean similarity index (SIMI_CV) is higher than average for this site (181.02 % compared to 173.27 %) and this variable is also a positive coefficient of the PH1-EAG2 model. High variability in the size and distance between similar patch types (SIMI_CV) is evident between patches of broad-leaved semi-natural woodland (PH-1), broad-leaved plantation woodland (PH-2), mixed plantation woodland (PH-3) and dense/ continuous scrub (PH-7) (Figure 5.7b).

High predictions are observed from the PH1-EAG1 model for this site as there is high coverage of dense/continuous scrub (PH-7; 0.43 ha), and above average coverage of semi-improved acidic grassland (PH-14; 0.06 ha), and semi-improved neutral grassland (PH-16; 34.69 ha) (Table 5.4; Figure 5.7b). These habitats have strong positive coefficients within the PH1-EAG1 model, particularly acidic grassland.

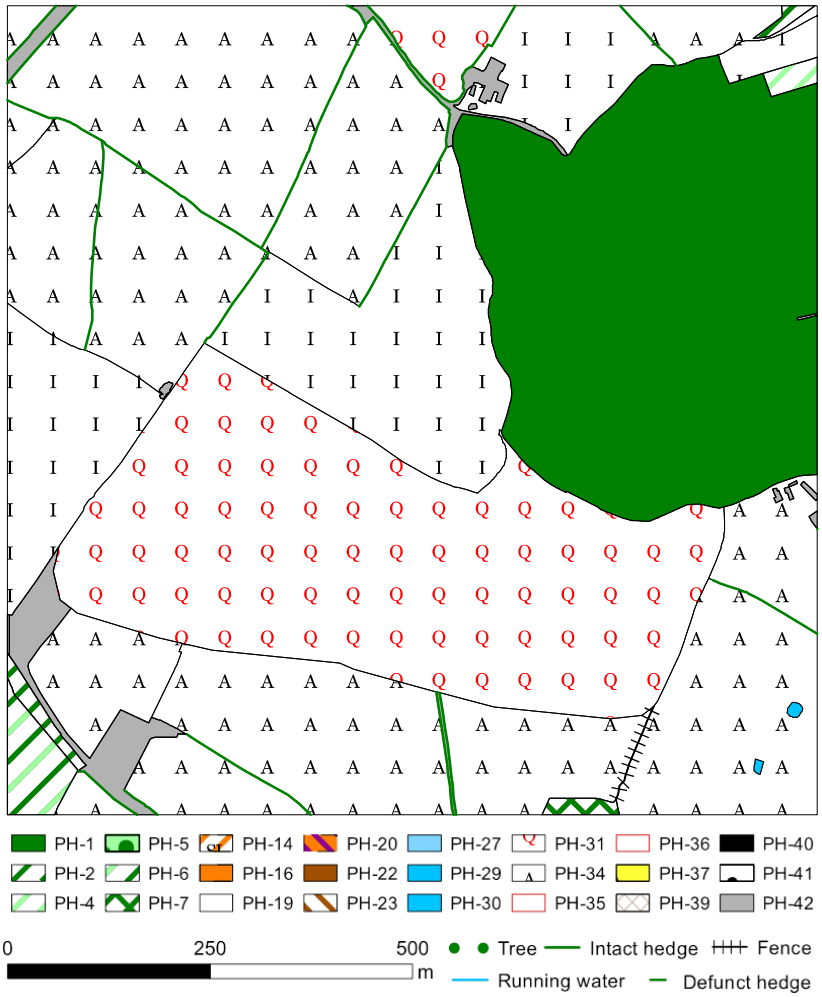
Correlations between models: Wappenbury

The predicted values for the site Wappenbury are consistently low between model predictions (PH1-ALL_{comb}, PH1-$EAG1_{comb}$, PH1-$EAG2_{comb}$) and as such this site is grouped within habitat suitability group 1 for the PH1-ALL_{comb} and PH1-$EAG1_{comb}$ models and within habitat suitability group 2 for the PH1-$EAG2_{comb}$ model. A low predicted value is obtained for this site from the PH1-ALL_{comb} model because this site is characterised by above average coverage of mixed semi-natural woodland (PH-5; 0.19 ha), and arable land (PH-34; 53.7 ha) (Table 5.4; Figure 5.7c), both of which have negative coefficients in the PH1-ALL_{comb} model. Furthermore, this site does not comprise any habitats which have strong positive coefficients in the PH1-ALL_{comb} model (Table 4.26). The structure of this site is characterised by above average range in patch extent (GYRATE_RA; 307.39 ha), with a few small patches of plantation woodland (PH-2; PH-6) and tall ruderal (PH-23) amongst very large

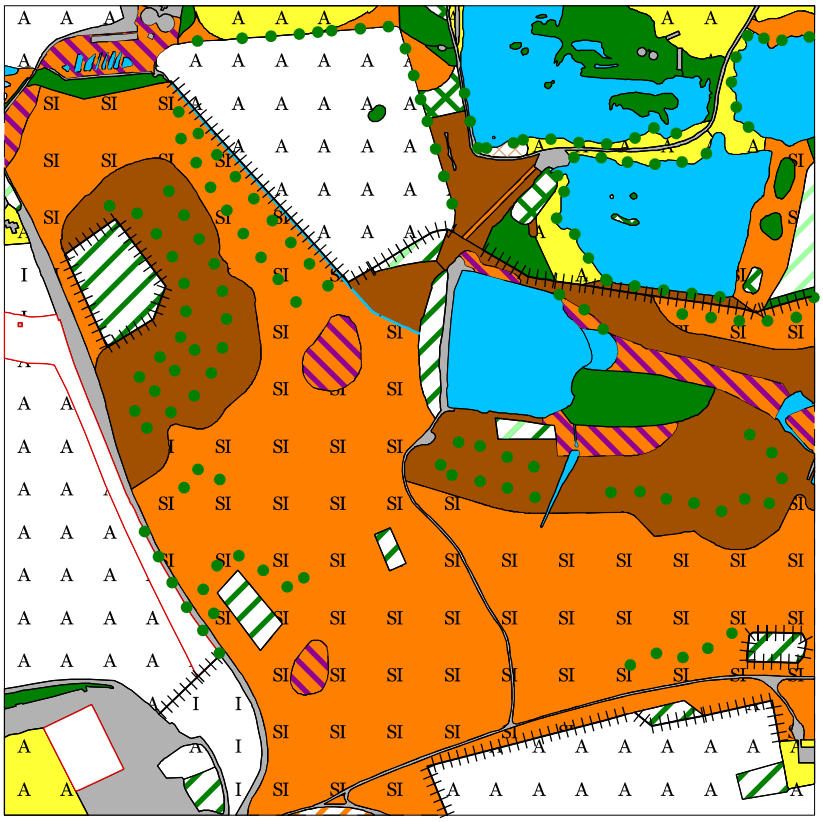
patches of improved grassland (PH-19) and arable land (PH-34) (Figure 5.7c; Table 5.4). This landscape structure metric has a negative coefficient in the PH1-ALL_{comb} model.

A low prediction is obtained for Wappenbury from the PH1-$EAG1_{comb}$ model as this site comprises only one of the seven habitats within the model which has a positive relationship with EAG1 species occurrence, and this is semi-improved neutral grassland (PH-16), which has below average coverage (1.56 ha) (Table 5.4; Figure 5.7c). Above average value is obtained for the landscape structure metric mean edge contrast (ECON_MN), with 84.61 % of patch perimeter at maximum edge contrast within the landscape, and this metric has a negative coefficient in the PH1-$EAG1_{comb}$ model (Figure 5.7c; Table 5.4). The high degree of contrast is evident from the high proportion of arable land within the site, bordered by improved grassland and semi-natural habitats (Figure 5.7c). Additionally, a low prediction is also obtained for Wappenbury from the PH1-$EAG2_{comb}$ model because it does not comprise any of the habitats within this model, being dominated by arable land and improved grassland (Figure 5.7c).

(a)


PH-1
PH-2
PH-4
PH-5
PH-6
PH-7
PH-14
PH-16
PH-19
PH-20
PH-22
PH-23
PH-27
PH-29
PH-30
PH-31
PH-34
PH-35
PH-36
PH-37
PH-39
PH-40
PH-41
PH-42
0
250
500
m
Tree
Intact hedge
Fence
Running water
Defunct hedge

(b)



(c)

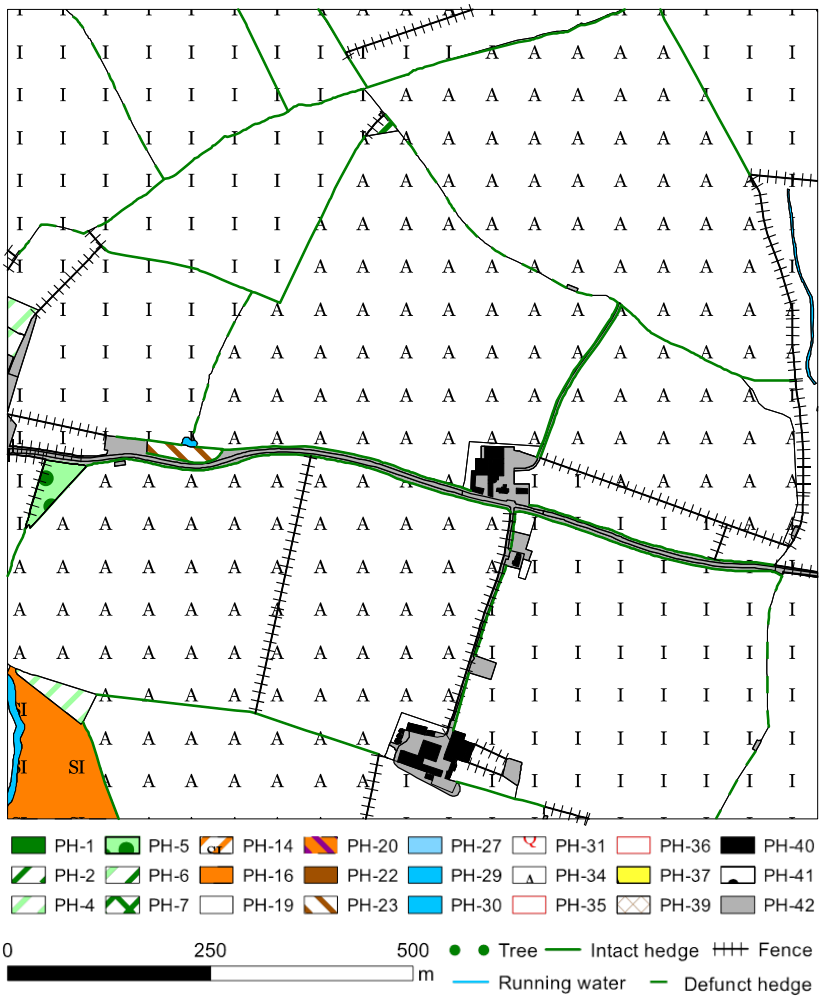


Figure 5.7: The contrasting composition and structure of three sample sites as classified by the PH1 2010: (a) Bubbenhall (b) Packington and (c) Wappenbury. See table 2.5 for PH1 habitat descriptions.

5.3.4 Phase 1 predictive model: relationships between combined models and butterfly community measurements

Significant positive correlations were observed between the predictions from the 'all butterfly' species model (PH1-ALL_{comb}) and a number of the butterfly community measurements. In particular, the strongest correlation occurred between the site predicted values and the Shannon's diversity index for all species (r = 0.597, p = 0.007) (Figure 5.8a). The predicted values for the 19 sites from the PH1-ALL_{comb} model were also significantly correlated with the overall species richness for each of the 19 sites (r = 0.544, p = 0.016), the abundance of EAG1 species (r = 0.561, p = 0.012), and the abundance of EAG3 species (r = 0.521, p = 0.022). No significant correlations were observed between the predicted site values and the overall species abundance, the species richness of each EAG or the abundance of EAG2 species. When considering the clustering of the 19 sites into habitat suitability groups based on the distribution of the predicted values from the PH1-ALL_{comb} model (Figure 5.2), only the Shannon's diversity index significantly differed between the suitability groups ($F_{3,15}$ = 4.269, p = 0.023), with significantly higher Shannon's diversity index in suitability group 4 in comparison to suitability group 1 ($p<0.05$).

Similar relationships were observed for the predicted values from the PH1-$EAG1_{comb}$ model, with significant positive correlations with the overall species richness (r = 0.551, p = 0.015), Shannon's diversity index (r=0.560, p =0.013), species richness of EAG1 (r = 0.496, p =0.031) and the abundance of EAG3 (r = 0.467, p = 0.044). The strongest correlation was observed between the site predicted values and the abundance of EAG1 species (r = 0.623, p =0.004) (Figure 5.8b). When considering the four habitat suitability groups based on the distribution of the predicted values from the PH1-$EAG1_{comb}$ model (Figure 5.2), only the average abundance of the EAG1 species differed significantly between the suitability groups ($F_{3,15}$ = 4.070, p = 0.027), with significantly higher abundance of EAG1 species in suitability group 4 in comparison to suitability group 1 ($p<0.05$). For the predictions obtained from the PH1-EAG2 model, only the abundance of EAG2 species significantly correlated with model predictions (r = 0.545, p = 0.016) (Figure 5.8c), and differed significantly between the four habitat suitability groups ($F_{3,15}$ = 7.031, p = 0.004), with significantly higher abundance of EAG2 species in suitability group 4 in comparison to suitability group 1 ($p<0.05$).

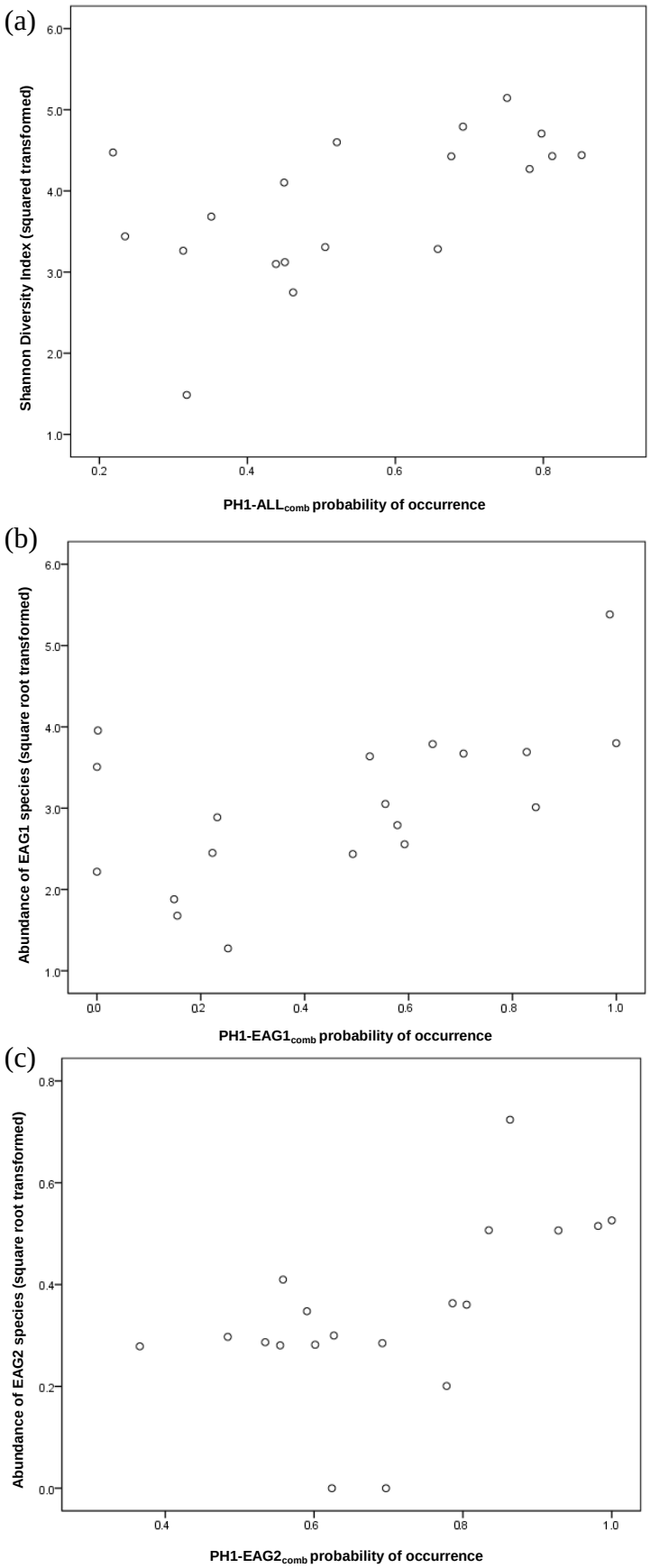


Figure 5.8: The relationship between model predictions and observed butterfly data for (a) the all butterfly species (PH1-ALL$_{comb}$) model predictions and overall Shannon's Diversity Index (b) the EAG 1 (PH1-EAG1$_{comb}$) model predictions and the abundance of EAG 1 species and (c) the EAG 2 (PH1-EAG2$_{comb}$) model predictions and the abundance of EAG 2 species.

5.3.5 LCM 2000 predictive model: relationships between combined models and butterfly community measurements

Significant positive correlations were observed between the predictions from the 'all butterfly' species model (LCM-ALL_{comb}) for the 19 sites and the butterfly Shannon's diversity index (r = 0.526, p = 0.021) (Figure 5.9a), and the abundance of EAG1 species (r = 0.524, p = 0.021). No significant correlations were observed between the predicted site values and the following butterfly community measures; overall abundance, species richness of EAG1, EAG2 or EAG3 or the abundance of EAG2 or EAG3.

The grouping of the 19 sample sites into four habitat suitability groups, based on the distribution of the predicted values from the LCM-ALL_{comb} model (Figure 5.6), revealed significant differences in the average Shannon's diversity index between suitability groups ($F_{3,15}$ = 5.708, p = 0.008), with significantly higher diversity in suitability group 4 in comparison to group 1 ($p<0.05$). The average abundance of EAG1 species also differed significantly between suitability groups ($F_{3,15}$ = 5.460, p = 0.010), with group 3 comprising significantly higher abundance in comparison to group 1 ($p<0.05$).

For the predicted values from the LCM-$EAG1_{comb}$ model significant positive correlations were observed with the abundance of all species (r = 0.633, p =0.004) (Figure 5.9b), overall species richness (r = 0.475, p =0.040), species richness of EAG4 (r = 0.458, p =0.049) and the abundance of EAG1 (r = 0.499, p =0.030). When considering differences in butterfly community measures between the four habitat suitability groups, based on the distribution of predicted values from the LCM-$EAG1_{comb}$ model (Figure 5.6), the average abundance of 'all butterfly' species significantly differed between suitability groups ($F_{3,15}$ = 3.628, p = 0.038). Significant differences between suitability groups were also obtained for the average species richness ($F_{3,15}$ = 4.129, p = 0.025), and abundance of EAG1 species ($F_{3,15}$ = 4.757, p = 0.016). In all cases suitability group 4 comprised significantly higher average total abundance, species richness, and abundance of EAG1 species in comparison to suitability group 1 ($p<0.05$).

For the predicted values from the LCM-$EAG2_{comb}$ model no significant correlations were observed with the community measurements for the observed EAG2 species across the 19 sample sites. However, a significant negative correlation was observed

between the predicted values from this model and species richness of EAG3 (r = -0.464, p = 0.045). Sites grouped by their habitat suitability (based on the distribution of the predicted values from the PH1-EAG2$_{comb}$ model; Figure 5.6) were found not to significantly differ in terms of the butterfly community measurements of average abundance, species richness, and diversity of all species and species EAGs.

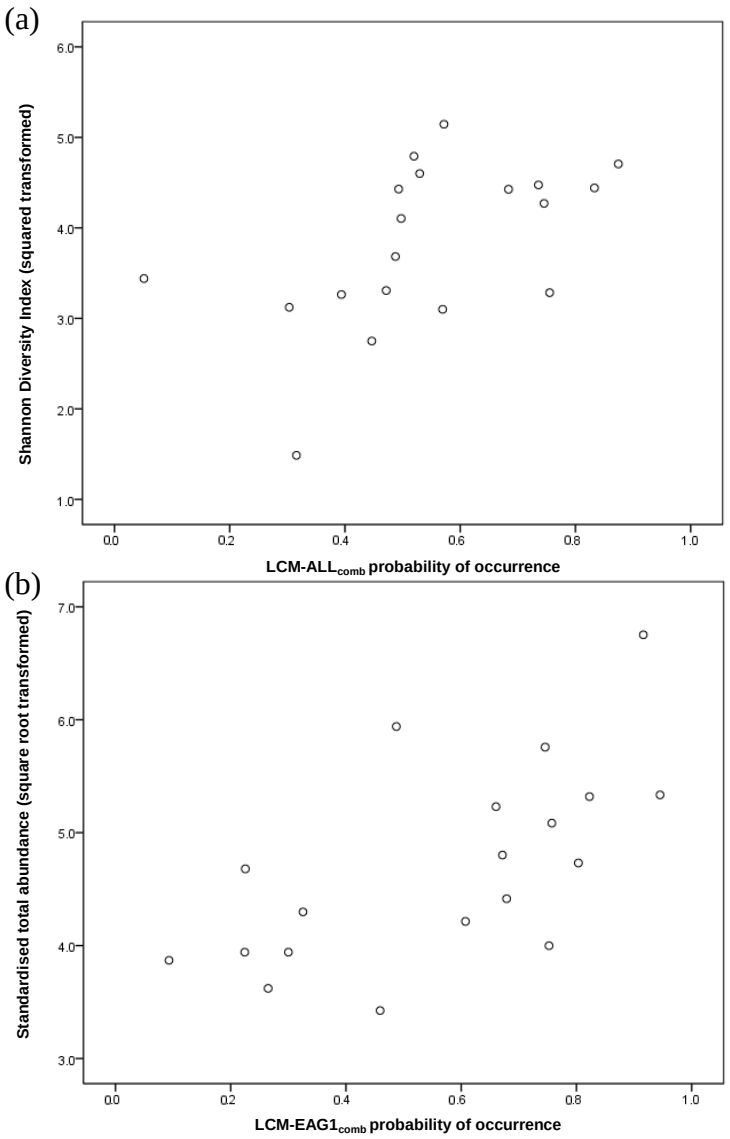


Figure 5.9: The relationship between LCM model predictions and observed butterfly data for (a) the all butterfly species model (LCM-ALL$_{comb}$) predictions and overall Shannon's Diversity Index and (b) the EAG 1 model (LCM-EAG1$_{comb}$) predictions and the standardised total abundance of all species.

