University of Southampton Research Repository

UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Centre for Biological Sciences

Prediction of Avian Species Composition from Assemblage Structure

by

Jing-Lun Huang

Thesis for the degree of Doctor of Philosophy

January 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Centre for Biological Sciences

Thesis for the degree of Doctor of Philosophy

Prediction of Avian Species Composition from Assemblage Structure

Jing-Lun Huang

This research focuses on how capitalising on community pattern, a character of ecological communities, could improve the predictability of community models, thus facilitating research in conservation. Patterns of communities not only depict phenomena but are also useful for predicting potential changes in species composition when patterns are governed by specific mechanisms. Most conventional prediction models do not take community pattern into consideration, despite the fact that incorporating community patterns into conventional models for predicting species richness and composition may enhance predictability. In this thesis, I assessed if incorporating two community patterns, nestedness and species co-occurrence, into conventional prediction models could improve the model predictability. Nestedness is a non-random species distribution pattern in which species in depauperate sites are contained in species-rich sites. Co-occurrence networks categorise species assemblages that reflect differential habitat requirements. I demonstrate that capitalising on nestedness provides a novel approach for improving the predictive power of species accumulation curves for species richness in unsampled areas. Specifically, while species richness is usually overestimated when the data are inputted in random order (the conventional approach), species richness is underestimated when the data are inputted in nested order. Taking an average of projected species richness of these two inputting orders dramatically lowers the prediction error rate, indicating that using nestedness in addition to random orders can greatly improve the predictive power of species distribution curves. I also show that network analysis can improve the ability to correctly classify site groups, which is the basis for calculating the indicator species value, by accurately reflecting similar ecological requirements of co-occurred species. Indicator species identified by network modularity, comparing to conventionally based on the k-means clustering method, can more successfully assign unsampled sites to the correct species groups and recognise representative species for the groups. These methods were tested using both British and Taiwanese bird assemblages. Both case studies supported the above conclusions, suggesting that the methods developed in this thesis have real promise for conservation applications. However, further work is required to assess whether these two novel pattern-based approaches are similarly applicable in other geographic regions or taxas.

To my grandpa and grandma

Table of Contents

Table of Tables

Table of Figures

DECLARATION OF AUTHORSHIP

I, Jing-Lun Huang, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Prediction of Avian Species Composition from Assemblage Structure

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signed: ....

Date: ....

Acknowledgements

Firstly, I would like to thank my supervisors, Felix Eigenbrod and Patrick Osborne. Thanks for your supports, inspirations and encouragements all the way throughout my PhD life. I really appreciate the chance that I can learn from both of you. Thank you both for always giving me hope to carry on. Thank you Felix that you always let me keep my own pace on the research. Especially thank you for always being very understanding to my family circumstances and helped me to exclude many obstacles and problems. Thank you Patrick, your simple question became the pillar of my PhD research. It is really fun to answer this question and I really enjoy these processes.

Many thanks to Ms Lorraine Prout in Graduate School Office of Biological Sciences, for giving me so many logistical supports. Thanks to the School Heads, Dr Geremy Webb and Dr Rob Ewing, who provided very generous and warm helps when I experienced the administrative issues.

I would like to thank to all who provided me academic supports in both Britain and Taiwan, they are: Dr Patrick Doncaster in University of Southampton, Dr Richard Field in University of Nottingham, Dr Tsung-Jen Shen in National Chun-Hsing University, Dr Anne Chao in National Tsing Hua University, Dr Pierre Legendre in Université de Montréal, Dr Tzung-Su Ding in National Taiwan University, Dr Laura Graham in University of Southampton and Dr Chi-Chien Kuo in National Taiwan Normal University. Also thanks to Dr Chris Jackson in University of Southampton, Dr I-Hsin, Sung in National Chiayi University, Dr Chi-Ling Chen and Dr Chien-Chung Chen in Taiwan Agricultural Research Institute for your advices and support in the early stage of developing my PhD research topic.