5.3.6 Relationships between butterfly observations and local habitat characteristics

There were several positive correlations between spring vegetation variables and the butterfly community measurements with understorey and ground layer species diversity and species richness being particularly important (Table 5.6). For example, the species richness and abundance of EAG3 butterfly species is significantly positively correlated with ground layer diversity (G_SD; r = 0.576; r = 0.573 respectively), ground layer species richness (G_SR; r = 0.608; r = 0.650), the percentage cover of ground layer herbaceous plants (GHERB; r = 0.543; r = 0.570) and understorey layer diversity (U_SD; r = 0.493; r = 0.544). Furthermore, the percentage cover of herbaceous plants in the ground layer and the ground layer species diversity is significantly positively correlated with the overall diversity of butterfly species (r = 0.513; r = 0.633 respectively). However, field layer diversity (F_SD) and the cover of field layer herbaceous plants (FHERB) is positively correlated with the total abundance of butterfly species (r = 0.616; r = 0.514 respectively).

When considering the summer vegetation characteristics different positive correlations are observed, with field layer diversity (F_SD), species richness (F_SR) and percentage cover of field layer herbaceous plants (FHERB) being important for overall butterfly species diversity and richness, in addition to ground layer and understorey layer species diversity (G_SD; U_SD) (Table 5.6b). Specifically, the abundance of EAG1 and EAG3 species is significantly positively correlated with the cover of field layer herbaceous plants (FHERB; r = 0.514; r = 0.795), field layer diversity (F_SD; r = 0.527; r = 0.591), and field layer species richness (F_SR; r = 0.603; r = 0.510). In contrast the cover of shrub in the field layer (FSHRUB) is significantly negatively correlated with the species richness and abundance of EAG3 species (r = -0.523; r = -0.665) and the richness of EAG4 species (r = -0.484).

(a)

Butterfly Metric	FGRASS	FHERB	FSHRUB	F_SD	F_SR	GBARE	GGRASS	GHERB	G_SD	G_SR	FERN	UNDER	U_SD	U_SR
Abundance	0.063	0.514	-0.053	0.616	0.306	-0.113	0.015	0.098	-0.056	0.357	-0.075	-0.010	0.393	0.132
Species richness	0.268	0.409	-0.004	0.387	0.165	0.006	-0.046	0.441	0.464	0.428	0.441	0.416	0.340	0.550
Shannon's Diversity	0.404	0.336	-0.153	0.088	-0.034	-0.121	-0.281	0.513	0.633	0.451	0.368	0.369	0.414	0.590
EAG 1 richness	0.178	0.243	0.080	0.055	0.065	0.011	0.146	0.339	0.262	0.332	0.347	0.202	0.339	0.462
EAG 2 richness	0.132	0.152	0.153	0.069	0.157	-0.202	0.418	0.252	0.086	0.129	0.293	0.057	-0.445	-0.073
EAG 3 richness	0.357	0.306	-0.197	0.489	0.192	-0.131	-0.307	0.543	0.576	0.608	0.222	0.281	0.493	0.463
EAG 4 richness	0.270	0.157	-0.136	0.534	0.065	-0.043	-0.271	0.423	0.561	0.225	0.179	0.261	0.385	0.372
EAG 1 abundance	0.253	0.428	0.193	0.355	0.225	0.071	-0.067	0.286	0.244	0.543	0.416	0.261	0.566	0.482
EAG 2 abundance	0.362	0.008	-0.124	0.074	-0.025	-0.271	0.220	0.323	0.259	0.046	0.122	-0.163	-0.148	-0.121
EAG 3 abundance	0.332	0.402	-0.186	0.393	0.170	-0.124	0.034	0.570	0.573	0.650	0.209	0.092	0.544	0.317
EAG 4 abundance	-0.153	0.274	-0.234	0.416	0.156	-0.214	-0.015	-0.180	-0.310	-0.066	-0.499	-0.169	0.002	-0.165

(b)

Butterfly Metric	FGRASS	FHERB	FSHRUB	F_SD	F_SR	GBARE	GGRASS	GHERB	G_SD	G_SR	FERN	UNDER	U_SD	U_SR
Abundance	0.003	0.357	-0.175	0.480	0.476	0.223	-0.396	-0.132	0.313	0.265	-0.013	-0.156	0.327	0.169
Species richness	0.106	0.656	-0.355	0.275	0.460	0.088	-0.267	0.111	0.488	0.247	0.186	0.190	0.578	0.458
Shannon's Diversity	0.314	0.703	-0.296	0.463	0.534	0.207	-0.264	0.293	0.493	0.091	0.472	0.289	0.477	0.565
EAG 1 richness	0.105	0.201	-0.199	-0.010	0.172	-0.272	0.032	-0.011	0.204	0.039	0.222	0.050	0.513	0.300
EAG 2 richness	-0.057	-0.174	0.181	-0.181	-0.117	-0.229	0.366	0.337	0.205	0.299	-0.118	-0.059	-0.142	-0.138
EAG 3 richness	0.288	0.750	-0.523	0.517	0.524	0.115	-0.332	0.117	0.498	0.110	0.250	0.110	0.517	0.405
EAG 4 richness	0.086	0.644	-0.484	0.268	0.478	0.211	-0.358	0.189	0.533	0.205	0.291	0.111	0.579	0.435
EAG 1 abundance	0.247	0.514	-0.305	0.527	0.603	0.296	-0.325	0.001	0.535	0.166	0.403	0.129	0.623	0.470
EAG 2 abundance	0.151	-0.082	0.042	-0.008	-0.002	-0.071	0.068	0.307	0.321	-0.002	0.227	-0.092	-0.072	-0.035
EAG 3 abundance	0.520	0.795	-0.665	0.591	0.510	0.214	-0.393	-0.101	0.485	-0.013	0.100	0.001	0.489	0.290
EAG 4 abundance	-0.263	-0.048	0.124	0.105	0.056	0.009	-0.213	-0.128	-0.099	0.243	-0.373	-0.229	-0.118	-0.119

Table 5.6: Relationship between the local habitat characteristics and butterfly community measurements of the 19 sample sites for (a) spring and (b) summer. Pearson product-moment correlation coefficients are displayed with light grey indicating significance <0.05, medium grey ≤0.01 and dark grey <0.001.

5.3.7 Butterfly species composition

Considering the habitat suitability groups identified from the PH1 and LCM models (see sections 5.2.3 and 5.2.4), and the similarity of the abundance and composition of butterfly species within each habitat suitability group, as measured by the Bray-Curtis measure of dissimilarity, significant differences in the composition of butterfly species were observed between habitat suitability groups. The abundance and composition of butterfly species differed significantly between the habitat suitability groups identified from the PH1-ALL_{comb} model ($_{ANOSIM}$ R = 0.298, p = 0.016) and the LCM-ALL_{comb} model ($_{ANOSIM}$ R = 0.416, p = 0.004). From the PH1-ALL_{comb} model, habitat suitability group 4 differed significantly in species composition in comparison to group 1 ($_{ANOSIM}$ R = 0.491, p = 0.011), and group 2 ($_{ANOISM}$ R = 0.417, p = 0.036). The same pattern was observed for the LCM-ALL_{comb} model, with habitat suitability group 4 differing significantly from group 1 ($_{ANOSIM}$ R = 0.635, p = 0.01), and group 2 ($_{ANOSIM}$ R = 0.715, p = 0.015).

When considering the similarities between the 19 sample sites based on the butterfly species composition irrespective of the model predictions, five clusters of sites can be identified (Figure 5.10a). Sites within each cluster are similar in terms of butterfly species composition at a similarity threshold of 70 %, as identified from a hierarchical cluster analysis. Most notably cluster group *c* comprised seven of the 19 sites, and cluster group *e* comprised five of the 19 sites (Figure 5.10a). When considering the similarities between sites based on the coverage and composition of spring and summer vegetation, different sites are considered to be similar depending on the season (Figure 5.10b and 5.10c). For example, the sites ID 2702 and ID 905 are relatively similar in terms of spring vegetation composition (Figure 5.10b), but are relatively dissimilar in terms of summer vegetation composition (Figure 5.10c). The similarity of the butterfly species communities between sites is significantly correlated with the similarity of spring vegetation composition between sites (r = 0.432, p <0.001) (Figure 5.11a). The butterfly community groups *c* and *e* remain clustered when the sites are plotted according to their degree of similarity in spring vegetation composition (Figure 5.10b). However, the similarity of the butterfly communities between sites is not significantly correlated with similarities of summer vegetation composition (r = 0.063, p = 0.404) (Figure 5.11b), and the butterfly community groups, including groups *c* and *e* separate when the sample sites are

plotted according to their degree of similarity in summer vegetation composition (Figure 5.10c).

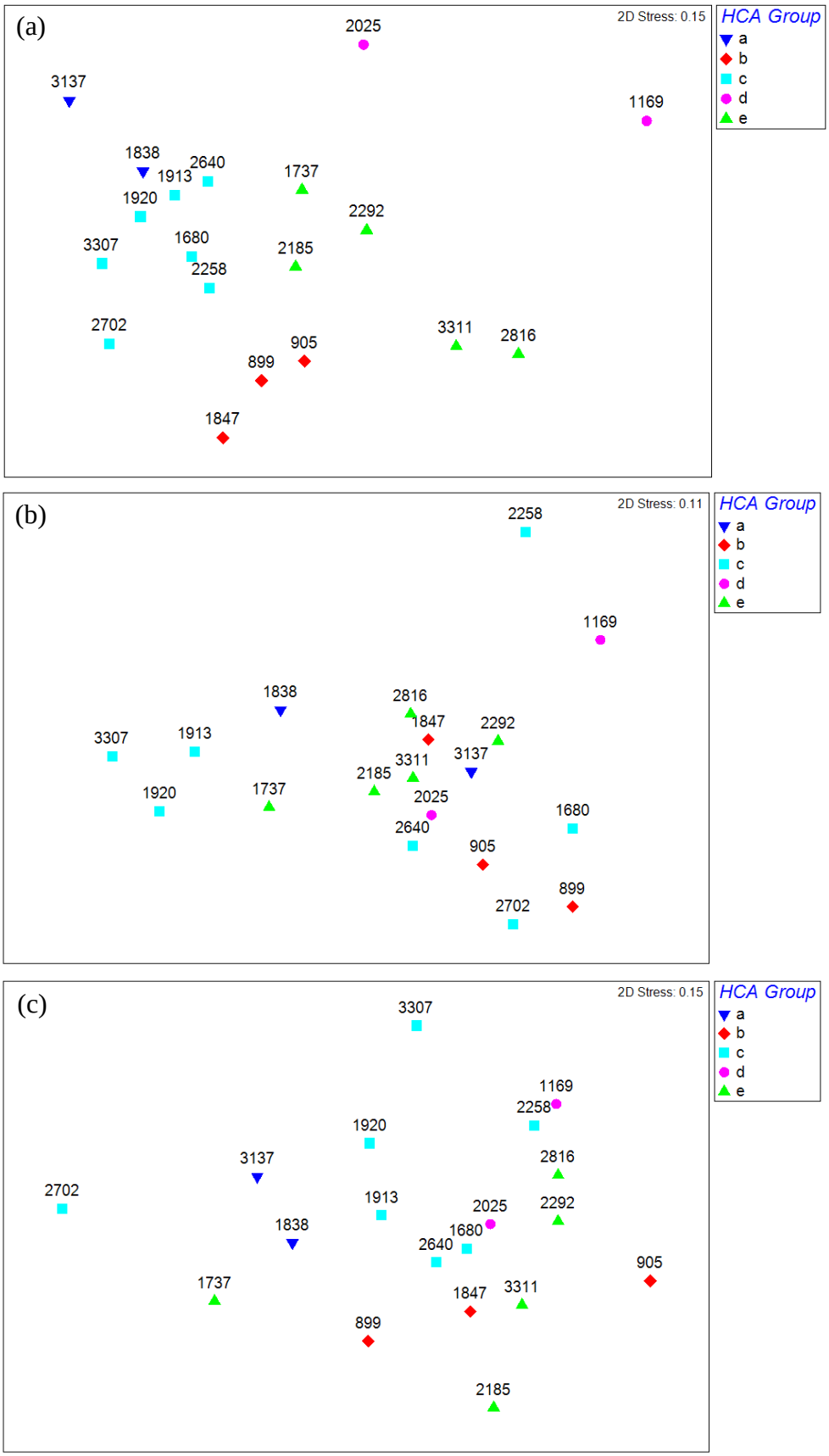


Figure 5.10: Non-metric Multi-dimensional Scaling ordination for the 19 sample sites (axes 1 and 2) based on dissimilarity in (a) butterfly composition, (b) spring vegetation composition and (c) summer vegetation composition. Position of and distances between symbols represent relative dissimilarity between sites in 2D space. Different symbols/colours indicate groups identified from a hierarchical cluster analysis (HCA) on the Bray-Curtis butterfly dissimilarity matrix using a similarity threshold of 70 %. Labels represent the ID for each of the 19 sample sites.

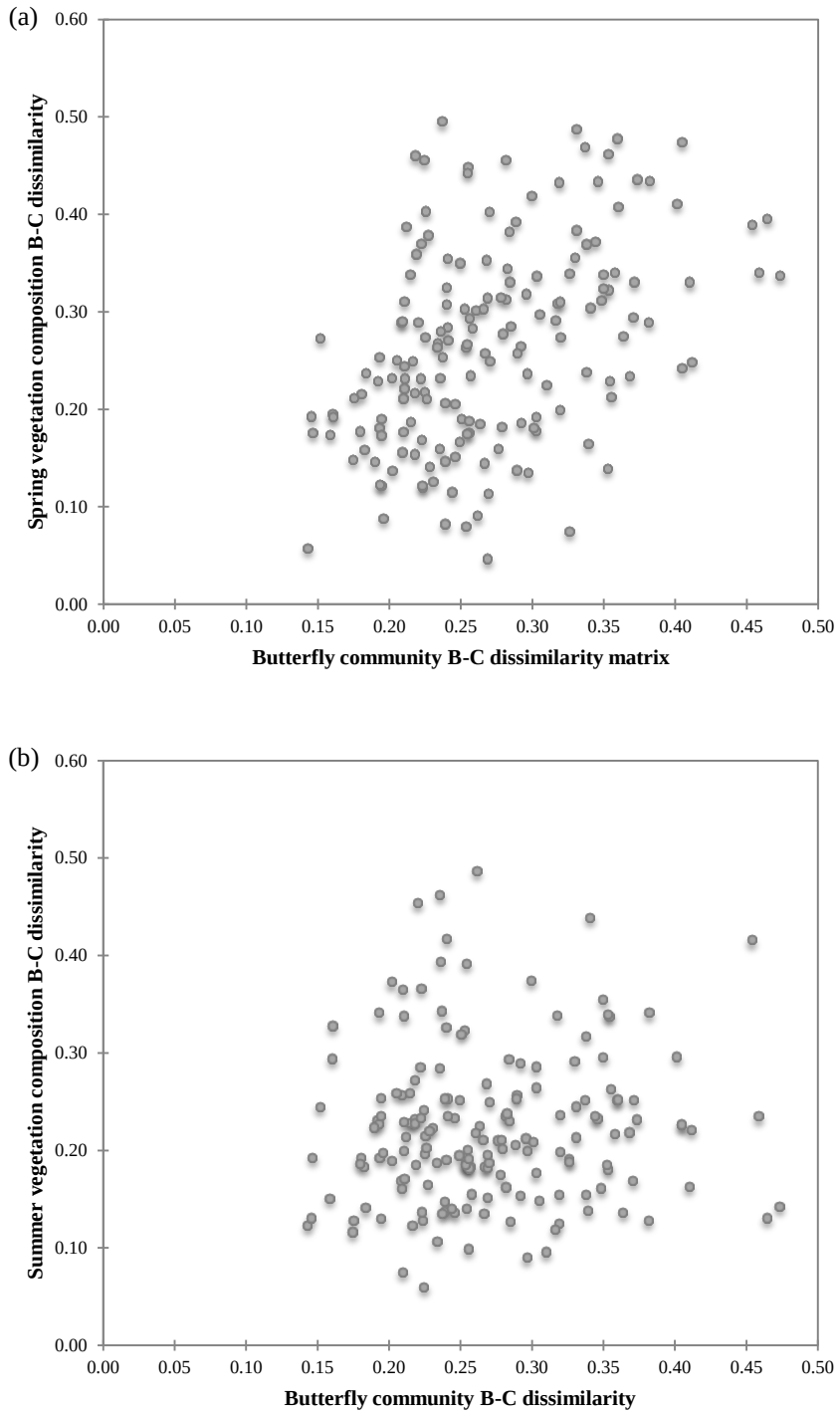


Figure 5.11: Pearson Product Moment Correlation derived from a Mantel test between the Bray-Curtis (B-C) dissimilarity matrix of butterfly species composition and the Bray-Curtis (B-C) dissimilarity matrix of (a) spring vegetation composition and (b) summer vegetation composition.

5.4 Discussion

The accuracy of the PH1 combined models in predicting the Warwickshire butterfly 2000-2009 distribution data using the PH1 2010 habitat map are comparable to the accuracy of the models developed using the initial training data (developed in section 4.2.9). The PH1 combined models for predicting all butterfly species, EAG1 and EAG2 species have proven to be transferable temporally, with comparable model accuracy and performance to models developed using the training data**.** There is little difference in the specificity and the sensitivity of the models between predictions for 1990-1999 and 2000-2009, with PH1-EAG1$_{comb}$ being the best at predicting distribution accurately in terms of specificity/ sensitivity and the Receiver Operating Characteristic (ROC) area under the curve (AUC). In all cases, the models are considered to be 'fair' predictors of observed butterfly data according to classification of the ROC area under the curve (AUC) by Araujo *et al.*, (2005), which provides a more reliable estimation of model performance as it is threshold independent. This measure of model accuracy is particularly important as all possible thresholds have been used for determining the presence/absence of grid squares based on the predictions from individual models, and as such allows for more reliable comparisons between model performances (Allouche, *et al.*, 2006; Manel, *et al.*, 2001).

In contrast the accuracy of the LCM combined models (LCM-ALL$_{comb}$, LCM-EAG1$_{comb}$ and LCM-EAG2$_{comb}$) in predicting the Warwickshire butterfly 2000-2009 distribution data is not as good as that obtained for predicting the observed 1990-1999 data set, particularly the LCM-ALL$_{comb}$ model, with a 44.6% decrease in the sensitivity of the model. However, this may not be a reflection of changes that may have occurred in the landscape during this time period, moreover strong correlations between the PH1 2000 model predictions and the PH1 2010 model predictions suggest that there has been little change in the suitability of the landscape based on those model parameters. The differences in model accuracy may be attributed to the proportion of prevalence in the data set, which directly influences the threshold set for determining presence or absence. The training data for the development of the LCM 2000 model was based on a proportion of 0.212, whereas the proportion of prevalence of the 2000-2009 data is 0.108. This means a higher threshold value was

used for the 1990-2000 data set than is appropriate for the 2000–2009 dataset, which emphasises the importance of using other measures of model validation such as ROC AUC values (Luoto, *et al.*, 2006). ROC values obtained for the two time periods are similar for all LCM combined models, but the LCM-ALL_{comb} and LCM-$EAG2_{comb}$ models are both considered to be 'poor' in predicting the observed data set (Araujo, *et al.*, 2005). The LCM-$EAG1_{comb}$ model maintains 'fair' discrimination of the two temporal butterfly data sets, indicating that this model can be used for accurately predicting the presence-absence of EAG1 species.