Many thanks to University of Southampton and Ministry of Education of Taiwan Government for funding my PhD study. Also many thanks to British Trust for Ornithology and Taiwan Endemic Species Research Insititue for providing me very precious avian data to be the main base of my PhD research. Thanks to Dr Simon Gillings and Dr Humphrey Crick in British Trust for Ornithology, Dr Meng-Wen Fan, Mr Chie-Jen Ko and Ms Yu-Chen Lu in Taiwan Endemic Species Research Institute for helping me solve the data related issues.

My beloved Dad and Mum, thank you so much for your unconditional support. Also thanks to my Parents-in-law, I really appreciate your emotional and economical supports. Thanks to Emma Wright, Rebecca Spake, Judith Lock, Neil Gostling, Robert Holland and all the B85 Level-6 cohorts. Thanks for your academic and emotional support and helped me get used to the life in England very quickly. It is very lucky to meet you in Britain. Thanks to all who be in company with me in I-Lan, especially my Auntie and cousin Cathy, when I experience this special life sapm. Your companies give me strengths to carry on. Finally, the very special thanks to my dearest partner (in both family and research), and my three lovely children's Dad, Chi-Chien. Thanks for your supports all the ways. It is through you that I can complete this research and live in my dream life.

Abbreviations

Abbreviation	Term
AIC	Akaike’s Information Criterion
BBS	Breeding bird survey
BTO	British Trust for Ornithology
CCA	Canonical correspondence analysis
Clh	Clench
Exp	Exponential
FM index	Fowlkes-Mallows index
IndVal	Indicator species value
Log	Logarithmic
NDVI	Normalized difference vegetation index
RDA	Redundancy analysis
SDMs	Species distribution models

Chapter 1: General Introduction -- Ecological communities, community pattern and the theories derived from these patterns

1.1 Ecological communities and community patterns

Community ecology is one of the major disciplines in ecology, focusing primarily on how and why specific species occur in one area and the interactions among species. Communities support the biodiversity that underpins ecosystem function, which regulates the ecosystem services on which human wellbeing depends (Cardinale et al. 2012, Balvanera et al. 2014, Allan et al. 2015). Changes to community structure therefore have repercussions on the survival and welfare of wildlife on earth including human beings. Therefore research on community pattern, which is one way to detect environmental change, matters. A community can be characterised by some attributes, including diversity, species richness, species-abundance relationships, and species composition (Morin 1999). The former three depict the numerical relations among species and provide the basis for comparison among communities. For example, alpha-diversity measures species richness and abundance in one site while beta-diversity is defined as species turnover among communities (Morin 1999, Socolar et al. 2016) . Community composition, on the other hand, is the collection of species that occur in each community.

Community composition is typically determined by abiotic factors and biotic interactions among species (Watt 1947). For example, species composition of a grassland community can be affected by soil properties (abiotic factor) and competition among grass species for similar resources (biotic factor). Patterns of community composition will vary with the processes involved in the formation of a community. Some well-known patterns of community include, for instance, the distribution-abundance relationship which describes how broadly distributed species are also more locally abundant (Holt et al. 2002); island biogeography theory which postulates that species number is determined by habitat size and remoteness (MacArthur 1967); and the species rank-abundance distribution which illustrates different descending distribution slopes (e.g. broken stick distribution, geometric series, and lognormal distribution) for species abundance in a community (Whittaker 1975). In general, the main goal of community ecology is not only to identify patterns of community, but also to elucidate the underlying mechanisms of these patterns.

1.2 Predictive models simulated at the community level

Patterns of communities not only depict phenomena but are also useful for predicting potential changes in species composition. This is particularly true when patterns are governed by specific mechanisms. If there are changes in the background mechanism, changes in the patterns can be predicted (D’Amen et al. 2017). For example, understanding how food resources are allocated would allow the prediction of the distribution of an animal which relies on particular food resources. There are various approaches for constructing species distribution models based on community patterns or the processes leading to them. Conventional methods rely on environmental attributes for the prediction of species distributions and are currently the most prevalent method for constructing species distribution models (Sinclair et al. 2010, Zimmermann et al. 2010, Ruhí et al. 2014). However, it has been argued that ecological processes, such as inter-specific interactions, are also crucial for building predictive models (Austin 2002, Godsoe and Harmon 2012) and that the determinants of species distributions may change over time (Skelly et al. 1999, Ackerly 2003). Different approaches have been utilised for building species distribution models but there are strengths and weaknesses inherent in each method, which I review here.