Comparing the PH1 combined models in time, model predictions are strongly significantly correlated suggesting that there has been little change in the PH1 habitats within individual grid squares over the last 10 years. Yet logic would suggest current landscape data is important for predicting butterfly population dynamics and distribution – butterflies have a very rapid response rate to disturbance events and have a short lifespan and as such responses to habitat and land use change are expected (Kumar, *et al.*, 2009; Thomas, 2005). Thus it is important to have continually up to date data, as small changes in the landscape can be expected to dramatically change butterfly population dynamics. In particular, changes in the area of landscape variables with high coefficients in the predictive models will have a large influence on the suitability of the landscape for butterfly occurrence. For example, holding all other variables constant a one-unit increase in the area of unimproved calcareous grassland (PH-17), will increase the log-odds of butterfly occurrence by 2.410 ($p = 0.004$).

For both the LCM and PH1 combined models the transferability of the EAG3 species model is very poor with low sensitivity and AUC values obtained in both instances. This could be due to the limited prevalence of the species which have a poor distribution, thus skewing the model probabilities to be extremely low, and as such model performances were poor when temporally transferred. Realistically the use of EAG3 species in models at a county wide scale are impractical due to the specific requirements needed for EAG3 butterfly species to thrive, with particular association with calcareous grasslands (Shreeve, *et al.*, 2001), and the poor coverage and distribution of this grassland type across Warwickshire occurring in less than 1 % of

grid squares. From this we can conclude that landscape can be used to predict butterfly occurrence for 'All butterflies', EAG1 and EAG2, but not for EAG3.

Considering the predicted values from logistic regression presence-absence models as an indicator of habitat suitability or as a probability of occurrence has been widely suggested as a more preferable use of such models than simply determining presence-absence (Hirzel, *et al.*, 2006; Vaughan and Ormerod, 2005). Determining the relationship between probabilities of occurrence and abundance data then provides a desirable approach for model validation in addition to the threshold independent AUC method (Guarino, *et al.*, 2012; Luoto, *et al.*, 2006). The Warwickshire butterfly validation dataset was supplemented by the sample site data (19 sites), to provide more accurate relationships with butterfly community measurements, measures of abundance in particular. The collection of butterfly data across the 19 sample sites was standardised and transect width maintained, which are both important elements in controlling for differences in detectability amongst species, which although uncounted for, will be constant amongst surveys as a consequence (Nowicki *et al.*, 2008). Maintaining constant species detectability and survey effort facilitates the inference of abundance measures which are comparable across sites (Gross, *et al.*, 2007; Nowicki, *et al.*, 2008).

The probability of occurrence obtained from the PH1-ALL_{comb} and LCM-ALL_{comb} models are significantly correlated with standardised abundance, species richness and diversity of 'all butterfly' species from the Warwickshire 2000-2009 data set. When considering the relationships observed from the 19 sample sites, however, site predictions from the PH1-ALL_{comb} model only correlated with overall species diversity, and species richness and predictions from the LCM-ALL_{comb} model correlated only with overall diversity. Correlation between the PH1-ALL_{comb} and LCM-ALL_{comb} model predictions with overall abundance may not have been detected for the sample sites, as sampling was conducted during one year (2013), which was characterised by a prolonged period of highly suitable weather for butterfly species. As such, each site was characterised by high overall species abundance. Weather directly affects relative abundance, with overestimation of abundance often occurring in good weather seasons (Nowicki, *et al.*, 2008). When considering the similarity in species composition between sites however, significant

differences were observed between sample sites of high suitability (group 4) compared to sites of low suitability (group 1) according to the predictions from both the LCM-ALL_{comb} and PH1-ALL_{comb} models, indicating that different butterfly communities were supported by sample sites which differed in their suitability. Furthermore, this is supported by the fact that the predictions from the PH1-ALL model were significantly correlated with the abundance of EAG1 and EAG3 species in the sample sites, and predictions from the LCM-ALL model were correlated with the abundance of EAG1 species.

Significant differences in the diversity of all species occurred between the PH1 and LCM suitability group 4 in comparison to suitability group 1 when considering both the Warwickshire data set and the sample data from the 19 sites. Furthermore, for the Warwickshire data set, suitability group 4 (derived from both the PH1-ALL_{comb} and LCM-ALL_{comb} models) was significantly different from all other groups, in terms of overall abundance, diversity and species richness. This is a reinforcement of the ability of the 'all butterfly species' models to clearly predict suitable habitats for butterflies in comparison to poor habitats for butterflies. When considering the findings from the Warwickshire data set and the sample sites together, these results suggest that the models for all butterfly species from both the PH1 and LCM data sets can be used to determine associated butterfly species composition, and overall diversity but not necessarily total abundance. Furthermore, the PH1-ALL_{comb} model predictions exhibited strong relationships with overall species richness.

Community measurements for individual EAGs have stronger correlations with model predicted values from both the PH1 and LCM models in comparison to the 'All butterfly models', suggesting that using inferred absences rather than classing 'no record' as an absence can create a stronger model, a likely consequence of the reliability of using 'no record' data.

When considering the two sets of butterfly validation data, the probabilities of occurrence from the PH1-$EAG1_{comb}$ model exhibited strong correlations with EAG1 species richness. Relationships were also observed with the abundance of EAG1 species, with significant differences in average abundance between suitability group 4 in comparison to group 1 for both butterfly data sets. Furthermore, a strong

correlation was observed between PH1 predicted values and EAG1 abundance from the sample site data. The probabilities of occurrence from the LCM-EAG1$_{comb}$ model exhibit a much weaker relationship with butterfly community measurements in comparison to that observed for the PH1-EAG1$_{comb}$ model. The observed relationships between LCM model predicted values and butterfly community measurements differ between the two sets of butterfly validation data. For the Warwickshire data set, significant correlations only occur between LCM predicted values and EAG1 species richness and EAG1 diversity. Furthermore, although significant differences were detected between the suitability groups (based on the predicted values from this model) in terms of EAG1 abundance, richness and diversity, suitability group 4 comprised similar community measurements to group 1.

When considering the sample site data, the LCM predicted values correlated with the abundance of all species, overall species richness, and the abundance of EAG1. Relationships between LCM model predictions and these community measurements were further supported by suitability group 4 comprising significantly higher average overall abundance, species richness, and abundance of EAG1 species in comparison to suitability group 1. The similarity between suitability group 1 and group 4 obtained for the Warwickshire data set, may then be a reflection of sampling bias. Exceptionally high abundance, richness and diversity of EAG1 species was associated with a few grid squares classified within suitability group 3. These grid squares were located next to grid squares characterised by high suitability for EAG1 species, with a UKBMS transect extending across both grid squares. As such, not only do they have a wealth of data associated with them collected over several years, which may then result in an over estimation of the population of EAG1 species in comparison to other sites of comparable suitability, but high observations of EAG1 species may be due to daily dispersal events from neighbouring grid squares of higher suitability. Similar diversity between sites classified within groups 4 and 1 may have occurred due to overestimation of diversity when considering EAG1 species separately from all species. High diversity was obtained for several grid squares within suitability group 1 characterised by low species richness of EAG1 species which were observed in equally low abundance, resulting in a high diversity index for EAG1 species. Due to these discrepancies, the relationship between model

predictions and the community measurements from the 19 sample sites can be considered to be more accurate and as such, the LCM-EAG1$_{comb}$ model can be considered to provide measurements of the abundance of EAG1 species, as well as overall abundance and species richness.

The predictions from both the LCM-EAG1$_{comb}$ and PH1-EAG1$_{comb}$ models also showed associations with the community measurements of other EAGs within the sample butterfly data. The predictions from the PH1-EAG1$_{comb}$ model correlated with the abundance of EAG3 species and the predictions from the LCM-EAG1 $_{comb}$ model correlated with the species richness of EAG4. EAG 1 and EAG 3 species are both associated with grassland biotopes, however, Shreeve *et al.,* (2001) found that the two different EAGs were associated with different structural variations in grassland vegetation, with EAG1 species associated with tall grassland and EAG3 with short sward grassland. It is likely then that heterogeneity within grassland habitats observed along transect sections results in variations in grassland structure and as a result provides resources required by both grassland groups. Indeed transect sections were observed to include short sward grasslands which were bordered by tall grasslands, particularly along field margins and hedgerows. For the LCM-EAG1 model the relationship with the species richness of EAG4 species is unsurprising considering that predicted values from this model were also associated with overall species richness, and EAG4 species were the most abundant and species rich of all the EAG species observed, occurring within every sample site. EAG4 species are typically generalist species with strong associations with tall ruderal vegetation. Shreeve *et al*., (2001) remark that heterogeneity in vegetation often provides resources suitable for EAG 4 species in addition to tall grassland for EAG1 species.

When considering the two butterfly validation data sets, the predicted values from the PH1-EAG2$_{comb}$ model significantly correlated with the abundance of EAG2 species, and this was further supported by significant differences between high suitability (group 4) and low suitability (group 1) grid squares in terms of EAG2 abundance. Several more relationships between the PH1 predicted values and the community measurements of EAG2 species were observed within the Warwickshire butterfly data set, with significant correlations with EAG2 species richness and significant differences in average EAG2 abundance, richness and diversity between

habitat suitability group 4 and all other suitability groups. Furthermore, predictions from the LCM-EAG2$_{comb}$ model were significantly correlated with the abundance of EAG2 species within the Warwickshire butterfly data set and significant differences were observed between suitability groups 4 and all other groups in terms of EAG2 abundance and richness. It is evident from the Warwickshire validation data set that grid squares which are more suitable according to the PH1-EAG2$_{comb}$ and LCM-EAG2$_{comb}$ models support a higher and more diverse population of EAG2 species. No relationships were observed between the LCM-EAG2$_{comb}$ model predictions and the observed EAG2 data from the 19 sample sites, and furthermore, no relationships were detected between the PH1-EAG2$_{comb}$ model predictions and the observed EAG2 species richness and diversity from the sample site data. However, this lack of relationships may reflect the low number of EAG2 species observed across the sample sites (two species). As strong relationships between predicted values (PH1) and species richness of EAG2 butterflies were observed for the Warwickshire data set, this suggests potential observer error in the detection of EAG2 species during the collection of sample data. EAG2 species, such as the purple emperor (*Apatura iris*) and the purple hairstreak (*Neozephyrus quercus*) are elusive species with low detectability amongst the woodland canopy, with woodlands species generally harder to detect in comparison to species associated with more open habitats (Liley, *et al.*, 2004; Pearce and Ferrier, 2001).

In this study significant differences were observed in abundance and other community measurements between highly suitable and unsuitable sites (suitability group 4 in comparison to group1) based on predicted probabilities. Some authors have argued that relationships between predicted values and abundance are due to differences in the mean predicted probability between occupied and unoccupied sites (Pearce and Ferrier, 2001), and that when removing unoccupied squares the idea of presence-absence data predicting abundance breaks down (Nielsen, *et al.*, 2005; Pearce and Ferrier, 2001). However this is somewhat counterintuitive as the main feature of a presence-absence model is to predict presence and absence, thus removing the absence squares prior to relating it to abundance takes away a key asset of the model which is to predict absence or 'zero abundance.' The significant differences in community measurements in this study are not necessarily occurring

however between present/absent sites, as the threshold for determining presence-absence varied with each model. In particular, the threshold for the PH1-EAG2 models (0.4) corresponds with the lower quartile of suitability group 3 and as such significant differences observed between groups 4 and 3 in EAG2 species richness and abundance for the Warwickshire data set occurred between occupied squares. Furthermore, the relationships between predicted probabilities and community measurements were considered for the Warwickshire data set, in addition to the occupied 19 sample sites, and as such the relationships which occur within both butterfly validation data sets, as described, can be considered to be robust.

Overall, the predicted values from the eight landscape models exhibited relationships with a range of butterfly community characteristics, including abundance. Predictions from presence-absence models have also been found to be successful predictors of butterfly abundance in the literature. Gutierrez *et al.,* (2013) found that distribution models based on presence-absence data for apollo butterfly *(Parnassius Apollo)* performed better in predicting abundance than quasi-Poisson regression models based on abundance data. Similarly, Guarino *et al.,* (2012) found that presence-absence models of plant species outperformed abundance models, with observed abundance correlated with probabilities of occurrence.

The observed relationship between probabilities of occurrence and butterfly community measurements suggests that these community characteristics of butterfly species (abundance, richness and diversity) are associated with similar landscape aspects as those which determine the distribution of butterfly species. Several studies have also found species distribution and abundance to be associated with mutual factors, with high similarity between the variables selected for in presence-absence and abundance models (Gaston, *et al.*, 2000; Gutierrez, *et al.*, 2013). Furthermore, sample sites which were similar in their composition of butterfly species were found to be similar also in their composition of spring vegetation. Spring vegetation is particularly important for the provision of larval host plants and several species have a strong association with specific plant species for the location of egg laying and the development of larvae (Garcia-Barros and Fartmann, 2009; Thomas, *et al.*, 2001).

Furthermore, the richness and coverage of larval host plants and nectar food plants were strongly correlated with the species richness of the ground layer in spring.

The abundance, diversity and richness of butterfly species across the 19 sample sites exhibited a strong relationship with the structure and composition of vegetation, both in spring and summer. Most notably the species richness and abundance of EAG3 species was shown to be correlated with spring ground layer diversity, richness and coverage of herbs, and species of this group have been documented to have strong associations with species rich short turf for basking, roosting, egg laying and larval development (Shreeve, *et al.*, 2001). Understorey species richness and diversity was also important vegetation characteristic for species richness and diversity of all butterfly species. Van Halder *et al.* (2008) suggested that structurally diverse deciduous woodlands provide suitable environments for mate location, diverse herbaceous vegetation cover and suitable variations in micro-climate which together provide complementary resources for several butterfly species.

5.4.1 Conclusion

This chapter demonstrates that the PH1 combined presence-absence models developed in section 4.2.9 can be transferred to a temporally independent landscape data set. Furthermore, the PH1 combined models are more accurate in discriminating between presence-absence squares for the 1990-1999 and 2000-2009 butterfly data sets in comparison to the LCM models, with only the LCM-EAG1$_{comb}$ model considered to be 'fair' in its discriminating ability.

This chapter also demonstrates that the predicted probabilities from the presence-absence models are associated with butterfly community measurements from both butterfly validation data sets, including species abundance, richness and diversity and overall species composition. In particular, stronger relationships are observed with the predicted values from the PH1 models in comparison to the LCM models and specific relationships with community measurements differ between the models of 'all butterfly species' and EAGs. Overall, simple presence-absence models can be related to butterfly characteristics and in turn provide an indication of habitat suitability. Furthermore, the strong associations detected between the observed butterfly community and the local habitat characteristics, particularly within the

spring, suggest that predictions of habitat suitability derived from landscape based models do reflect local habitat characteristics.

The landscape based models derived from the PH1 data can be used to predict the distribution and community characteristics of all butterfly species, and species comprising EAG1 and EAG2 across Warwickshire. The landscape based models derived from the LCM data, are limited in their application, however the accuracy of the model predicting the distribution of EAG1 species is comparable to the PH1 model. The strong relationship between the model predictions and the butterfly community characteristics, which are in turn intrinsically linked to diverse vegetation communities, suggests the potential for the models to provide predictions of wider biodiversity. The models could therefore be applied to the landscape of Warwickshire to identify sites with high biodiversity potential. The development of models for EAG3 species requires future exploration, potentially with greater spatial coverage of sites with ‘presence’. The transferability of the PH1 models temporally provides a positive indication that these models have the potential to be transferred spatially. Further development of the models would be required to ensure accurate spatial transformation in order to consider the range of habitats which are not present within the Warwickshire region used to develop the models. The modelling approach for development of the landscape based models has the potential for application to further species including birds and bats.

Chapter 6: General discussion

Predicting changes in the levels of biodiversity in response to changes in land cover is essential considering current and future rates of land use change (Lawler, *et al.*, 2011; Rushton, *et al.*, 2004; Vaughan and Ormerod, 2005). The role and value of landscape-based models or metrics as surrogate measures of biodiversity is unclear due to (1) conflicting responses of species to changes in the landscape at different spatial scales and (2) different spatial scales at which the landscape can be measured (grain size and thematic resolution) (Dover and Settele, 2009; Rossi and van Halder, 2010; Shreeve and Dennis, 2011). The aim of this thesis was to enhance our understanding of the application of metrics to measure landscape characteristics, and to provide answers to three main questions underpinning the development and use of landscape-based models in predicting patterns of biodiversity in a changing landscape:

1. Does grain size impact on the ability to characterise and discriminate between landscapes and what is the necessary grain-size to measure landscape characteristics?
2. Which measures of landscape structure, composition and connectivity should be used to predict indicators of biodiversity, and can we understand why they are good predictors?
3. Can predictions derived from landscape based models of species presence-absence predict community characteristics such as species richness or abundance?

Butterflies were used in this study as indicators of biodiversity; their value as indicators of biodiversity has been widely documented due to their rapid response to environmental change, broad geographic range and well-studied life histories (Asher, *et al.*, 2001; Fox, *et al.*, 2011; Thomas, 2005). In addition, extensive records of species presence and abundance are available for butterflies relative to many other taxa, though the reliability and spatial coverage of the data is still not ideal.

Two sources of landscape data were chosen for investigation in this study; the Land Cover Map (LCM 2000) obtained from the Centre for Ecology and Hydrology (Fuller, *et al.*, 2002) and the Warwickshire Habitat Biodiversity Audit (HBA) Phase 1 Habitat Map (PH1 2000 and 2010), obtained from Warwick County Council.

Landscape extents defined by the National Character Areas (NCAs) across the UK and 1 km grid squares across the County of Warwickshire were used to investigate the impact of changes in grain size on the discriminating ability of metrics and in turn to identify the best grain size to maximise landscape discrimination (Chapter 3). For the development of landscape based models of butterfly distribution landscape extents defined by 1 km grid squares across Warwickshire were used (Chapter 4).

To date, no study has investigated the impact of changes in grain size on the use of landscape structure metrics derived from widely available landscape data sources in discriminating between different landscapes. The first component of this thesis (Chapter 3) investigated how changes in grain size impacted the ability of 32 landscape structure metrics, both individually and in combination, to discriminate between landscapes. Previous studies have identified several metrics to have unpredictable or erratic responses to increases in grain size (Simova and Gdulova, 2012; Wu, 2004; Wu, *et al.*, 2002); however in Chapter 3 the same metrics identified in these previous studies appeared to have a discriminating ability between landscapes which was consistent across scales. At the national level, landscape structure metrics measuring both habitat fragmentation and patch shape consistently discriminated between different landscapes at grain sizes of between 25 m and 250 m. Furthermore the relationships between the values of these metrics were similar across the different grain sizes. With grain sizes of 250 m and above, different combinations of landscapes and metrics became more similar, almost certainly due to the simplification in landscape patterns associated with coarser grain sizes.

Several landscape structure metrics had the ability to discriminate between different landscapes across scales at the county level for both landscape data sources (LCM 2000 and PH1 2000). In particular, a higher proportion of landscape structure metrics derived from the PH1 2000 (in comparison to the LCM 2000) discriminated between landscapes on the basis of the landscape characteristics of mean patch size (MPS), diversity of land covers (DLC) and number of land covers (NLC). A greater variation in landscape structure is therefore captured amongst grid squares by the PH1 2000 landscape structure metrics, and this is because of the higher level of precision associated with this data set in comparison to the broad habitats defined by the LCM 2000. Despite the consistent discrimination between different landscapes with

increases in grain size (up to 250 m), the relationships between metric values for grid square landscapes were inconsistent with grain sizes of 50 m upwards. Although metrics are able to maintain their discriminatory ability with increases in grain size, different patterns are detected within the landscape from a grain size of 50 m and higher due to landscape generalisation.

When identifying the most appropriate grain size for detecting landscape patterns and developing relationships to predict biodiversity as a function of landscape characteristics it is important to consider the grain size that detects landscape patterns within the perceptual range of the species in question (Baguette and Van Dyck, 2007; Tischendorf and Fahrig, 2000). Butterflies have been documented to respond to fine grain landscape patterns with daily movements ranging from 200 – 600 m (Davis, *et al.*, 2007). Due to this limited range in daily dispersal of butterfly species, insufficient levels of detail in landscape data has been identified as a major limiting factor of several landscape based models of butterfly species richness and abundance (Flick, *et al.*, 2012). From Chapter 3 it was found that the smallest grain size of 25 m is most applicable for detecting landscape patterns due to the inconsistencies which occur when using grain-sizes over 50 m in comparison to that at a 25 m to 50 m level. This grain size was also deemed appropriate to detect landscape patterns which occur within the perceptual range of butterfly species.

Utilising landscape data rasterised at a grain size of 25 m Chapter 4 aimed to model butterfly presence-absence as a function of landscape based metrics. Empirical logistic regression models were developed considering the relationship between records of butterfly presence and absence (all species and groups of species based on their ecological attributes) and landscape metrics (composition, connectivity and structure). The grouping of species by their ecological attributes enabled the modelling of multiple species simultaneously. This work improves upon current understandings of butterfly-landscape relationships derived from empirical modelling in the literature. Several studies have either focused on the individual response of species to the landscape or not considered the contrasting response of different species to the landscape in multi-species models (Rossi and van Halder, 2010; Schweiger, *et al.*, 2006; Shreeve and Dennis, 2011). The grouping of species by their ecological attributes accounts for these differing responses, as species with similar

attributes are more likely to respond similarly to their surrounding landscape (Shreeve, *et al.*, 2001). Such an approach also enabled the inference of absence data for the Ecological Attribute Group (EAG) models. This second component of the project (Chapter 4) compared the species-landscape associations, model performance and accuracy between the four species models and two data sources (PH1 2000 and LCM 2000).