1.3 Predictive models based on environmental constraints

The last few decades have seen much interest in the development of species distribution models (SDMs) (Guisan and Thuiller 2005, Elith and Leathwick 2009, Calabrese et al. 2014, Guillera-Arroita et al. 2015) and their application in the prediction of potential change in distribution following environment variations such as climate or land-use changes (Dirnböck 2003). Typically, SDMs associate environmental variables to distribution of individual species to help reveal how environmental factors may govern distribution of wildlife (Guisan and Zimmermann 2000). SDMs have proven effective in associating and predicting species' distributions (Zimmermann et al. 2010) and have been widely applied to ecological and environmental studies, such as the validation of biogeographical (Romdal et al. 2005, Randin et al. 2006, McInerny and Purves 2011) and ecological (Mouton et al. 2010, Meier et al. 2010, Godsoe and Harmon 2012) hypotheses, prediction of range expansion of invasive species (Ward 2006, Ficetola et al. 2007, Václavík et al. 2012), and provision of predictions and solutions for conservation related issues, such as potential shifts in species distributions following land use or global climate change (Pearson and Dawson 2003, Thuiller et al. 2004, Watt et al. 2011). SDMs could also be very useful for ecological surveys and for establishing conservation strategies. For example, the efficiency of detecting rare species (LeLay et al. 2010) can be improved by modelling potentially suitable habitat. SDMs also help predict how species distribution may shift with climate change, such as projection of amphibian and reptile distributions under global warming (Araújo et al. 2006).

SDMs are also known as environmental niche models or ecological niche models (Elith and Leathwick 2009). The environmental attributes that constrain the distribution of a species reflect the preference and the realistic niche of a species (Austin 2007). The commonly used algorithms for SDMs are generalised linear models (GLM) (Thuiller 2003, Segurado and Araújo 2004, Pearson et al. 2006, Guisan et al. 2007, Oppel et al. 2012, García-Callejas and Araújo 2016), generalised additive models (GAM) (Lehmann et al. 2002, Thuiller 2003, Segurado and Araújo 2004, Pearson et al. 2006, Guisan et al. 2007, Oppel et al. 2012, García-Callejas and Araújo 2016), maximum entropy (Maxent) (Phillips et al. 2006, Guisan et al. 2007, Oppel et al. 2012, Merow et al. 2013, García-Callejas and Araújo 2016), artificial neural networks (ANN) (Thuiller 2003, Araújo et al. 2006, Olden et al. 2008, Larsen et al. 2012), classification and regression trees (CART) (Thuiller 2003, Olden et al. 2008), and boosted regression trees (BRT) (Oppel et al. 2012, García-Callejas and Araújo 2016). Developing these models is challenging and each has its own statistical background and limitation. Some models (especially mechanistic models that synthesise individual responses into demographic models) require further field experiments or observations to estimate key parameters (Johnson and Omland 2004, Guisan and Thuiller 2005, Holt 2009).

The SDM approach has two major limitations in terms of predictions. One is that these distribution models focus almost exclusively on separate species instead of considering all species as a whole. Detailed information on abundance and distribution of the majority of species is thus a prerequisite for a satisfactory performance of SDM, which may not be available for less common species or places with less extensive survey (Ferrier and Guisan 2006). Moreover, SDMs typically consider only abiotic factors and treat the study as a closed and isolated system while neglecting the importance of biotic interactions, such as competition, predation, or mutualism, in shaping species distribution and the reality is that ecological systems are more or less interconnected (Cassini 2011, Gavish et al. 2017). This is similar to the contrast between the Grinnellian niche and the Eltonian niche, with the former considering environmental variables only (e.g. temperature, precipitation, solar radiation, etc.) but neglecting biotic interactions (e.g. competition, predation and parasitism) as emphasised in the Eltonian niche (Soberón 2007). SDMs only reflect the Grinnelian but not the Eltonian niche, thus missing important components of the full niche (Hutchinsonian n-dimensional hyperspace).