Research in Chapter 4 demonstrated that model performance was improved when the landscape metrics were considered together in a combined model. Models based on landscape structure alone were particularly poor in terms of both model fit and model discrimination in comparison to the combined models. This is in agreement with several studies which have demonstrated that structural aspects of the landscape (such as habitat heterogeneity, isolation or patch shape) are important when considered in addition to habitat area (landscape composition) (Dover and Settele, 2009; Ockinger and Smith, 2006; van Halder, *et al.*, 2008). Findings from Chapter 3 demonstrated that several landscape structure metrics can discriminate between landscapes, with variation in landscape structure evident across the grid square landscapes. The fact that very few relationships between structural metrics and butterfly occurrence were included within the combined landscape models, suggests that only specific structural aspects of the landscape influence the distribution of butterflies. Most notably measures of patch shape and aggregation were important in the combined landscape models, and such measures have been found to be important within the literature for influencing population dynamics of a range of species including butterflies (Ewers and Didham, 2007; Saura, *et al.*, 2008; Yamaura, *et al.*, 2008).

Different combinations of landscape parameters were important across the EAG models, reflecting the resource requirements and habitat associations of the different EAGs. For example different combinations of grassland habitats occurred within the four PH1 models, in conjunction with different woodland habitat types which occurred in most models. The importance of woodland habitats for several butterfly species are well documented (Clarke, *et al.*, 2011; van Halder, *et al.*, 2008; Warmington and Vickery, 2003) as is the importance of semi-natural grasslands (Ockinger and Smith, 2006). However, a small number of habitats were specific to

the models for particular attribute groups, for example, only for the classical calcareous grassland ecological attribute group EAG3 (Shreeve, *et al.*, 2001) was calcareous grassland included in the PH1 model. Habitat specificity within models can be expected as butterfly occurrence would have been likely to occur in conjunction with habitats providing their resource requirements (Dennis, 2010; Shreeve and Dennis, 2011; Shreeve, *et al.*, 2001).

Chapter 4 (and Chapter 5) also explored the most appropriate level of precision at which landscape characteristics are recorded for predicting butterfly species distribution by comparing the predictive power of models based on two landscape data sources introduced in Chapter 3 (PH1 2000 and LCM 2000). For both data sources (PH1 2000 and LCM 2000) greatest model accuracy was associated with the EAG3 and EAG1 models. Furthermore, model discrimination of the 'all species' model was comparable to the EAG2 models for both data sources, suggesting that probabilities of butterfly occurrence derived from presence-only data are just as accurate as models based on inferred absences. This contradicts general consensus within the literature, where it is assumed that pseudo-absences based on either inferred absences using the distribution of other species, or from back ground or environmental data, is required to obtain reliable predictions derived from presence-only data (Hirzel, *et al.*, 2006; Schröder, *et al.*, 2009).

Chapter 4 also demonstrates that species-landscape associations within each of the four species models differed between the two data sources (PH1 2000 and LCM 2000), reflecting differences in the thematic resolution of these data types. These findings are in accordance with the literature; it has been widely documented that the classification of land covers/ habitats influences the relationships detected between landscapes and butterfly species distribution (Flick, *et al.*, 2012; Shreeve and Dennis, 2011). Several more compositional parameters were included in the PH1 models in comparison to the LCM models, particularly relating to a combination of grassland habitats. Different and contrasting relationships between butterfly occurrence and landscape structure metrics also occurred between the LCM and PH1 models. Although species-landscape associations were different between the two data sources, they were all considered to be ecologically plausible. The discrimination between presence-absence squares was more accurate, however, for all four models

derived from the PH1 2000 data in comparison to those derived from the LCM 2000 data. Higher model accuracy for the PH1 models is expected due to the higher level of precision associated with this data set, and as such a greater level of detail within the landscape can be detected in terms of landscape composition and configuration. Landscape patterns occurring at fine spatial scales have been found to be directly related to increased butterfly species richness in contrast to coarse landscape patterns (Schneider and Fry, 2001).

To date no study has investigated the influence of different land cover classification systems, widely used in landscape planning and habitat mapping in the UK, on the development of landscape based models for predicting butterfly distribution (or other indicators of biodiversity). Findings from Chapter 4 therefore provide useful inferences on the influences of the LCM broad habitats and PH1 habitats on butterfly occurrence and the predictive performance of models based on metrics describing these habitats.

Landscape models developed in Chapter 4 were validated in Chapter 5 by assessing the ability of the models to correctly predict temporally independent butterfly data (2000 – 2009) (Dormann, *et al.*, 2012). Furthermore, temporally independent PH1 habitat data for 2010 were used to obtain more realistic predicted values. Model validation is essential for drawing conclusions on the variability of the models under different conditions, determining potential problems, and identifying areas for further research (Vaughan and Ormerod, 2005). Predictions of butterfly occurrence based on landscape data from PH1 2010 were compared to observed butterfly data for 2000-2009. Temporally independent data were available for the LCM (LCM 2007); however this could not be used due to inconsistencies in the designation of land cover types between the LCM 2000 and the LCM 2007 (Morton, *et al.*, 2011). As such, predictions of butterfly occurrence derived from the LCM 2000 were compared to observed butterfly data for 2000-2009.

Chapter 5 demonstrates that the PH1 combined models which predict the distribution of ‘all species’, EAG1 and EAG2 species can be transferred temporally, with comparable model discrimination obtained when using the testing and training data. In particular, performance of the PH1-EAG1 model was most accurate in predicting

the distribution of the 2000-2009 butterfly data when considering the AUC test statistic and model specificity and sensitivity. Predictions derived from the LCM 2000 poorly predicted the 2000-2009 butterfly observations in comparison to the PH1 2010 models.

When considering the training data, the LCM and PH1 models which predicted the distribution of EAG3 species were the most accurate in terms of model discrimination. The transferability of the EAG3 models, however, was poor with a decrease in model discrimination and sensitivity when predictions were compared to the observed 2000-2009 butterfly data. The use of EAG3 models for different regions and time periods is therefore impractical, and this could be due to the specific requirements needed for EAG3 species to thrive, and as such the low prevalence of these species across Warwickshire (Shreeve, *et al.*, 2001; Warmington and Vickery, 2003). Low prevalence has been previously identified to be a limiting factor in obtaining accurate predictions from presence-absence models, as it is often difficult to detect strong relationships as a result (Manel, *et al.*, 2001; Santika, 2011; Vaughan and Ormerod, 2005).

The ability of the presence-absence models to predict butterfly species composition, species richness, diversity and abundance (referred to collectively as community characteristics hereon) was also investigated in Chapter 5. Spatially independent butterfly data (no previous records of butterfly presence) collected from a sample of 19 1 km grid squares was used to assess this relationship in addition to the observed butterfly data for 2000-2009. The predictive probabilities of butterfly occurrence from the eight models developed in Chapter 4 were used to select 19 sites which ranged in their suitability and habitat characteristics. Relationships between probabilities of butterfly occurrence and butterfly community characteristics, which occurred between the two validation data sets, are considered to be robust and were specific to each model, with strongest relationships observed with predictions from the PH1 models in comparison to the LCM models.

Predictions obtained from the ‘all species’ models for both LCM and PH1 data sets were associated with overall butterfly species richness, diversity and species composition. Probabilities of occurrence for EAG1 and EAG2 species derived from

the PH1 models were also associated with the abundance of each species group respectively. For the LCM models, relationships were only observed between EAG1 predicted probabilities and EAG1 abundance. When considering the 19 validation sites, those which had similar butterfly community composition also had similar community composition of spring vegetation, suggesting that model predictions reflect local habitat quality.

Chapter 5 demonstrated that probabilities of butterfly occurrence derived from presence-absence models can be related to butterfly characteristics including abundance, diversity, species richness and overall species composition. In turn probabilities of butterfly occurrence can be used to provide an indication of habitat suitability. Predictions from presence-absence models have been found to be successful predictors of butterfly abundance in the literature (Pearce and Ferrier, 2001). For example, Gutierrez *et al.,* (2013) found that distribution models based on presence-absence data for the apollo butterfly *(Parnassius Apollo)* performed better in predicting abundance than quasi-Poisson regression models based on abundance data. Similarly, Guarino *et al.,* (2012) found that presence-absence models of plant species outperformed abundance models, with observed abundance correlated with probabilities of occurrence. The association between probabilities of occurrence and butterfly community characteristics suggests that environmental factors which determine the distribution and community composition of butterfly species are similar. Several studies have found common factors which determine both distribution and abundance, with similar variables selected in both presence-absence and abundance models (Gaston, *et al.*, 2000; Gutierrez, *et al.*, 2013).

Combined Chapters 3, 4 and 5 demonstrate that the PH1 data is the most appropriate level of precision for detecting landscape patterns (Chapter 3), for developing landscape based models of butterfly occurrence (Chapter 4) and for predicting butterfly distribution and abundance in space and time (Chapter 5). However, the ability of the LCM 2000 models to successfully predict butterfly presence-absence for the four species groups cannot be disregarded. Furthermore, the models proved to be transferable, with similar discrimination obtained for the training and testing data and fair discrimination associated with the predictions of EAG1 species. PH1 habitat maps are not readily available across the whole of UK, or even at the county level as

they are labour intensive to produce (Cherrill, *et al.*, 1995; Lucas, *et al.*, 2011). In the absence of PH1 data, it is evident that the LCM models are sufficient for predicting butterfly distribution and community characteristics of EAG1 species.

Chapter 5 demonstrated that the landscape based models (PH1 combined models in particular) are temporally transferable, and therefore an interesting future direction of this work would be to apply the developed landscape models to spatially independent regions across the UK, to further assess their transferability. In particular, application of the models to regions with differing compositional characteristics will enable further development and the identification of additional important habitats not present within Warwickshire. The widespread coverage of the LCM data and the standardisation of the PH1 mapping technique enable the production of LCM and PH1 maps for different areas which are comparable (Cherrill, *et al.*, 1995; Fuller, *et al.*, 2002; Stevens, *et al.*, 2004). As previously mentioned the major disadvantage of PH1 maps is that they are often labour intensive to produce, and as such when spatially transferring the model the LCM has the advantage of availability of land cover data for the whole of the UK (Cherrill, *et al.*, 1995). Furthermore, despite the standardisation of the PH1 mapping technique differences in the classification of habitats between organisations have been documented (Cherrill and McClean, 1999). The LCM models would require further development, however, to consider the broad habitats considered within the LCM 2007 to facilitate the spatial transfer of the models with the most recent land cover data.

The PH1 and LCM landscape-based models would require further development in order to ensure accurate transfer to spatially independent regions. Widespread coverage of butterfly data across the UK is available (Asher, *et al.*, 2001), and the launch of the wider countryside survey and the big butterfly count will improve data coverage for butterfly species in a wide range of habitats (BC, 2014; UKBMS, 2014). The application and use of this data, however, would require further exploration due to potential limitations associated with species detectability bias and potential observer error (Kery and Schmid, 2004; Lele, *et al.*, 2012). Application of the models to different regions considering additional data sets from the UKBMS may also be useful for improving the models for EAG3 species which have low prevalence in Warwickshire.

Another possible future direction of this work could be the application of the modelling approach developed in this thesis to observed data for a range of indicator species. In particular, the grouping of species by their ecological or functional attributes will enable the consideration of multi-species and the inference of absence data where this is not available. The consideration of landscape characteristics (composition, connectivity and structure) will enable the identification of important landscape characteristics for species considered. It is evident from the findings of this thesis, that widespread coverage of butterfly species data, and a reasonable prevalence rate for specific species groups, is required for identifying relationships with landscape characteristics that are transferable. Wide spread species data is available for the taxonomic groups of plants, birds and bats through the NBN, particularly for birds as a result of co-ordinated survey efforts. Consideration of a range of taxonomic groups which differ in dispersal capabilities, and scales at which they perceive their environment will provide an enhanced understanding of the influence of landscape characteristics on biodiversity (Billeter, *et al.*, 2008). The grouping of species by their ecological attributes as developed by Shreeve *et al.*, (2001) and implemented in this study facilitates the consideration of this behavioural response to scale between different species groups. Furthermore, the calculation of habitat connectivity by the IIC metric considers the dispersal capabilities of species capturing the perceptual range of species (Pascual-Hortal and Saura, 2006).

The predicted probabilities of occurrence obtained from the presence-absence models developed in this thesis have been demonstrated to provide appropriate indications of habitat suitability, with relationships developed with butterfly community measurements and in turn local habitat characteristics. Predicted probabilities can therefore be reliably used in the generation of habitat suitability maps across Warwickshire (as developed in Chapter 4 and 5), which can be used to aid the identification of sites which are important for butterfly community assemblage, hence biodiversity (Bayliss, *et al.*, 2005). Such an application of the models highlights their potential for use alongside landscape planning and within ecological consultancy to identify biodiversity hot spots and guide the management of land use developments, associated habitat recommendations and location of sites for habitat creation.

In conclusion this thesis addresses three main areas of uncertainty regarding the development and application of landscape-based models for predicting indicators of biodiversity. Spatial scale (grain size) was shown to influence the patterns detected by landscape structure metrics, and a grain size of 25 m was demonstrated to be most appropriate for discriminating between landscapes and measuring landscape characteristics within the perceptual range of butterfly species. Using butterfly species as biodiversity indicators, the presence-absence of butterfly species groups were successfully predicted as a function of landscape metrics. Landscape metrics (composition, structure and connectivity) were most successful at predicting butterfly presence-absence when they were considered in combination, and the model parameters were demonstrated to have ecologically sound relationships with butterfly species and community characteristics. The successful predictions and transferability of developed models demonstrates their potential for future application in monitoring biodiversity and facilitating targeted conservation efforts.

Appendix

LCM	11	21	41	42	43	51	52	61	71	81	91	101	102	110	121	131	151	161	171	172	181	191	201	211	212	221
11	1.00	0.05	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.11	0.07	0.07	0.00	0.00	0.00	0.00	0.00	0.00
21	0.05	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.06	0.06	0.00	0.01	0.00	0.01	0.01	0.03	0.03	0.00	0.01	0.00	0.00	0.00	0.00
41	0.00	0.00	1.00	1.00	1.00	0.03	0.03	0.01	0.01	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.12	0.12	0.00	0.03	0.00	0.00	0.00	0.00
42	0.00	0.00	1.00	1.00	1.00	0.03	0.03	0.01	0.01	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.12	0.12	0.00	0.03	0.00	0.00	0.00	0.00
43	0.00	0.00	1.00	1.00	1.00	0.03	0.03	0.01	0.01	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.12	0.12	0.00	0.03	0.00	0.00	0.00	0.00
51	0.00	0.00	0.03	0.03	0.03	1.00	1.00	0.14	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.01	0.02	0.02	0.02	0.00
52	0.00	0.00	0.03	0.03	0.03	1.00	1.00	0.14	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.01	0.02	0.02	0.02	0.00
61	0.02	0.00	0.01	0.01	0.01	0.14	0.14	1.00	0.10	0.01	0.00	0.01	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.01	0.01	0.00
71	0.02	0.00	0.01	0.01	0.01	0.02	0.02	0.10	1.00	0.04	0.00	0.01	0.01	0.02	0.00	0.00	0.06	0.10	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
81	0.00	0.01	0.02	0.02	0.02	0.00	0.00	0.01	0.04	1.00	0.05	0.09	0.09	0.01	0.02	0.00	0.01	0.04	0.01	0.01	0.01	0.02	0.00	0.00	0.00	0.00
91	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
101	0.02	0.06	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.09	0.00	1.00	1.00	0.01	0.12	0.00	0.06	0.04	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
102	0.02	0.06	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.09	0.00	1.00	1.00	0.01	0.12	0.00	0.06	0.04	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
110	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.02	0.02	0.01	0.00	0.01	0.01	1.00	0.03	0.13	0.03	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
121	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.12	0.12	0.03	1.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
131	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.13	0.02	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
151	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.01	0.00	0.06	0.06	0.03	0.02	0.00	1.00	0.15	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00
161	0.11	0.01	0.02	0.02	0.02	0.00	0.00	0.01	0.10	0.04	0.01	0.04	0.04	0.02	0.00	0.00	0.15	1.00	0.07	0.07	0.01	0.03	0.00	0.00	0.00	0.00
171	0.07	0.03	0.12	0.12	0.12	0.02	0.02	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	1.00	1.00	0.02	0.06	0.00	0.00	0.00	0.00
172	0.07	0.03	0.12	0.12	0.12	0.02	0.02	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	1.00	1.00	0.02	0.06	0.00	0.00	0.00	0.00
181	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.02	0.02	1.00	0.17	0.06	0.06	0.06	0.00
191	0.00	0.01	0.03	0.03	0.03	0.01	0.01	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.03	0.06	0.06	0.17	1.00	0.06	0.06	0.06	0.00
201	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.06	0.06	1.00	1.00	1.00	0.00
211	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.06	0.06	1.00	1.00	1.00	0.00
212	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.06	0.06	1.00	1.00	1.00	0.00
221	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Appendix A1: Similarity between 26 broad habitats classified by the LCM 2000. Similarities are defined in relation to the similarity in plant community composition between the 26 broad habitats, with 1 representing the highest level of similarity and 0 no similarity. Source: Skrivin and Mead (*pers. comm.)*

PH1	1	2	3	4	5	6	7	8	9	11	12	13	14	15	16	17	18	19	20	22
1	1.0	0.6	0.3	0.1	0.7	0.4	0.4	0.4	0.8	0.1	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.6	1.0	0.4	0.6	0.3	0.8	0.5	0.5	0.9	0.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.3	0.4	1.0	0.6	0.7	0.4	0.2	0.2	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.1	0.6	0.6	1.0	0.3	0.8	0.5	0.5	0.1	0.7	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.7	0.3	0.7	0.3	1.0	0.6	0.4	0.4	0.7	0.1	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.4	0.8	0.4	0.8	0.6	1.0	0.5	0.5	0.3	0.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.4	0.5	0.2	0.5	0.4	0.5	1.0	0.9	0.4	0.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7
8	0.4	0.5	0.2	0.5	0.4	0.5	0.9	1.0	0.2	0.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7
9	0.8	0.9	0.3	0.1	0.7	0.3	0.4	0.2	1.0	0.2	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.1	0.1	0.1	0.7	0.1	0.1	0.2	0.2	0.2	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.6	0.9	0.1	0.1	0.7	0.9	0.4	0.4	0.6	0.0	1.0	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.4	0.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	1.0	0.9	0.8	0.9	0.8	0.9	0.8	0.9	0.0
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.9	1.0	0.9	1.0	0.9	1.0	0.8	0.8	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.8	0.9	1.0	0.9	1.0	0.9	0.8	0.8	0.0
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.9	1.0	0.9	1.0	0.9	1.0	0.8	0.8	0.0
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.8	0.9	1.0	0.9	1.0	0.9	0.8	0.5	0.0
18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.9	1.0	0.9	1.0	0.9	1.0	0.8	0.5	0.0
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.8	0.8	0.8	0.8	0.8	0.8	1.0	0.7	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.9	0.8	0.8	0.8	0.5	0.5	0.7	1.0	0.0
22	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
23	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.6	0.0	0.0	0.7	0.6	0.7	0.8	0.8	0.8	0.8	0.9	0.8	0.8
24	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.6	0.0	0.0	0.4	0.6	0.6	0.6	0.6	0.6	0.6	0.8	0.6	0.9
25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.9	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.7
26	0.3	0.3	0.0	0.0	0.3	0.3	0.3	0.3	0.3	0.0	0.0	0.8	0.8	0.7	0.7	0.6	0.6	0.6	0.9	0.4
27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.6	0.5	0.5	0.0	0.0	0.0	0.9	0.0
28	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.9	0.9	0.9	0.0	0.0	0.0	0.8	0.0
29	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
30	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
32	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.3	0.3	0.3	0.3	0.3	0.4	0.3	0.0
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.2	0.2	0.2	0.4	0.2	0.4	0.4	0.2	0.0
36	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.9	0.8	0.0
37	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.9	0.7	0.0
38	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.8	0.4	0.0
39	0.4	0.5	0.2	0.5	0.4	0.5	0.9	0.9	0.4	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7
40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
41	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.6	0.1	0.6	0.6	0.6	0.6	0.6	0.6	0.0	0.6	0.0
43	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.8	0.7	0.7	0.0	0.0	0.0	0.8	0.0
44	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Appendix A2*(cont.)*

PH1	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	43	44
1	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.0
2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.1	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.0
6	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.1	0.0	0.0
7	0.6	0.6	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.1	0.0	0.0
8	0.6	0.6	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.1	0.0	0.0
9	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0
12	0.7	0.4	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.7	0.7	0.4	0.4	0.0	0.1	0.0	0.0
13	0.6	0.6	0.9	0.8	0.6	0.9	0.0	0.0	0.0	0.0	0.0	0.4	0.2	0.8	0.8	0.4	0.0	0.0	0.6	0.8	0.0
14	0.7	0.6	0.8	0.8	0.6	0.9	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.8	0.8	0.4	0.0	0.0	0.6	0.8	0.0
15	0.8	0.6	0.8	0.7	0.5	0.9	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.8	0.8	0.4	0.0	0.0	0.6	0.7	0.0
16	0.8	0.6	0.8	0.7	0.5	0.9	0.0	0.0	0.0	0.0	0.0	0.3	0.4	0.8	0.8	0.4	0.0	0.0	0.6	0.7	0.0
17	0.8	0.6	0.7	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.8	0.8	0.4	0.0	0.0	0.6	0.0	0.0
18	0.8	0.6	0.7	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.4	0.8	0.8	0.4	0.0	0.0	0.6	0.0	0.0
19	0.9	0.8	0.8	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.9	0.9	0.8	0.0	0.0	0.0	0.0	0.0
20	0.8	0.6	0.8	0.9	0.9	0.8	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.8	0.7	0.4	0.0	0.0	0.6	0.8	0.0
22	0.8	0.9	0.7	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0
23	1.0	0.9	0.7	0.6	0.4	0.7	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.9	0.5	0.8	0.8	0.0	0.0	0.8	0.0
24	0.9	1.0	0.7	0.6	0.4	0.7	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.9	0.5	0.6	0.8	0.0	0.0	0.8	0.0
25	0.7	0.7	1.0	0.9	0.8	0.7	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.7	0.8	0.6	0.0	0.0	0.5	0.8	0.0
26	0.6	0.6	0.9	1.0	0.9	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.5	0.3	0.0	0.0	0.0	0.8	0.0
27	0.4	0.4	0.8	0.9	1.0	0.9	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0
28	0.7	0.7	0.7	0.7	0.9	1.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0
29	0.0	0.0	0.0	0.0	0.3	0.2	1.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
30	0.0	0.0	0.0	0.0	0.2	0.1	0.4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
31	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.9	0.1	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.7	0.0	0.0
32	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	1.0	0.1	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.7	0.0	0.0
33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	1.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.0
34	0.4	0.4	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.5	0.5	0.4	0.4	0.0	0.0	0.0	0.0	0.0
35	0.4	0.4	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.0	0.5	0.4	0.4	0.7	0.0	0.6	0.0	0.0
36	0.9	0.9	0.7	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.5	1.0	0.7	0.6	0.7	0.0	0.6	0.0	0.0
37	0.5	0.5	0.8	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.7	1.0	0.8	0.3	0.0	0.6	0.4	0.0
38	0.8	0.6	0.6	0.3	0.0	0.0	0.0	0.0	0.5	0.3	0.3	0.4	0.4	0.6	0.8	1.0	0.3	0.0	0.8	0.4	0.0
39	0.8	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.7	0.3	0.3	1.0	0.0	0.6	0.0	0.0
40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.3
41	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.7	0.7	0.1	0.0	0.6	0.6	0.6	0.8	0.6	0.0	1.0	0.6	0.7
43	0.8	0.8	0.8	0.8	0.9	0.8	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.0	0.0	0.6	1.0	0.0
44	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.7	0.0	1.0

Appendix A2: Similarity between 44 Phase 1 habitats classified by the PH1 2000. Similarities are defined in relation to the similarity in habitat provision and plant community composition between the 44 Phase 1 habitats, with 1 representing the highest level of similarity and 0 no similarity. Developed in reference to the similarity matrix developed considering the 72 level 3 habitats identified from the LCM 2000 by Skirvin and Mead (*pers. comm.)*.

(a)

	Area	1	2	3	4	5	6	7	8	9	10	11	12	13
1	AREA_AM													
2	AREA_CV	0.988												
3	AREA_MD	-0.155	-0.252											
4	AREA_MN	0.617	0.490	0.406										
5	AREA_RA	0.960	0.964	-0.114	0.515									
6	GYRATE_AM	0.993	0.971	-0.130	0.666	0.931								
7	GYRATE_CV	0.924	0.883	-0.214	0.726	0.802	0.950							
8	GYRATE_MD	-0.167	-0.271	0.970	0.443	-0.123	-0.141	-0.207						
9	GYRATE_MN	-0.172	-0.301	0.856	0.575	-0.166	-0.137	-0.114	0.915					
10	GYRATE_RA	0.917	0.920	-0.058	0.490	0.976	0.891	0.745	-0.071	-0.133				
11	PD	-0.594	-0.467	-0.401	-0.992	-0.484	-0.645	-0.717	-0.445	-0.587	-0.459			
12	LPI	0.897	0.905	-0.319	0.465	0.797	0.893	0.889	-0.332	-0.334	0.712	-0.448		
13	ED	-0.687	-0.577	-0.190	-0.949	-0.550	-0.737	-0.833	-0.248	-0.380	-0.502	0.954	-0.602	
14	LSI	-0.237	-0.192	0.169	-0.373	0.013	-0.292	-0.490	0.151	0.040	0.122	0.393	-0.544	0.519

(b)

	Isolation	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	PROX_AM														
2	PROX_CV	0.515													
3	PROX_MD	-0.519	-0.396												
4	PROX_MN	0.896	0.443	-0.508											
5	PROX_RA	0.863	0.713	-0.476	0.924										
6	SIMI_AM	0.725	0.562	-0.321	0.769	0.812									
7	SIMI_CV	-0.312	0.195	0.264	-0.473	-0.300	-0.094								
8	SIMI_MD	0.692	0.253	-0.445	0.881	0.751	0.756	-0.505							
9	SIMI_MN	0.677	0.458	-0.350	0.806	0.785	0.947	-0.203	0.890						
10	SIMI_RA	0.675	0.560	-0.319	0.704	0.769	0.936	0.096	0.701	0.890					
11	ENN_AM	-0.352	-0.239	-0.049	-0.522	-0.503	-0.601	0.059	-0.553	-0.636	-0.617				
12	ENN_CV	-0.234	-0.285	0.118	-0.121	-0.140	-0.152	-0.257	0.012	-0.067	-0.084	0.160			
13	ENN_MD	0.554	0.127	-0.676	0.572	0.463	0.241	-0.455	0.510	0.285	0.222	0.098	-0.105		
14	ENN_MN	0.491	0.137	-0.632	0.569	0.491	0.394	-0.540	0.578	0.441	0.349	0.038	0.274	0.778	
15	ENN_RA	-0.064	-0.118	0.099	0.014	0.045	0.035	-0.262	0.089	0.075	0.066	0.066	0.914	-0.030	0.260

(c)

	Contrast	1	2	3	4	5	6	7	8	9	10
1	CWED										
2	TECI	0.596									
3	ECON_AM	0.586	0.961								
4	ECON_CV	-0.573	-0.881	-0.794							
5	ECON_MD	0.459	0.892	0.794	-0.744						
6	ECON_MN	0.498	0.945	0.845	-0.890	0.963					
7	ECON_RA	0.065	-0.006	-0.028	-0.033	-0.041	-0.030				
8	PRD	-0.062	-0.152	-0.011	0.205	-0.108	-0.152	-0.383			
9	RPR	-0.045	-0.266	-0.226	0.331	-0.308	-0.351	0.108	0.028		
10	SIDI	0.650	0.168	0.131	-0.132	0.151	0.116	-0.028	-0.211	0.141	
11	SIEI	0.669	0.206	0.164	-0.180	0.191	0.163	-0.040	-0.211	0.003	0.990

(d)

	Aggregation	1	2	3	4	5	6
1	CONTAG						
2	IJI	-0.847					
3	MESH	0.731	-0.440				
4	SPLIT	-0.712	0.407	-0.896			
5	AI	0.772	-0.480	0.683	-0.689		
6	CONNECT	0.139	-0.022	0.019	-0.454	0.266	
7	COHESION	0.763	-0.483	0.971	-0.919	0.739	0.147

Appendix A3 ***(cont.)***

(e)

Shape	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 SHAPE_AM																			
2 SHAPE_CV	0.856																		
3 SHAPE_MD	-0.436	-0.410																	
4 SHAPE_MN	-0.421	-0.242	0.928																
5 SHAPE_RA	0.842	0.679	-0.330	-0.365															
6 FRAC_AM	0.979	0.892	-0.437	-0.376	0.763														
7 FRAC_CV	-0.107	0.149	0.088	0.265	-0.252	-0.017													
8 FRAC_MD	-0.458	-0.444	0.942	0.892	-0.321	-0.450	0.186												
9 FRAC_MN	-0.460	-0.384	0.950	0.950	-0.339	-0.435	0.204	0.978											
10 FRAC_RA	0.720	0.596	-0.312	-0.350	0.966	0.629	-0.249	-0.284	-0.304										
11 CIRCLE_AM	0.753	0.749	-0.322	-0.211	0.540	0.780	0.369	-0.309	-0.299	0.449									
12 CIRCLE_CV	0.186	0.334	-0.312	-0.166	-0.070	0.251	0.811	-0.293	-0.277	-0.114	0.570								
13 CIRCLE_MD	-0.281	-0.265	0.769	0.759	-0.249	-0.248	0.364	0.791	0.776	-0.263	0.028	0.032							
14 CIRCLE_MN	-0.260	-0.214	0.748	0.769	-0.273	-0.211	0.433	0.766	0.765	-0.294	0.103	0.110	0.983						
15 CIRCLE_RA	0.026	-0.116	0.102	0.073	0.091	-0.007	0.272	0.178	0.155	0.106	0.191	0.243	0.206	0.238					
16 CONTIG_AM	0.607	0.678	-0.098	0.066	0.259	0.712	0.236	-0.182	-0.116	0.100	0.700	0.434	0.150	0.211	-0.060				
17 CONTIG_CV	-0.029	0.082	-0.361	-0.279	-0.200	0.025	0.763	-0.219	-0.257	-0.192	0.312	0.803	-0.015	0.019	0.248	0.067			
18 CONTIG_MD	-0.086	-0.038	0.441	0.435	-0.105	-0.109	-0.329	0.160	0.240	-0.168	-0.185	-0.238	0.136	0.147	-0.194	0.228	-0.620		
19 CONTIG_MN	-0.064	-0.107	0.473	0.420	0.013	-0.104	-0.633	0.251	0.314	-0.021	-0.311	-0.627	0.121	0.103	-0.207	0.064	-0.904	0.872	
20 CONTIG_RA	-0.030	0.070	0.390	0.492	-0.069	0.053	0.348	0.419	0.439	-0.064	0.253	0.192	0.492	0.508	0.228	0.470	0.074	0.076	0.009

Appendix A3: Pearson Product Moment Correlation coefficients between landscape structure metrics derived from the LCM 2000 for the 32 National Character Areas (NCAs) grouped by landscape aspect: (a) area metrics (b) isolation metrics (aggregation) (c) contrast metrics (d) patch aggregation metrics and (e) shape metrics. Correlations coefficients shaded in grey indicate strongly significant correlations ($p<0.01$) with coefficients >0.6 or <-0.6.

Metric	Acronym	Units (range)	Component	Scaling relation	Description
Area metrics					*Represent the number/ density of patches, average size and variation in patch size*
Mean Patch Size	AREA_MN	ha (>0)	Composition	Type I	**Mean area** of all patches within the landscape, irrespective of class type, providing a measure of habitat **fragmentation**.
Range in Patch Size	AREA_RA	ha	Composition	Unassessed	Difference between the maximum and minimum patch size in the landscape.
Mean Radius of Gyration	GYRATE_MN	m (≥ 0)	Composition	Unassessed	Measure of **patch extent**, calculated as the mean distance between each cell in the patch and the patch centroid, averaged across all patches in the landscape.
Area-weighted Mean Radius of Gyration	GYRATE_AM	m (≥ 0)	Composition	Unassessed	Measure of **landscape continuity** (correlation length). Mean radius of gyration after weighting patches according to size, larger patches weighted more heavily.
Radius of Gyration Coefficient of Variation	GYRATE_CV	**%**	Composition	Unassessed	Relative variability about the mean radius of gyration.
Shape metrics					*Capture the complexity of patch shape*
Mean Shape Index	SHAPE_MN	None (≥1)	Configuration	Type III	Measures **complexity** of patch shape compared to standard square shape (maximally compact patch). Patch shape is averaged across all patches in the landscape.
Shape Index Coefficient of Variation	SHAPE_CV	%	Configuration	Unassessed	Relative variability about the mean shape index.
Mean Related Circumscribing Circle	CIRCLE_MN	None (0-1)	Configuration	Unassessed	Measure of patch **elongation**. The ratio of patch area to the area of the smallest circumscribing circle around the patch, averaged across all patches in the landscape.

Appendix A4 *(cont.)*

Metric	Acronym	Units (range)	Component	Scaling relation	Description
Area-weighted Mean Related Circumscribing Circle	CIRCLE_AM	None (0-1)	Configuration	Unassessed	Mean circumscribing circle index after weighting patches according to size, larger patches weighted more heavily.
Range in Related Circumscribing Circle	CIRCLE_RA	None (0-1)	Configuration	Unassessed	Difference between the maximum and minimum circle index in the landscape.
Area-weighted Mean Fractal Dimension Index	FRAC_AM	None (1-2)	Configuration	Type I	Index of shape **complexity** where patch perimeter is related to patch area. Mean patch fractal dimension after weighting patches according to size, larger patches weighted more heavily.
Fractal Dimension Coefficient of Variation	FRAC_CV	%	Configuration	Unassessed	Relative variability about the mean fractal dimension.
Mean Contiguity Index	CONTIG_MN	None (0-1)	Configuration	Unassessed	Measures the **spatial connectedness** of **cells** within a patch providing an index of patch boundary configuration. Contiguity index is averaged across all patches in the landscape.
Area-weighted Mean Contiguity Index	CONTIG_AM	None (0-1)	Configuration	Unassessed	Mean contiguity index after weighting patches according to size, larger patches weighted more heavily.
Range in Contiguity Index	CONTIG_RA	None (0-1)	Configuration	Unassessed	Difference between the maximum and minimum contiguity index in the landscape.
Contrast metrics					*Degree of edge contrast between adjacent patch types*
Area-weighted Mean Edge Contrast Index	ECON_AM	% (0-100)	Configuration	Unassessed	**Relative** measure of the % of the patch perimeter which is in contrast with its neighbourhood. Each segment of the patch perimeter is weighted by the **degree of contrast** with the adjacent patch. This index is averaged across all patches in the landscape, after weighting patches according to size, larger patches weighted more heavily.

Appendix A4 ***(cont.)***

Metric	Acronym	Units (range)	Component	Scaling relation	Description
Edge Contrast Index coefficient of variation	ECON_CV	%	Configuration	Unassessed	Relative variability about the mean edge contrast index.
Contrast weighted edge density	CWED	m/ha (0>)	Configuration	Unassessed	Standardises edge to a per unit area, incorporating both **edge density and edge contrast**, and reduces length of each edge segment proportionate to the degree of contrast.
Aggregation metrics					*Represent the dispersion, interspersion, subdivision and isolation of patch types.*
Contagion index	CONTAG	% (0-100)	Configuration	Type III	Summary of clumpiness measuring patch type **interspersion and dispersion**. Measures the extent to which patch types are aggregated by measuring how individual cells are arranged relative to each other. It is the probability that two random adjacent **cells** belong to patch type i and j.
Interspersion and juxtaposition index	IJI	% (0-100)	Configuration	Unassessed	Based on patch adjacencies, measures extent to which **patch types** are **interspersed** (not necessarily dispersed) as a percentage of maximum possible given total number of patch types.
Landscape Shape Index	LSI	None (≥1)	Configuration	Type I	Measure of **dispersion and complexity**, based on the perimeter-to-area ratio for the landscape as a whole. Provides a standardized measure of **edge density**.
Patch Cohesion Index	COHESION	None	Configuration	Unassessed	Measures **dispersion** and compaction of patches incorporating **patch area and perimeter**. Takes into consideration the number of patches, distribution and shape of patches.

Appendix A4 *(cont.)*

Metric	Acronym	Units (range)	Component	Scaling relation	Description
Effective Mesh Size	MESH	ha (cell size – TA)	Configuration	Unassessed	Measure of landscape **subdivision** independent of landscape size. Based on the cumulative **patch area** distribution, provides the area-weighted mean patch size based on total landscape area.
Mean Euclidean Nearest-Neighbour Distance	ENN_MN	m (>0)	Configuration	Unassessed	Measure of **patch isolation**, averaged across all patches in the landscape. Shortest **straight-line distance** from a patch to the nearest neighbouring patch of same type, based on edge to edge distance for patches comprising the class. Considers only patches that have neighbours.
Area-weighted Mean Euclidean Nearest-Neighbour Distance	ENN_AM	m (>0)	Configuration	Unassessed	Mean nearest neighbor distance after weighting patches according to size, larger patches weighted more heavily
Euclidean Nearest-Neighbour Distance co-efficient of variation	ENN_CV	%	Configuration	Unassessed	Measures relative variability about the mean Euclidean Nearest-Neighbour Distance
Area-weighted Mean Proximity Index	PROX_AM	None (≥ 0)	Configuration	Unassessed	Spatial context of a habitat patch in relation to its neighbours, considering the **size and proximity** distance of patches of the same patch type whose edges are within specified **search radius** of the focal patch. Average proximity index for all patches in the landscape.
Proximity Index co-efficient of variation	PROX_CV	%	Configuration	Unassessed	Measures relative variability about the mean Proximity Index
Area-weighted Mean Similarity Index	SIMI_AM	None (≥ 0)	Configuration	Unassessed	Incorporates landscape mosaic, quantifying the spatial context of a habitat patch in relation to neighbours of **similar class type**. Considers the **size and proximity** distance of patches regardless of patch type, whose edges are within specified search radius of the focal patch.

Appendix A4 ***(cont.)***

Metric	Acronym	Units (range)	Component	Scaling relation	Description
Similarity Index co-efficient of variation	SIMI_CV	%	Configuration	Unassessed	Measures relative variability about the Mean Similarity Index
Diversity Metrics					*Quantify diversity of patch types, incorporating richness and evenness*
Simpson's diversity index	SIDI	None (0-1)	Composition	Unassessed	Measure of class type diversity, incorporating **number and proportion** of class types in the landscape. It is the probability that any cells selected at random would belong to different class types.
Patch richness density (#/100ha)	PRD	No/100 ha (>0)	Composition	Type II	**Number of class types** present in the landscape standardized by landscape area.

Appendix A4:. Landscape structure metrics defined and summarised at the landscape level, ordered by associated landscape aspect (adapted from McGarigal and Marks 1995; 2002; 2014). The landscape component which each metric captures is provided in addition to the behavior of the metric in relation to increasing grain size as identified by Wu *et al.*, (2002): Type I – predictable response with simple scaling relations; Type II – staircase-like response with no simple scaling relation; Type III – erratic response with no scaling relation. The units and range for each metric are provided (ha – hectares; m – meters; % - percentage) in addition to associated acronyms which are used in the text.

Common name	Latin name
Alder	*Alnus glutinosa*
Alder buckthorn	*Frangula alnus*
Annual meadow-grass	*Poa annua*
Ash	*Fraxinus excelsior*
Aspen	*Populus tremula*
Beech	*Fagus sylvatica*
Betony	*Stachys officinalis*
Bird cherry	*Prunus padus*
Bittersweet	*Solanum dulcamara*
Black bryony	*Tamus communis*
Black horehound	*Ballota nigra*
Black medick	*Medicago lupulina*
Black nightshade	*Solanum nigrum*
Black-poplar	*Populus nigra*
Blackthorn	*Prunus spinosa*
Bluebell	*Hyacinthoides non-scripta*
Bracken	*Pteridium aquilinum*
Bramble	*Rubus fruticosus agg.*
Broad-leaved dock	*Rumex obtusifolius*
Broad-leaved willowherb	*Epilobium montanum*
Buckthorn	*Rhamnus cathartica*
Bugle	*Ajuga reptans*
Bulrush	*Typha latifolia*
Butterfly-bush	*Buddleia davidii*
Cat's-ear	*Hypochaeris radicata*
Charlock	*Sinapis arvensis*
Cherry Laurel	*Prunus laurocerasus*
Cleavers	*Galium aparine*
Cock's-foot	*Dactylis glomerata*
Common ash	*Fraxinus excelsior*
Common bent	*Agrostis capillaris*
Common Bird's-foot-trefoil	*Lotus corniculatus*
Common centaury	*Centaurium erythraea*
Common comfrey	*Symphytum officinale*
Common couch	*Elytrigia repens*
Common Evening-primrose	*Oenothera biennis*
Common field-speedwell	*Veronica persica*
Common ivy	*Hedera helix subsp. helix*
Common Mallow	*Malva sylvestris*
Common meadow-rue	*Thalictrum flavum*
Common nettle	*Urtica dioica*
Common poppy	*Papaver rhoeas*

Appendix A5 ***(cont.)***

Common name	Latin name
Common ragwort	*Senecio jacobaea*
Common sedge	*Carex nigra*
Compact rush	*Juncus conglomeratus*
Cow parsley	*Anthriscus sylvestris*
Cowslip	*Primula veris*
Crab apple	*Malus sylvestris*
Creeping buttercup	*Ranunculus repens*
Creeping cinquefoil	*Potentilla reptans*
Creeping soft-grass	*Holcus mollis*
Creeping thistle	*Cirsium arvense*
Cuckooflower	*Cardamine pratensis*
Daffodil	*Narcissus pseudonarcissus subsp. pseudonarcissus*
Daisy	*Bellis perennis*
Dandelion	*Taraxacum officinale agg.*
Devil's-bit scabious	*Succisa pratensis*
Dog-rose	*Rosa canina*
Dog's mercury	*Mercurialis perennis*
Dove's-foot crane's-bill	*Geranium molle*
Downy birch	*Betula pubescens*
Elder	*Sambucus nigra*
Enchanter's-nightshade	*Circaea lutetiana*
Eyebright	*Euphrasia officinalis agg.*
False oat-grass	*Arrhenatherum elatius*
False-brome	*Brachypodium sylvaticum*
Fat-hen	*Chenopodium album*
Fennel	*Foeniculum vulgare*
Field bindweed	*Convolvulus arvensis*
Field forget-me-not	*Myosotis arvensis*
Field horsetail	*Equisetum arvense*
Field Maple	*Acer campestre*
Field pansy	*Viola arvensis*
Field penny-cress	*Thlaspi arvense*
Garden parsley	*Petroselinum crispum*
Garden privet	*Ligustrum ovalifolium*
Garden Tulip	*Tulipa gesneriana*
Garlic mustard	*Alliaria petiolata*
Goat willow	*Salix caprea*
Gorse	*Ulex europaeus*
Greater bird's-foot-trefoil	*Lotus pedunculatus*
Greater burdock	*Arctium lappa*

Appendix A5 ***(cont.)***

Common name	Latin name
Greater Plantain	*Plantago major*
Greater stitchwort	*Stellaria holostea*
Greater tussock-sedge	*Carex paniculata*
Green alkanet	*Pentaglottis sempervirens*
Ground-elder	*Aegopodium podagraria*
Ground-ivy	*Glechoma hederacea*
Groundsel	*Senecio vulgaris*
Hairy sedge	*Carex hirta*
Hard rush	*Juncus inflexus*
Hawthorn	*Crataegus monogyna*
Hazel	*Corylus avellana*
Heath bedstraw	*Galium saxatile*
Heather	*Calluna vulgaris*
Hedge bindweed	*Calystegia sepium*
Hedge mustard	*Sisymbrium officinale*
Hedge woundwort	*Stachys sylvatica*
Hedgerow crane's-bill	*Geranium pyrenaicum*
Herb-robert	*Geranium robertianum*
Himalayan balsam	*Impatiens glandulifera*
Hoary plantain	*Plantago media*
Hogweed	*Heracleum sphondylium*
Holly	*Ilex aquifolium*
Honeysuckle	*Lonicera periclymenum*
Horse-chestnut	*Aesculus hippocastanum*
Hydrangea	*Hydrangea macrophylla*
Ivy	*Hedera helix*
Knotgrass	*Polygonum aviculare*
Lady-fern	*Athyrium filix-femina*
Lady's bedstraw	*Galium verum*
Lady's-mantle	*Alchemilla vulgaris agg.*
Lemon-scented fern	*Oreopteris limbosperma*
Lesser burdock	*Arctium minus*
Lesser celandine	*Ranunculus ficaria*
Lesser clubmoss	*Selaginella selaginoides*
Lesser stitchwort	*Stellaria graminea*
Lords-and-ladies	*Arum maculatum*
Marsh foxtail	*Alopecurus geniculatus*
Marsh thistle	*Cirsium palustre*
Meadow buttercup	*Ranunculus acris*
Meadow crane's-bill	*Geranium pratense*

Appendix A5 ***(cont.)***

Common name	Latin name
Meadow fescue	*Festuca pratensis*
Meadow foxtail	*Alopecurus pratensis*
Meadow thistle	*Cirsium dissectum*
Meadow vetchling	*Lathyrus pratensis*
Musk-mallow	*Malva moschata*
Narrow-leaved Michaelmas-daisy	*Aster lanceolatus*
Nipplewort	*Lapsana communis*
Norway spruce	*Picea abies*
Orange balsam	*Impatiens capensis*
Oxeye daisy	*Leucanthemum vulgare*
Pale willowherb	*Epilobium roseum*
Pedunculate Oak	*Quercus robur*
Perennial rye-grass	*Lolium perenne*
Perforate St John's-wort	*Hypericum perforatum*
Pheasant's-eye	*Adonis annua*
Pignut	*Conopodium majus*
Prickly sow-thistle	*Sonchus asper*
Primrose	*Primula vulgaris*
Ramsons	*Allium ursinum*
Red campion	*Silene dioica*
Red clover	*Trifolium pratense*
Red dead-nettle	*Lamium purpureum*
Red valerian	*Centranthus ruber*
Redshank	*Persicaria maculosa*
Ribwort plantain	*Plantago lanceolata*
Rosebay willowherb	*Chamerion angustifolium*
Rough chervil	*Chaerophyllum temulum*
Rough meadow-grass	*Poa trivialis*
Rowan	*Sorbus aucuparia*
Scarlet pimpernel	*Anagallis arvensis subsp. arvensis*
Scentless mayweed	*Tripleurospermum inodorum*
Scots pine	*Pinus sylvestris*
Selfheal	*Prunella vulgaris*
Sessile Oak	*Quercus petraea*
Sheep's sorrel	*Rumex acetosella*
Shepherd's-purse	*Capsella bursa-pastoris*
Shining crane's-bill	*Geranium lucidum*
Short-fruited willowherb	*Epilobium obscurum*
Silver Birch	*Betula pendula*
Silverweed	*Potentilla anserina*

Appendix A5 ***(cont.)***

Common name	Latin name
Slender speedwell	*Veronica filiformis*
Smooth cat's-ear	*Hypochaeris glabra*
Smooth hawk's-beard	*Crepis capillaris*
Smooth sow-thistle	*Sonchus oleraceus*
Spear thistle	*Cirsium vulgare*
Sweet vernal-grass	*Anthoxanthum odoratum*
Sweet violet	*Viola odorata*
Sycamore	*Acer pseudoplatanus*
Timothy	*Phleum pratense*
Tormentil	*Potentilla erecta*
Trailing St John's-wort	*Hypericum humifusum*
Tufted forget-me-not	*Myosotis laxa*
Tufted hair-grass	*Deschampsia cespitosa subsp. cespitosa*
Tufted vetch	*Vicia cracca*
Upright hedge-parsley	*Torilis japonica*
Water dock	*Rumex hydrolapathum*
Wavy hair-grass	*Deschampsia flexuosa*
White bryony	*Bryonia dioica*
White clover	*Trifolium repens*
White dead-nettle	*Lamium album*
Wild cherry	*Prunus avium*
Wild pansy	*Viola tricolor*
Wild teasel	*Dipsacus fullonum*
Wood anemone	*Anemone nemorosa*
Wood avens	*Geum urbanum*
Wood dock	*Rumex sanguineus*
Wood forget-me-not	*Myosotis sylvatica*
Wood meadow-grass	*Poa nemoralis*
Wood sage	*Teucrium scorodonia*
Wood-sedge	*Carex sylvatica*
Wood-sorrel	*Oxalis acetosella*
Wych Elm	*Ulmus glabra*
Yellow Archangel	*Lamiastrum galeobdolon*
Yellow pimpernel	*Lysimachia nemorum*
Yorkshire-fog	*Holcus lanatus*

Appendix A5: Plant species identified from the vegetation sampling conducted across the 19 sample sites in Warwickshire. Common name and latin name are provided. See section 2.3.4 for vegetation sampling methodology.

NCA	25_50m	25_100m	25_250m	25_500m	25_1000m
7	0.05	0.09	0.06	0.10	0.05
41	0.10	0.16	0.16	0.12	0.17
42	0.07	0.09	0.08	0.11	0.12
45	0.05	0.08	0.07	0.09	0.08
52	0.04	0.09	0.09	0.08	0.05
58	0.05	0.06	0.12	0.11	0.13
59	0.01	0.04	0.05	0.07	0.13
62	0.06	0.08	0.08	0.09	0.09
64	0.06	0.05	0.07	0.08	0.07
66	0.03	0.03	0.07	0.09	0.09
69	0.01	0.03	0.03	0.04	0.10
70	0.01	0.03	0.05	0.05	0.07
73	0.03	0.06	0.05	0.02	0.07
78	0.01	0.04	0.06	0.06	0.09
80	0.01	0.03	0.09	0.13	0.17
85	0.03	0.05	0.13	0.14	0.15
96	0.05	0.04	0.04	0.05	0.08
97	0.02	0.03	0.07	0.11	0.12
109	0.02	0.05	0.01	0.06	0.10
119	0.06	0.05	0.04	0.02	0.04
120	0.04	0.02	0.05	0.05	0.13
125	0.04	0.05	0.07	0.08	0.10
128	0.03	0.03	0.06	0.06	0.08
131	0.03	0.04	0.05	0.03	0.04
132	0.03	0.03	0.05	0.06	0.08
134	0.04	0.03	0.07	0.08	0.15
138	0.06	0.06	0.07	0.11	0.16
142	0.06	0.06	0.06	0.08	0.08
145	0.02	0.06	0.06	0.08	0.05
147	0.01	0.03	0.04	0.06	0.08
150	0.03	0.02	0.06	0.09	0.04
153	0.06	0.06	0.06	0.08	0.07

Metric	Code	25_50m	25_100m	25_250m	25_500m	25_1000m
AREA_MN	1	0.06	0.08	0.07	0.10	0.12
AREA_RA	2	0.02	0.03	0.04	0.04	0.06
CIRCLE_AM	3	0.06	0.12	0.11	0.13	0.12
CIRCLE_MN	4	0.05	0.12	0.09	0.11	0.13
CIRCLE_RA	5	0.04	0.07	0.12	0.12	0.12
COHESION	6	0.03	0.03	0.04	0.05	0.06
CONTAG	7	0.02	0.05	0.07	0.07	0.07
CONTIG_AM	8	0.07	0.08	0.07	0.12	0.13
CONTIG_MN	9	0.09	0.22	0.20	0.21	0.22
CONTIG_RA	10	0.06	0.06	0.18	0.19	0.19
CWED	11	0.04	0.05	0.06	0.09	0.12
ECON_AM	12	0.03	0.05	0.03	0.01	0.07
ECON_CV	13	0.05	0.08	0.06	0.07	0.12
ENN_AM	14	0.07	0.09	0.15	0.11	0.12
ENN_CV	15	0.11	0.04	0.18	0.14	0.16
ENN_MN	16	0.05	0.06	0.10	0.13	0.14
FRAC_AM	17	0.02	0.03	0.05	0.06	0.07
FRAC_CV	18	0.18	0.28	0.26	0.29	0.28
GYRATE_AM	19	0.01	0.04	0.05	0.05	0.06
GYRATE_CV	20	0.02	0.03	0.05	0.05	0.06
GYRATE_MN	21	0.05	0.08	0.07	0.08	0.10
IJI	22	0.08	0.05	0.10	0.07	0.08
LSI	23	0.01	0.04	0.09	0.10	0.09
MESH	24	0.01	0.05	0.06	0.07	0.07
PRD	25	0.03	0.06	0.11	0.15	0.16
PROX_AM	26	0.04	0.02	0.04	0.04	0.12
PROX_CV	27	0.05	0.12	0.16	0.19	0.18
SHAPE_CV	28	0.03	0.06	0.07	0.10	0.09
SHAPE_MN	29	0.06	0.08	0.07	0.08	0.09
SIDI	30	0.01	0.03	0.05	0.06	0.06
SIMI_AM	31	0.06	0.04	0.08	0.08	0.10
SIMI_CV	32	0.04	0.05	0.08	0.11	0.12

Appendix A6: Procrustes residuals obtained when comparing the 4-D configuration of the (a) NCA scores and (b) metric loadings for a grain size of 25 m with grain sizes of 50 m, 100 m, 250 m, 500 m and 1000 m. The range in colours represent the size of the Procrustes residual with green representing small residuals and red representing high residuals.

(a)

		LCM 2000		
Metric	Code	25_50m	25_100m	25_250m
AREA_MN	1	0.02	0.06	0.07
AREA_RA	2	0.01	0.09	0.04
CIRCLE_AM	3	0.01	0.06	0.05
CIRCLE_MN	4	0.02	0.05	0.07
CIRCLE_RA	5	0.05	0.14	0.12
COHESION	6	0.01	0.04	0.05
CONTAG	7	0.04	0.05	0.07
CONTIG_AM	8	0.03	0.04	0.08
CONTIG_MN	9	0.03	0.11	0.10
CONTIG_RA	10	0.07	0.25	0.24
CWED	11	0.03	0.11	0.07
ECON_AM	12	0.02	0.14	0.05
ECON_CV	13	0.04	0.09	0.11
ENN_AM	14	0.03	0.05	0.25
ENN_CV	15	0.02	0.23	0.09
ENN_MN	16	0.02	0.04	0.31
FRAC_AM	17	0.02	0.07	0.10
FRAC_CV	18	0.05	0.09	0.10
GYRATE_AM	19	0.01	0.06	0.05
GYRATE_CV	20	0.02	0.04	0.06
GYRATE_MN	21	0.02	0.06	0.11
IJI	22	0.06	0.06	0.11
LSI	23	0.02	0.06	0.06
MESH	24	0.01	0.09	0.05
PRD	25	0.05	0.06	0.05
PROX_AM	26	0.09	0.09	0.08
PROX_CV	27	0.04	0.07	0.20
SHAPE_CV	28	0.03	0.02	0.08
SHAPE_MN	29	0.03	0.08	0.12
SIDI	30	0.03	0.03	0.07
SIMI_AM	31	0.03	0.09	0.04
SIMI_CV	32	0.02	0.05	0.06

(b)

		PH1 2010		
Metric	Code	25_50m	25_100m	25_250m
AREA_MN	1	0.02	0.02	0.04
AREA_RA	2	0.03	0.04	0.10
CIRCLE_AM	3	0.06	0.08	0.14
CIRCLE_MN	4	0.02	0.05	0.06
CIRCLE_RA	5	0.08	0.09	0.09
COHESION	6	0.06	0.07	0.08
CONTAG	7	0.07	0.09	0.11
CONTIG_AM	8	0.02	0.03	0.05
CONTIG_MN	9	0.02	0.05	0.09
CONTIG_RA	10	0.06	0.10	0.10
CWED	11	0.02	0.02	0.04
ECON_AM	12	0.03	0.03	0.05
ECON_CV	13	0.05	0.07	0.23
ENN_AM	14	0.04	0.05	0.06
ENN_CV	15	0.07	0.10	0.14
ENN_MN	16	0.06	0.11	0.16
FRAC_AM	17	0.17	0.17	0.14
FRAC_CV	18	0.05	0.06	0.09
GYRATE_AM	19	0.06	0.07	0.09
GYRATE_CV	20	0.05	0.07	0.10
GYRATE_MN	21	0.03	0.04	0.07
IJI	22	0.09	0.13	0.13
LSI	23	0.03	0.04	0.06
MESH	24	0.03	0.04	0.09
PRD	25	0.03	0.04	0.09
PROX_AM	26	0.23	0.26	0.28
PROX_CV	27	0.05	0.07	0.14
SHAPE_CV	28	0.05	0.05	0.08
SHAPE_MN	29	0.03	0.05	0.08
SIDI	30	0.08	0.10	0.15
SIMI_AM	31	0.05	0.08	0.19
SIMI_CV	32	0.04	0.05	0.06

Appendix A7: Procrustes residuals obtained when comparing the 4-D configuration of the metric loadings for a grain size of 25 m with grain sizes of 50 m, 100 m and 250 m for the two landscape data sources; (a) LCM 2000 and (b) PH1 2000. the range in colours represent the size of the Procrustes residual with green representing small residuals and red representing high residuals.

Metric	Acronym	Units (range)	Component	Description
Area metrics				*Number/ density of patches, average size and variation*
Area-weighted Mean Patch Size	AREA_AM	ha (>0)	Composition	**Area weighted mean patch size** considering all patches within the landscape, irrespective of class type, providing a measure of habitat **fragmentation**
Range in Radius of Gyration	GYRATE_RA	m (≥ 0)	Composition	Difference between the maximum and minimum patch extent.
Shape metrics				*Capture the complexity of patch shape*
Area-weighted Mean Shape Index	SHAPE_AM	None (≥1)	Configuration	Measures **complexity** of patch shape compared to standard square shape (maximally compact patch). Patch shape is averaged across all patches in the landscape, after weighting patches according to size, larger patches weighted more heavily.
Shape Index Standard Deviation	SHAPE_SD	None	Configuration	Absolute variability around the mean shape index.
Contiguity Index Standard Deviation	CONTIG_SD	None	Configuration	Absolute variability around the mean contiguity index which measures the **spatial connectedness** of **cells** within a patch.
Contrast metrics				*Degree of edge contrast between adjacent patch types*
Mean Edge Contrast Index	ECON_MN	% (0-100)	Configuration	**Relative** measure of the % of the patch perimeter which is in contrast with its neighbourhood. Each segment of the patch perimeter is weighted by the **degree of contrast** with the adjacent patch. This index is averaged across all patches in the landscape.
Range in Edge Contrast Index	ECON_RA	% (0-100)	Configuration	Difference between the maximum and minimum edge contrast index.
Edge Contrast Standard Deviation	ECON_SD	%	Configuration	Absolute variability about the mean edge contrast index.

Appendix A8 ***(cont.)***

Metric	Acronym	Units (range)	Component	Description
Aggregation metrics				*Represent the dispersion, interspersion, subdivision and isolation of patch types.*
Connectance Index	CONNECT	% (0-100)	Configuration	**Functional connectivity** between patches based on specified **threshold distance**, measured as the percentage of maximum possible connections.
Euclidean Nearest-Neighbour Distance Standard Deviation	ENN_SD	m (>0)	Configuration	Absolute variability about the mean Euclidean nearest neighbour index which measures the straight line distance between patches of the same patch type.
Mean Proximity Index	PROX_MN	None (≥ 0)	Configuration	Spatial context of a habitat patch in relation to its neighbours, considering the **size and proximity** distance of patches of the same patch type whose edges are within specified **search radius** of the focal patch. Average proximity index for all patches in the landscape.
Median Proximity Index	PROX_MD	% (≥ 0)	Configuration	Mid proximity index value based on the distribution of proximity index values for all patches in the landscape.
Mean Similarity Index	SIMI_MN	None (≥ 0)	Configuration	Incorporates landscape mosaic, quantifying the spatial context of a habitat patch in relation to neighbours of **similar class type**. Considers the **size and proximity** distance of patches regardless of patch type, whose edges are within specified search radius of the focal patch.
Median Similarity Index	SIMI_MD	% (≥ 0)	Configuration	Mid similarity index value based on the distribution of similarity index values for all patches in the landscape.
Range in Similarity Index	SIMI_RA	% (≥ 0)	Configuration	Difference between the maximum and minimum similarity index.

Appendix A8: Additional landscape structure metrics considered during development of landscape based models. Landscape structure metrics are defined and summarised at the landscape level, and are ordered by associated landscape aspect (adapted from McGarigal and Marks 1995; 2002; 2014). The landscape component which each metric captures is provided in addition to the units and range for each metric (ha – hectares; m – meters; % - percentage) and the associated acronyms which are used in the text.

Phase 1 Habitat	PH1 2010	Biodiversity Action Plan Broad Habitat	LCM 2000
Broad-leaved semi-natural woodland	1	Broadleaved, mixed and yew woodland	11
Broad-leaved plantation	2	Broadleaved, mixed and yew woodland	11
Coniferous semi-natural woodland	3	Coniferous woodland	21
Coniferous plantation	4	Coniferous woodland	21
Mixed semi-natural woodland	5	Broadleaved, mixed and yew woodland	11
Mixed plantation	6	Broadleaved, mixed and yew woodland	11
Dense/continuous scrub	7	Broadleaved, mixed and yew woodland	11
Recently felled woodland	11	Broadleaved, mixed and yew woodland/ Coniferous woodland	11/21
Orchard (commercial)	12	Arable and horticultural/ Broadleaved, mixed and yew woodland	43, 11
Unimproved acidic grassland	13	Acid grassland	81
semi-improved acidic grassland	14	Acid grassland	81
Unimproved neutral grassland	15	Neutral grassland	61
Semi-improved neutral grassland	16	Neutral grassland	61
Unimproved calcareous grassland	17	Calcareous grassland	71
semi-improved calcareous grassland	18	Calcareous grassland	71
Improved grassland	19	Improved grassland	51
Marsh/marshy grassland	20	Neutral grassland Fen, marsh and swamp	61 111
Poor semi-improved grassland	21	Improved grassland	51
Continuous bracken	22	Bracken	91
Dry heath/acid grassland mosaic	25	Acid grassland Dwarf shrub heath	81 101
Acid/neutral flush	26	Fen, marsh and swamp	111
Swamp	27	Fen, marsh and swamp	111
Inundation vegetation	28	Standing open water and canals Rivers and streams	131 131
Standing water	29	Standing open water and canals	131
Running water	30	Rivers and streams	131
Quarry	31	Inland rock	161
Spoil	32	Inland rock	161
Refuse tip	33	Built up areas and gardens	171
Arable	34	Arable creals, horticulture and non-rotational	41/ 42/ 43
Allotments	35	Built up areas and gardens	171
Set-aside	36	Set-aside grassland	52
Amenity grassland	37	Improved grassland	51
Ephemeral/short perennial	38	Built up areas and gardens	19
Introduced shrub	39	Broadleaved, mixed and yew woodland	11
Buildings	40	Built up areas and gardens	172
Bare ground	41	Inland rock	161
Fen - basin mire	43	Fen, marsh and swamp	111

Appendix A9: Correspondence between PH1 2010 habitats and LCM 2000 broad habitats. The habitats scattered scrub (PH-8), broad-leaved parkland/ scattered trees (PH-9), Coniferous parkland/ scattered trees (PH-10), Tall ruderal (PH-23), Non-ruderal (PH-24), road/infrastructure (PH-44).

References

Aavik, T. and Liira, J. (2009). Agrotolerant and high nature-value species-Plant biodiversity indicator groups in agroecosystems. *Ecological Indicators* **9**(5), pp. 892-901.

Allouche, O., Tsoar, A. and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**(6), pp. 1223-1232.

Araujo, M.B., Pearson, R.G., Thuiller, W. and Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology* **11**(9), pp. 1504-1513.

ArcGIS. (2011). *Polygon to Raster (Conversion)*. [Online]. (http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//001200000030000000). (26/06/14)

ARCH. (2011). *ARCH Kent Wildlife Habitat Survey*. [Online]. (http://www.archnature.eu/news-item-6.html). (26/04/14)

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G. and Jeffcoate, S. (2001). *The Millennium Atlas of Butterflies in Britain and Ireland*. Oxford University Press: New York, USA.

Aurambout, J.P., Endress, A.G. and Deal, B.M. (2005). A spatial model to estimate habitat fragmentation and its consequences on long-term persistence of animal populations. *Environmental Monitoring and Assessment* **109**(1-3), pp. 199-225.

Azeria, E.T., Fortin, D., Hebert, C., Peres-Neto, P., Pothier, D. and Ruel, J.C. (2009). Using null model analysis of species co-occurrences to deconstruct biodiversity patterns and select indicator species. *Diversity and Distributions* **15**(6), pp. 958-971.

Baguette, M., Petit, S. and Queva, F. (2000). Population spatial structure and migration of three butterfly species within the same habitat network: consequences for conservation. *Journal of Applied Ecology* **37**(1), pp. 100-108.

Baguette, M. and Van Dyck, H. (2007). Landscape connectivity and animal behavior: functional grain as a key determinant for dispersal. *Landscape Ecology* **22**(8), pp. 1117-1129.

Bailey, D., Billeter, R., Aviron, S., Schweiger, O. and Herzog, F. (2007). The influence of thematic resolution on metric selection for biodiversity monitoring in agricultural landscapes. *Landscape Ecology* **22**(3), pp. 461-473.

Bailey, S. (2007). Increasing connectivity in fragmented landscapes: An investigation of evidence for biodiversity gain in woodlands. *Forest Ecology and Management* **238**(1-3), pp. 7-23.

Bailey, S.A., Haines-Young, R.H. and Watkins, C. (2002). Species presence in fragmented landscapes: modelling of species requirements at the national level. *Biological Conservation* **108**(3), pp. 307-316.

Baldwin, D.J.B., Weaver, K., Schnekenburger, F. and Perera, A.H. (2004). Sensitivity of landscape pattern indices to input data characteristics on real landscapes: implications for their use in natural disturbance emulation. *Landscape Ecology* **19**(3), pp. 255-271.

Barr, C.J., Bunce, R.G.H., Clarke, R.T., Fuller, R.M., Furse, M.T., Gillespie, M.K., Groom, G.B., Hallam, C.J., Hornung, M., Howard, D.C. and Ness, M.J. (1993). *Countryside Survey 1990: main report. (Countryside 1990 vol.2).*

Bayliss, J.L., Simonite, V. and Thompson, S. (2005). The use of probabilistic habitat suitability models for biodiversity action planning. *Agriculture Ecosystems & Environment* **108**(3), pp. 228-250.

Begon, M., Townsend, C.R. and Harper, J.L. (2006). *Ecology: From individuals to ecosystems*. 4th ed. Blackwell Publishing Ltd: Oxford, UK.

Bergerot, B., Merckx, T., Van Dyck, H. and Baguette, M. (2012). Habitat fragmentation impacts mobility in a common and widespread woodland butterfly: do sexes respond differently? *Bmc Ecology* **12**pp. 10.

Bergman, K.O. (2001). Population dynamics and the importance of habitat management for conservation of the butterfly Lopinga achine. *Journal of Applied Ecology* **38**(6), pp. 1303-1313.

Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek, R., Burel, F., Cerny, M., De Blust, G., De Cock, R., Diekötter, T., Dietz, H., Dirksen, J., Dormann, C., Durka, W., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., Le Coeur, D., Maelfait, J.P., Opdam, P., Roubalova, M., Schermann, A., Schermann, N., Schmidt, T., Schweiger, O., Smulders, M., Speelmans, M., Simova, P., Verboom, J., Van Wingerden, W., Zobel, M. and Edwards, P. (2008). Indicators for biodiversity in agricultural landscapes: a pan-European study. *Journal of Applied Ecology* **45**(1), pp. 141-150.

Blamires, D., de Oliveira, G., Barreto, B.D. and Diniz, J.A.F. (2008). Habitat use and deconstruction of richness patterns in Cerrado birds. *Acta Oecologica-International Journal of Ecology* **33**(1), pp. 97-104.

Bräuniger, C., Knapp, S., Kühn, I. and Klotz, S. (2010). Testing taxonomic and landscape surrogates for biodiversity in an urban setting. *Landscape and Urban Planning* **97**(4), pp. 283-295.

Bullock, J.M., Jefferson, R.G., Blackstock, T.H., Pakeman, R.J., Emmett, B.A., Pywell, R.J., Grime, J.P. and Silvertown, J. (2011). *Semi-natural grasslands*. p. 162‑195,
Technical Report: The UK National Ecosystem Assessment. UNEP‑WCMC, Cambridge, UK.

Bulte, E., Hector, A. and Larigauderie, A. (2005). *ecoSERVICES: Assessing the impacts of biodiversity changes on ecosystem functioning and services.* DIVERSITAS.

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vie, J.C. and Watson, R. (2010). Global Biodiversity: Indicators of Recent Declines. *Science* **328**(5982), pp. 1164-1168.

Butterfly Conservation. (2014). *Big Butterfly Count.* [Online]. (http://www.bigbutterflycount.org/about). (26/06/14).

Büchs, W. (2003). Biodiversity and agri-environmental indicators--general scopes and skills with special reference to the habitat level. *Agriculture, Ecosystems & Environment* **98**(1-3), pp. 35-78.

Cabeza, M., Araujo, M.B., Wilson, R.J., Thomas, C.D., Cowley, M.J.R. and Moilanen, A. (2004). Combining probabilities of occurrence with spatial reserve design. *Journal of Applied Ecology* **41**(2), pp. 252-262.

Castilla, G., Larkin, K., Linke, J. and Hay, G.J. (2009). The impact of thematic resolution on the patch-mosaic model of natural landscapes. *Landscape Ecology* **24**(1), pp. 15-23.

Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C. and Diaz, S. (2000). Consequences of changing biodiversity. *Nature* **405**(6783), pp. 234-242.

Cherrill, A. and McClean, C. (1999). The reliability of 'Phase 1' habitat mapping in the UK: the extent and types of observer bias. *Landscape and Urban Planning* **45**(2-3), pp. 131-143.

Cherrill, A.J., McClean, C., Lane, A. and Fuller, R.M. (1995). A comparison of land-cover types in an ecological field survey in northern England and a remotely-sensed land-cover map of Great-Britain. *Biological Conservation* **71**(3), pp. 313-323.

Christensen, N.L., Bartuska, A.M., Brown, J.H., Carpenter, S., Dantonio, C., Francis, R., Franklin, J.F., MacMahon, J.A., Noss, R.F., Parsons, D.J., Peterson, C.H., Turner, M.G. and Woodmansee, R.G. (1996). The report of the ecological society of America committee on the scientific basis for ecosystem management. *Ecological Applications* **6**(3), pp. 665-691.

Clarke, K.R. (1993). Non-parametri multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18**(1), pp. 117-143.

Clarke, S.A., Green, D.G., Bourn, N.A. and Hoare, D.J. (2011). *Woodland Management for butterflies and moths: a best practice guide.* Dorset: Butterfly Conservation.

Cowley, M.J.R., Wilson, R.J., Leon-Cortes, J.L., Gutierrez, D., Bulman, C.R. and Thomas, C.D. (2000). Habitat-based statistical models for predicting the spatial distribution of butterflies and day-flying moths in a fragmented landscape. *Journal of Applied Ecology* **37**pp. 60-72.

Cozzi, G., Muller, C.B. and Krauss, J. (2008). How do local habitat management and landscape structure at different spatial scales affect fritillary butterfly distribution on fragmented wetlands? *Landscape Ecology* **23**(3), pp. 269-283.

Croxton, P.J., Hann, J.P., Greatorex-Davies, J.N. and Sparks, T.H. (2005). Linear hotspots? The floral and butterfly diversity of green lanes. *Biological Conservation* **121**(4), pp. 579-584.

Dauber, J., Hirsch, M., Simmering, D., Waldhardt, R., Otte, A. and Wolters, V. (2003). Landscape structure as an indicator of biodiversity: matrix effects on species richness. *Agriculture, Ecosystems & Environment* **98**(1-3), pp. 321-329.

Davis, J.D., Debinski, D.M. and Danielson, B.J. (2007). Local and landscape effects on the butterfly community in fragmented Midwest USA prairie habitats. *Landscape Ecology* **22**(9), pp. 1341-1354.

Defra. (2011a). The natural choice: securing the value of nature. London, UK, Department for Environment, Food and Rural Affairs.

Defra. (2011b). Biodiversity 2020: A strategy for England's wildlife and ecosystem services. London, UK, Department for Environment, Food and Rural Affairs.

Defra. (2012). UK Biodiversity Indicators in Your Pocket 2012 London, UK, Department for Environment, Food and Rural Affairs.

Defra. (2013). Biodiversity 2020: a strategy for England's wildlife and ecosystem services Indicators 2013 London, UK, Department for Environment, Food and Rural Affairs.

Dennis, R. (2009). Ecology of butterflies in Europe - where we are now and where to go? In: Settele, J., Shreeve, T., Konvicka, M. and Van Dyck, H. *Ecology of Butterflies in Europe*, pp. 1-5. Cambridge: Cambridge University Press.

Dennis, R.L.H. (2004). Butterfly habitats, broad-scale biotope affiliations, and structural exploitation of vegetation at finer scales: the matrix revisited. *Ecological Entomology* **29**(6), pp. 744-752.

Dennis, R.L.H. (2010). *A resource-based habitat view for conservation: butterflies in the British landscape*. United Kingdom: Wiley-Blackwell.

Dennis, R.L.H. and Hardy, P.B. (2007). Support for mending the matrix: resource seeking by butterflies in apparent non-resource zones. *Journal of Insect Conservation* **11**(2), pp. 157-168.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Romermann, C., Schroder, B. and Singer, A. (2012). Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography* **39**(12), pp. 2119-2131.

Dover, J. and Settele, J. (2009). The influences of landscape structure on butterfly distribution and movement: a review. *Journal of Insect Conservation* **13**(1), pp. 3-27.

Dover, J., Sparks, T., Clarke, S., Gobbett, K. and Glossop, S. (2000). Linear features and butterflies: the importance of green lanes. *Agriculture Ecosystems & Environment* **80**(3), pp. 227-242.

Dover, J.W. and Sparks, T.H. (2001). Green lanes: biodiversity reservoirs in farmland? In: Barr, C. and Petit, S. *Hedgerows of the World: their ecological functions in different landscapes*, pp. 241-250. IALE(UK): Lymm.

Dunning, J.B., Danielson, B.J. and Pulliam, H.R. (1992). Ecological Proccesses that affect populations in complex landscapes. *Oikos* **65**(1), pp. 169-175.

EEA-FOEN. (2011). Landscape fragmentation in Europe Joint EEA-FOEN Report. p. 92, vol. 2. European Environment Agency/Federal Office for the Environment, Copenhagen.

Ehrlich, P.R. and Wilson, E. (1991). Biodiversity Studies: Science and Policy. *Science* **253**(5021), pp. 758-762.

Emmet, A.M. and Heath, J. (1990). *The Butterflies of Great Britain and Ireland: Hesperiidae-Nymphalidae - Vol 7, part1 of The Moths and Butterflies of Great Britain & Ireland*. Harley Books: Colchester.

England Biodiversity Group. (2011). *Think Big: How and why landscape-scale conservation benefits wildlife, people and the wider economy*. *Natural*

England Report (309). [Online]. (http://publications.naturalengland.org.uk/publication/30047). (26/06/14).

Erhardt, A. and Mevi-schutz, J. (2009). Adult food resources in butterflies. In: Settele, J., Shreeve, T., Konvicka, M. and Van Dyck, H. *Ecology of Butterflies in Europe*, pp. 9-16. Cambridge: Cambridge University Press.

Evans, K.L., Bradbury, R.B. and Wilson, J.D. (2003). Selection of hedgerows by Swallows Hirundo rustica foraging on farmland: the influence of local habitat and weather. *Bird Study* **50**pp. 8-14.

Ewers, R.M. and Didham, R.K. (2006). Confounding factors in the detection of species responses to habitat fragmentation. *Biological Reviews* **81**(1), pp. 117-142.

Ewers, R.M. and Didham, R.K. (2007). The effect of fragment shape and species' sensitivity to habitat edges on animal population size. *Conservation Biology* **21**(4), pp. 926-936.

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology Evolution and Systematics* **34**pp. 487-515.

Failing, L. and Gregory, R. (2003). Ten common mistakes in designing biodiversity indicators for forest policy. *Journal of Environmental Management* **68**(2), pp. 121-132.

Falk, S. (2003). *Warwickshire Local BAP – Quarries and gravel pits*. [Online]. (http://apps.warwickshire.gov.uk/api/documents/WCCC-863-179). (26/06/14)

Falk, S., J. (2009). *Warwickshire's wildflowers: The wildflowers, shrubs and trees of historic Warwickshire.* Warwickshire: Brewin Books Ltd.

Feld, C.K., da Silva, P.M., Sousa, J.P., de Bello, F., Bugter, R., Grandin, U., Hering, D., Lavorel, S., Mountford, O., Pardo, I., Partel, M., Rombke, J., Sandin, L., Jones, K.B. and Harrison, P. (2009). Indicators of biodiversity and ecosystem services: a synthesis across ecosystems and spatial scales. *Oikos* **118**(12), pp. 1862-1871.

Fielding, A.H. and Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**(1), pp. 38-49.

Firbank, L.G., Petit, S., Smart, S., Blain, A. and Fuller, R.J. (2008). Assessing the impacts of agricultural intensification on biodiversity: a British perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1492), pp. 777-787.

Fischer, J. and Lindenmayer, D.B. (2007). Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography* **16**(3), pp. 265-280.

Fjellstad, W.J., Dramstad, W.E., Strand, G.-H. and Fry, G.L.A. (2001). Heterogeneity as a measure of spatial pattern for monitoring agricultural landscapes. *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography* **55**(2), pp. 71 - 76.

Fleishman, E., Mac Nally, R. and Fay, J.P. (2003). Validation tests of predictive models of butterfly occurrence based on environmental variables. *Conservation Biology* **17**(3), pp. 806-817.

Flick, T., Feagan, S. and Fahrig, L. (2012). Effects of landscape structure on butterfly species richness and abundance in agricultural landscapes in eastern Ontario, Canada. *Agriculture Ecosystems & Environment* **156**pp. 123-133.

Fortin, M.J., Boots, B., Csillag, F. and Remmel, T.K. (2003). On the role of spatial stochastic models in understanding landscape indices in ecology. *Oikos* **102**(1), pp. 203-212.

Fox, R., Brereton, T.M., Asher, J., Botham, M.S., Middlebrook, I., Roy, D.B. and Warren, M.S. (2011). The State of the UK's Butterflies 2011 Wareham, Dorset, Butterfly Conservation and the Centre for Ecology & Hydrology.

Franklin, A.B., Noon, B.R. and George, T.L. (2002). What is habitat fragmentation? *Effects of Habitat Fragmentation on Birds in Western Landscapes: Contrasts with Paradigms from the Eastern United States*(25), pp. 20-29.

Fuller, R.M. (1987). The changing extent and conservation interest of lowland grasslands in England and Wales – A review of grassland surveys. *Biological Conservation* **40**(4), pp. 281-300.

Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G., Cox, R., Brown, N.J., Clarke, R.T., Rothery, P. and Gerard, F.F. (2002). Countryside Survey 2000 Module 7: Land Cover Map 2000 Final Report. Cambridgeshire, UK, Centre for Ecology and Hydrology, Natural Environment Research Council.

Garcia-Barros, E. and Fartmann, T. (2009). Butterfly oviposition: sites, behaviour and modes. In: Settele, J., Shreeve, T., Konvicka, M. and Van Dyck, H. *Ecology of Butterflies in Europe*, pp. 29-42. Cambridge: Cambridge University Press.

Garcia-Feced, C., Saura, S. and Elena-Rossello, R. (2010). Assessing the effect of scale on the ability of landscape structure metrics to discriminate landscape types in Mediterranean forest districts. *Forest Systems* **19**(2), pp. 129-140.

Gaston, K., J., Blackburn, T.M., Greenwood, J.J.D., Gregory, R.D., Quinn, R.M. and Lawton, J.H. (2000). Abundance-occupancy relationships. *Journal of Applied Ecology* **37**(1), pp. 39-59.

Gaucherel, C. (2007). Multiscale heterogeneity map and associated scaling profile for landscape analysis. *Landscape and Urban Planning* **82**(3), pp. 95-102.

Gil-Tena, A., Lecerf, R. and Ernoult, A. (2012). Disentangling community assemblages to depict an indicator of biological connectivity: A regional study of fragmented semi-natural grasslands. *Ecological Indicators* **24**(0), pp. 48-55.

Gillings, S. and Fuller, R.J. (1998). Changes in bird populations on sample lowland English farms in relation to loss of hedgerows and other non-crop habitats. *Oecologia* **116**(1-2), pp. 120-127.

Goodwin, B.J. (2003). Is landscape connectivity a dependent or independent variable? *Landscape Ecology* **18**(7), pp. 687-699.

Goodwin, B.J. and Fahrig, L. (2002). How does landscape structure influence landscape connectivity? *Oikos* **99**(3), pp. 552-570.

Gregory, R.D., van Strien, A., Vorisek, P., Gmelig Meyling, A.W., Noble, D.G., Foppen, R.P.B. and Gibbons, D.W. (2005). Developing indicators for European birds. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**(1454), pp. 269-288.

Gross, K., Kalendra, E.J., Hudgens, B.R. and Haddad, N.M. (2007). Robustness and uncertainty in estimates of butterfly abundance from transect counts. *Population Ecology* **49**(3), pp. 191-200.

Guarino, E.D.G., Barbosa, A.M. and Waechter, J.L. (2012). Occurrence and abundance models of threatened plant species: Applications to mitigate the impact of hydroelectric power dams. *Ecological Modelling* **230**pp. 22-33.

Guisan, A. and Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* **135**(2-3), pp. 147-186.

Gustafson, E.J. (1998). Quantifying Landscape Spatial Pattern: What Is the State of the Art? *Ecosystems* **1**(2), pp. 143-156.

Gutierrez, D., Harcourt, J., Diez, S.B., Illan, J.G. and Wilson, R.J. (2013). Models of presence-absence estimate abundance as well as (or even better than) models of abundance: the case of the butterfly Parnassius apollo. *Landscape Ecology* **28**(3), pp. 401-413.

Haila, Y. (2002). A conceptual genealogy of fragmentation research: From island biogeography to landscape ecology. *Ecological Applications* **12**(2), pp. 321-334.

Haines-Young, R. (2009). Land use and biodiversity relationships. *Land Use Policy* **26**pp. S178-S186.

HainesYoung, R. and Chopping, M. (1996). Quantifying landscape structure: A review of landscape indices and their application to forested landscapes. *Progress in Physical Geography* **20**(4), pp. 418-445.

Hanski, I. (1994). A practical model of metapopulation dynamics. *Journal of Animal Ecology* **63**(1), pp. 151-162.

Harrison, S. and Bruna, E. (1999). Habitat fragmentation and large-scale conservation: what do we know for sure? *Ecography* **22**(3), pp. 225-232.

Heikkinen, R.K., Luoto, M., Kuussaari, M. and Toivonen, T. (2007). Modelling the spatial distribution of a threatened butterfly: Impacts of scale and statistical technique. *Landscape and Urban Planning* **79**(3-4), pp. 347-357.

Heink, U. and Kowarik, I. (2010). What criteria should be used to select biodiversity indicators? *Biodiversity and Conservation* **19**(13), pp. 3769-3797.

Hilty, J. and Merenlender, A. (2000). Faunal indicator taxa selection for monitoring ecosystem health. *Biological Conservation* **92**(2), pp. 185-197.

Hinsley, S.A. and Bellamy, P.E. (2000). The influence of hedge structure, management and landscape context on the value of hedgerows to birds: A review. *Journal of Environmental Management* **60**(1), pp. 33-49.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C. and Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* **199**(2), pp. 142-152.

Hoffman, H. and Marktanner, T. (1995). *Butterflies and moths of Britain and Europe*. London: HarperCollinsPublishers.

Jackson, D.L. (2000). *Guidance on the interpretation of the Biodiversity Broad Habitat Classification (terrestrial and freshwater types): Definitions and the relationship with other classifications*.

JNCC. (2010a). *Handbook for Phase 1 habitat survey - a technique for environmental audit*. [Online]. (http://jncc.defra.gov.uk/page-2468).

JNCC. (2010b). *The UK Biodiversity Action Plan: Highlights from the 2008 reporting round*. [Online]. (http://jncc.defra.gov.uk/pdf/pub2010_UKBAPHighlightsReport2008.pdf). (24/06/14)

JNCC. (2014). *Joint Nature Conservation Committe: UK surveillance schemes*. [Online]. (http://jncc.defra.gov.uk/page-3721). (26/06/14)

Jongman, R.H.G., Ter Braak, C.J.F. and Van Tongeren, O.F.R. (1995). *Data Analysis in Community and Landscape Ecology*. Cambridge: Cambridge University Press.

Kadoya, T. and Washitani, I. (2011). The Satoyama Index: A biodiversity indicator for agricultural landscapes. *Agriculture Ecosystems & Environment* **140**(1-2), pp. 20-26.

Kennedy, C.E.J. and Southwood, T.R.E. (1984). The number of species of insects associated wth British trees- a re-analysis. *Journal of Animal Ecology* **53**(2), pp. 455-478.

Kery, M. and Schmid, H. (2004). Monitoring programs need to take into account imperfect species detectability. *Basic and Applied Ecology* **5**(1), pp. 65-73.

Kindlmann, P. and Burel, F. (2008). Connectivity measures: a review. *Landscape Ecology* **23**(8), pp. 879-890.

Kirby, K.J., Smart, S.M., Black, H.I.J., Bunce, R.G.H., Corney, P.M. & Smithers, R.J. (2005). *Long-Term Ecological Change in British Woodland*. English Nature Research Report no. 653.

Kumar, S., Simonson, S.E. and Stohlgren, T.J. (2009). Effects of spatial heterogeneity on butterfly species richness in Rocky Mountain National Park, CO, USA. *Biodiversity and Conservation* **18**(3), pp. 739-763.

Lawler, J., Wiersma, Y. and Huettmann, F. (2011). *Using Species Distribution Models for Conservation Planning and Ecological Forecasting*. New York: Springer New York.

Lawton, J.H., Brotherton, P.N.M., Brown, V.K., Elphick, C., Fitter, A.H., Forshaw, J., Haddow, R.W., Hilborne, S., Leafe, R.N., Mace, G.M., Southgate, M.P., Sutherland, W.A., Tew, T.E., Varley, J. and Wynne, G.R. (2010). Making Space for Nature: a review of England's wildlife sites and ecological network. p. 107. Defra.

Lele, S.R., Moreno, M. and Bayne, E. (2012). Dealing with detection error in site occupancy surveys: what can we do with a single survey? *Journal of Plant Ecology* **5**(1), pp. 22-31.

Levins, R. (1969). Some deomographic and genetic consequences of enviornmental heterogenity for biological control. *Bulletin of the Entomological Society of America* **15**(3), pp. 237-240.

Lewis, T. (1969). The distribution of insects near a low hedgerow. *Journal of Applied Ecology* **6**(3), pp. 443-452.

Li, H. and Wu, J. (2004). Use and misuse of landscape indices. *Landscape Ecology* **19**(4), pp. 389-399.

Li, X., He, H.S., Bu, R., Wen, Q., Chang, Y., Hu, Y. and Li, Y. (2005). The adequacy of different landscape metrics for various landscape patterns. *Pattern Recognition* **38**(12), pp. 2626-2638.

Liley, D., Brereton, T. and Roy, D. (2004). *The Current Level of Butterfly Monitoring in UK Woodlands. Butterfly Conservation Report to the Forestry CommissionReport no. SO4-35*. [Online]. (http://www.forestry.gov.uk/pdf/BC-ButterfliesForestBioIndicatorsRep.pdf/$FILE/BC-ButterfliesForestBioIndicatorsRep.pdf). (26/06/14)

Lindenmayer, D.B. and Likens, G.E. (2011). Direct Measurement Versus Surrogate Indicator Species for Evaluating Environmental Change and Biodiversity Loss. *Ecosystems* **14**(1), pp. 47-59.

Lindenmayer, D.B., Margules, C.R. and Botkin, D.B. (2000). Indicators of biodiversity for ecologically sustainable forest management. *Conservation Biology* **14**(4), pp. 941-950.

Lobo, J.M., Jimenez-Valverde, A. and Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**(2), pp. 145-151.

Lucas, R., Medcalf, K., Brown, A., Bunting, P., Breyer, J., Clewley, D., Keyworth, S. and Blackmore, P. (2011). Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *Isprs Journal of Photogrammetry and Remote Sensing* **66**(1), pp. 81-102.

Luoto, M., Heikkinen, R.K., Poyry, J. and Saarinen, K. (2006). Determinants of the biogeographical distribution of butterflies in boreal regions. *Journal of Biogeography* **33**(10), pp. 1764-1778.

Mac Nally, R., Fleishman, E., Thomson, J.R. and Dobkin, D.S. (2008). Use of guilds for modelling avian responses to vegetation in the Intermountain West (USA). *Global Ecology and Biogeography* **17**(6), pp. 758-769.

MacArthur, R. and Wilson, E. (1967). *The theory of Island Biogeography*. Princeton: Princeton University Press.

Maddock, A. (2008). UK Biodiversity Action Plan Priority Habitat Descriptions: Lowland Meadows. JNCC, UK.

Manel, S., Williams, H.C. and Ormerod, S.J. (2001). Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* **38**(5), pp. 921-931.

Marquet, P.A., Fernandez, M., Navarrete, S.A. and Valdovinos, C. (2004). *Diversity Emerging: Toward a deconstruction of biodiversity patterns*. Sunderland: Sinauer Assoc.

Maurer, B.A. and McGill, B.J. (2011). *Measurement of species diversity*. United States: Oxford University Press.

McGarigal, K. (2014). *FRAGSTATS metrics.* [Online]. (http://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.2.pdf). (24/06/14)

McGarigal, K., Cushman, S. and Ene, E. (2012). *FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst*. [Online]. (http://www.umass.edu/landeco/research/fragstats/fragstats.html). (24/06/14)

McGarigal, K. and Marks, B.J. (1995). FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure. *U S Forest Service General Technical Report PNW* **0**(351), pp. I-122.

Merckx, T. and Van Dyck, H. (2007). Habitat fragmentation affects habitat-finding ability of the speckled wood butterfly, Pararge aegeria L. *Animal Behaviour* **74**pp. 1029-1037.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R. and Simpson, I.C. (2011). Final Report for LCM2007 - the new UK Land Cover Map p. 113. Centre for Ecology and Hydrology, Natural Environment Research Council, Cambridgeshire, UK.

Munguira, M., L, Garcia-Barros, E. and Martin Cano, J. (2009). Butterfly herbivory and larval ecology. In: Settele, J., Shreeve, T., Konvicka, M. and Van Dyck, H. *Ecology of Butterflies in Europe*, pp. 43-54. Cambridge: Cambridge University Press.

Natural England. (2012a) *National Character Areas – defining England's natural boundaries.* [Online]. (http://www.naturalengland.org.uk/publications/nca/default.aspx). (26/06/14).

Natural England. (2012b) *NCA Profile: 97 Arden (NE337).* [Online]. (http://publications.naturalengland.org.uk/publication/1819400). (26/06/14).

Natural England. (2012c) *96 Dunsmore and Feldon.* [Online]. (http://www.naturalengland.org.uk/publications/nca/dunsmore_and_feldon.aspx). (26/06/14).

NBN. (2011). *The National Biodiversity Network*. [Online]. (http://www.nbn.org.uk/The-NBN/What-is-the-NBN.aspx). (26/06/14)

Nielsen, S.E., Johnson, C.J., Heard, D.C. and Boyce, M.S. (2005). Can models of presence-absence be used to scale abundance? - Two case studies considering extremes in life history. *Ecography* **28**(2), pp. 197-208.

Noss, R.F. (1990). Indicators for Monitoring Biodiversity: A Hierarchical Approach. *Conservation Biology* **4**(4), pp. 355-364.

Nowicki, P., Settele, J., Henry, P.Y. and Woyciechowski, M. (2008). Butterfly monitoring methods: The ideal and the real world. *Israel Journal of Ecology & Evolution* **54**(1), pp. 69-88.

O'Neill, R.V., Krummel, J.R., Gardner, R.H., Sugihara, G., Jackson, B., Deangelis, D.L., Milne, B.T., Turner, M.G., Zygmunt, B. and et al. (1988). Indices of landscape pattern. *Landscape Ecology* **1**(3), pp. 153-162.

Ockinger, E. and Smith, H.G. (2006). Landscape composition and habitat area affects butterfly species richness in semi-natural grasslands. *Oecologia* **149**(3), pp. 526-534.

Ockinger, E. and Van Dyck, H. (2012). Landscape Structure Shapes Habitat Finding Ability in a Butterfly. *Plos One* **7**(8), pp. 6.

Ouin, A., Aviron, S., Dover, J. and Burel, F. (2004). Complementation/supplementation of resources for butterflies in agricultural landscapes. *Agriculture Ecosystems & Environment* **103**(3), pp. 473-479.

Pascual-Hortal, L. and Saura, S. (2006). Comparison and development of new graph-based landscape connectivity indices: towards the priorization of habitat patches and corridors for conservation. *Landscape Ecology* **21**(7), pp. 959-967.

Pearce, J. and Ferrier, S. (2001). The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation* **98**(1), pp. 33-43.

Peng, J., Wang, Y., Zhang, Y., Wu, J., Li, W. and Li, Y. (2010). Evaluating the effectiveness of landscape metrics in quantifying spatial patterns. *Ecological Indicators* **10**(2), pp. 217-223.

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M. and Sexton, J.O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science (New York, N.Y.)* **344**(6187), pp. 1246752.

Pollard, E. (1977). Method for assessing changes in abundance of butterflies. *Biological Conservation* **12**(2), pp. 115-134.

Pywell, R.F., Warman, E.A., Sparks, T.H., Greatorex-Davies, J.N., Walker, K.J., Meek, W.R., Carvell, C., Petit, S. and Firbank, L.G. (2004). Assessing habitat quality for butterflies on intensively managed arable farmland. *Biological Conservation* **118**(3), pp. 313-325.

Quine, C., Cahalan, C., Hester, A., Humphrey, J., Kirby, K., Moffat, A. and Valatin, G. (2011). Woodlands. p. 241-291. Technical Report: The UK National Ecosystem Assessment. UNEP‑WCMC, Cambridge, UK.

Quine, C.P. and Watts, K. (2009). Successful de-fragmentation of woodland by planting in an agricultural landscape? An assessment based on landscape indicators. *Journal of Environmental Management* **90**(1), pp. 251-259.

Rackham, O. (1986). The History of the Countryside J.M. Dent, London.

Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A. and Peterson, A.T. (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature* **426**(6968), pp. 837-841.

Riitters, K.H., Oneill, R.V., Hunsaker, C.T., Wickham, J.D., Yankee, D.H., Timmins, S.P., Jones, K.B. and Jackson, B.L. (1995). A factor analysis of landscape pattern and structure metrics. *Landscape Ecology* **10**(1), pp. 23-39.

Riley, A.M. (2007). *British and Irish Butterflies: The complete Identification, Field and Site Guide to the Species, Subspecies and Forms*. London: Brambleby Books.

Robinson, N., Kadlec, T., Bowers, M.D. and Guralnick, R.P. (2014). Integrating species traits and habitat characteristics into models of butterfly diversity in a fragmented ecosystem. *Ecological Modelling* **281**pp. 15-25.

Rodwell, J.S. (2006). *National Vegetation Classification: Users' Handbook*. Peterborough: Joint Nature Conservation Committee.

Rose, F. (1989). *Colour Identification Guide to the Grasses, Sedges, Rushes and Ferns of the British Isles and North Western Europe*. London: Viking.

Rose, F. (2006). The Wild Flower Key (Revised Edition) - How to identify wild plants, trees and shrubs in Britain and Ireland. London: Warne.

Rossi, J.P. and van Halder, I. (2010). Towards indicators of butterfly biodiversity based on a multiscale landscape description. *Ecological Indicators* **10**(2), pp. 452-458.

Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C. and Robinson, L.D. (2012). Understanding Citizen Science & Environmental Monitoring, NERC Centre for Ecology & Hydrology and Natural History Museum

Rushton, S.P., Ormerod, S.J. and Kerby, G. (2004). New paradigms for modelling species distributions? *Journal of Applied Ecology* **41**(2), pp. 193-200.

Santika, T. (2011). Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography* **20**(1), pp. 181-192.

Saura, S. and Pascual-Hortal, L. (2007). A new habitat availability index to integrate connectivity in landscape conservation planning: Comparison with existing indices and application to a case study. *Landscape and Urban Planning* **83**(2-3), pp. 91-103.

Saura, S. and Rubio, L. (2010). A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. *Ecography* **33**(3), pp. 523-537.

Saura, S. and Torne, J. (2009). Conefor Sensinode 2.2: A software package for quantifying the importance of habitat patches for landscape connectivity. *Environmental Modelling & Software* **24**(1), pp. 135-139.

Saura, S., Torras, O., Gil-Tena, A. and Pascual-Hortal, L. (2008). *Shape Irregularity as an Indicator of Forest Biodiversity and Guidelines for Metric Selection*. Dordrecht: Springer.

Schindler, S., Poirazidis, K. and Wrbka, T. (2008). Towards a core set of landscape metrics for biodiversity assessments: A case study from Dadia National Park, Greece. *Ecological Indicators* **8**(5), pp. 502-514.

Schneider, C. and Fry, G.L.A. (2001). The influence of landscape grain size on butterfly diversity in grasslands. *Journal of Insect Conservation* **5**(3), pp. 163-171.

Schroeder, B. and Richter, O. (2000). Are habitat models transferable in space and time? *Zeitschrift fuer Oekologie und Naturschutz* **8**(4), pp. 195-205.

Schröder, B., Strauss, B., Biedermann, R., Binzenhöfer, B. and Settele, J. (2009). Predictive species distribution modelling in butterflies. In: Settele, J., Shreeve, T., Konvicka, M. and Van Dyck, H. *Ecology of Butterflies in Europe*. Cambridge: Cambridge University Press.

Schweiger, O., Dormann, C.F., Bailey, D. and Frenzel, M. (2006). Occurrence pattern of Pararge aegeria (Lepidoptera : Nymphalidae) with respect to local habitat suitability, climate and landscape structure. *Landscape Ecology* **21**(7), pp. 989-1001.

Sekar, S. (2012). A meta-analysis of the traits affecting dispersal ability in butterflies: can wingspan be used as a proxy? *Journal of Animal Ecology* **81**(1), pp. 174-184.

Shreeve, T.G. and Dennis, R.L.H. (2011). Landscape scale conservation: resources, behaviour, the matrix and opportunities. *Journal of Insect Conservation* **15**(1-2), pp. 179-188.

Shreeve, T.G., Dennis, R.L.H., Roy, D.B. and Moss, D. (2001). An ecological classification of British butterflies: Ecological attributes and biotope occupancy. *Journal of Insect Conservation* **5**(3), pp. 145-161.

Simova, P. and Gdulova, K. (2012). Landscape indices behavior: A review of scale effects. *Applied Geography* **34**pp. 385-394.

Spencer, J.W. and Kirby, K.J. (1992). An inventory of ancient woodland for England Wales. *Biological Conservation* **62**(2), pp. 77-93.

Sterry, P. (2008). *British Trees: A photographic guide to every common species (Collins Complete Guide).* London: Collins.

Stevens, J.P., Blackstock, T.H., Howe, E.A. and Stevens, D.P. (2004). Repeatability of Phase I habitat survey. *Journal of Environmental Management* **73**(1), pp. 53-59.

Stevens, V.M., Turlure, C. and Baguette, M. (2010). A meta-analysis of dispersal in butterflies. *Biological Reviews* **85**(3), pp. 625-642.

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M.C., Schwager, M. and Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography* **31**(1), pp. 79-92.

Thomas, J.A. (1983). A quick method for estimating butterfly numbers during surveys. *Biological Conservation* **27**(3), pp. 195-211.

Thomas, J.A. (2005). Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philosophical Transactions of the Royal Society B-Biological Sciences* **360**(1454), pp. 339-357.

Thomas, J.A., Bourn, N.A.D., Clarke, R.T., Stewart, K.E., Simcox, D.J., Pearman, G.S., Curtis, R. and Goodger, B. (2001). The quality and isolation of habitat patches both determine where butterflies persist in fragmented landscapes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**(1478), pp. 1791-1796.

Tischendorf, L. (2001). Can landscape indices predict ecological processes consistently? *Landscape Ecology* **16**(3), pp. 235-254.

Tischendorf, L. and Fahrig, L. (2000). On the usage and measurement of landscape connectivity. *Oikos* **90**(1), pp. 7-19.

Tolman, T. (2004) *Collins field guide – butterflies of Britain and Europe*. 2nd Edition. London: Collins.

Tscharntke, T., Klein, A.M., Kruess, A., Steffan-Dewenter, I. and Thies, C. (2005). Landscape perspectives on agricultural intensification and biodiversity - ecosystem service management. *Ecology Letters* **8**(8), pp. 857-874.

Tscharntke, T., Steffan-Dewenter, I., Kruess, A. and Thies, C. (2002). Contribution of small habitat fragments to conservation of insect communities of grassland-cropland landscapes. *Ecological Applications* **12**(2), pp. 354-363.

Tudor, O., Dennis, R.L.H., Greatorex-Davies, J.N. and Sparks, T.H. (2004). Flower preferences of woodland butterflies in the UK: nectaring specialists are species of conservation concern. *Biological Conservation* **119**(3), pp. 397-403.

Turner, M.G. (2005). Landscape ecology: What is the state of the science? *Annual Review of Ecology Evolution and Systematics* **36**pp. 319-344.

Turner, M.G., Gardner, R.H. and O'Neill, R.V. (2001). *Landscape Ecology in Theory and Practice: Pattern and Process*. Springer-Verlag: New York, USA.

UKBMS. (2014). *Wider countryside butterfly survey*. [Online]. (http://www.ukbms.org/wcbs.aspx). (26/06/14)

UKEOF. (2014). *About UK EOF*. [Online]. (http://www.ukeof.org.uk/about/about-ukeof). (24/06/14)

van Halder, I., Barbaro, L., Corcket, E. and Jactel, H. (2008). Importance of semi-natural habitats for the conservation of butterfly communities in landscapes dominated by pine plantations. *Biodiversity and Conservation* **17**(5), pp. 1149-1169.

van Strien, A.J., van Duuren, L., Foppen, R.P.B. and Soldaat, L.L. (2009). A typology of indicators of biodiversity change as a tool to make better indicators. *Ecological Indicators* **9**(6), pp. 1041-1048.

Vandewalle, M., de Bello, F., Berg, M.P., Bolger, T., Doledec, S., Dubs, F., Feld, C.K., Harrington, R., Harrison, P.A., Lavorel, S., da Silva, P.M., Moretti, M., Niemela, J., Santos, P., Sattler, T., Sousa, J.P., Sykes, M.T., Vanbergen, A.J. and Woodcock, B.A. (2010). Functional traits as indicators of biodiversity response to land use changes across ecosystems and organisms. *Biodiversity and Conservation* **19**(10), pp. 2921-2947.

Vaughan, I.P. and Ormerod, S.J. (2005). The continuing challenges of testing species distribution models. *Journal of Applied Ecology* **42**(4), pp. 720-730.

Vogt, P., Ferrari, J.R., Lookingbill, T.R., Gardner, R.H., Riitters, K.H. and Ostapowicz, K. (2009). Mapping functional connectivity. *Ecological Indicators* **9**(1), pp. 64-71.

Warmington, K. and Vickery, M. (2003). *Warwickshire's Butterflies: Their Habitats and Where to Find Them*. The Warwickshire Branch of Butterfly Conservation: Warwickshire, UK.

Watson, R. and Albon, S. (2011). The UK National Ecosystem Assessment: Synthesis of the Key Findings *In* NEA, U. (ed.). Cambridge, UNEP-WCMC.

Watts, K. (2006). British forest landscapes: The legacy of woodland fragmentation. *Quarterly Journal of Forestry* **100**(4), pp. 273-279.

WCC. (2014). *Habitat Biodiversity Audit*. [Online]. (http://heritage.warwickshire.gov.uk/ecology/data-and-ecological-records/habitat-biodiversity-audit/). (26/06/14)

Wu, J.G. (2004). Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecology* **19**(2), pp. 125-138.

Wu, J.G., Shen, W.J., Sun, W.Z. and Tueller, P.T. (2002). Empirical patterns of the effects of changing scale on landscape metrics. *Landscape Ecology* **17**(8), pp. 761-782.

Yamaura, Y., Kawahara, T., Iida, S. and Ozaki, K. (2008). Relative Importance of the Area and Shape of Patches to the Diversity of Multiple Taxa. *Conservation Biology* **22**(6), pp. 1513-1522.