Modelling at the community level investigates environmental correlations of biodiversity from a different perspective than single-species modelling. While species-level modelling could be useful for predicting occurrence of single species, taxonomic groups are typically considered for real world conservation issues. For example, when investigating the impact of climate change on a community containing numerous plant and animal species, the response of each individual species may vary, and the overall effect is not simply the sum of separate species but also includes the interactions among them. Species-level modelling may thus become inappropriate under such circumstance (Mokany and Ferrier 2011).

Species-level data can still assist community-level models because: (a) species-level data can serve as groundwork, providing individual distribution maps to be assembled and analysed; (b) predictors for each individual distribution can be established first, then with the species-level base maps assembled; or (c) species-level data can be applied to disparate statistical models in order to detect the predictor variables and ensemble at the same time (Ferrier and Guisan 2006). However, none of these methods could accurately predict community composition (Baselga and Araújo 2010). Moreover, these three predicting methods rely largely on environmental factors, which will lead to low predictability when environments cannot reliably reflect species distribution. Furthermore, exclusion of biotic factors may further reduce predictability of these models.

1.4 Community-level modelling based on interspecific interaction

Biotic interactions are usually neglected in classical SDMs (Elith and Leathwick 2009, Kissling et al. 2012, De Araújo et al. 2014). Predictive models with limited incorporation of biotic interactions might fail to predict the consequences of global change for species and ecosystems (Guisan and Thuiller 2005, Gilman et al. 2010, Kissling et al. 2012, Clark et al. 2014). Different from species distribution modelling based on environmental factors, some ecologists attempt to predict distribution of species by focusing on interspecific interactions. A classic example of how community composition can be affected by interspecific interaction is Robert Paine’s observations from the tidal pools of the Pacific rocky shore that presence of top predators could lead to the coexistence of its prey (Paine 1966). Starfish (*Pisaster ochraceus*) is a generalist predator preying on most mollusc species in the tidal pool. With intermediate predation intensity the biodiversity is the highest since no species can dominate the resources, thus leading to coexistence of several species. However, species richness is greatly reduced after manual removal of starfish (Paine 1974). This example, termed keystone predation (Gilman et al. 2010, Rudolf and Rasmussen 2013), demonstrates how biotic factors such as predation could shape community composition.

Besides predation, competition has also been included in the modelling of species distributions. Diamond (1975) stresses the importance of competition for species assemblage, which leads to much debate on the generality of this theme (Gotelli and McCabe 2002, Yackulic et al. 2014, Rollinson et al. 2016). According to Diamond, related species would be less likely to coexist due to competition, thus fewer species are likely to occur at a given site than by chance. Leathwick and Austin (2001) evaluated the importance of competition for the distribution of New Zealand tree species and found that including competition leads to higher predictability in species distribution model than relying simply on environmental factors. The model with inclusion of competition also helps explain disjunct species distributions in the middle of the temperature gradient.

Examples of including biotic factors in SDMs, such as applying Lotka–Volterra models (Solé and Bascompte 2006, Holt and Barfield 2009), population dynamic models (Mutshinda et al. 2009, 2011), epidemiological models (Keeling 2001, Estrada-peña and Fuente 2016), multivariate regression models (Banerjee et al. 2008, Latimer et al. 2009, Ovaskainen et al. 2010, Ovaskainen and Soininen 2011), trophic interaction distribution model (Trainor et al. 2014) and dynamic vegetation models (Lischke et al. 2006, Prentice et al. 2007) can be found in the literature (Kissling et al. 2012).

A contrasting approach is neutral theory (Hubbell 2001), which posits that the role of trophically similar species in the assembly of ecological community is neutral (equal). The distributions of species are determined by chance and are unrelated to their niche. It is a null simplification of competitive interactions that works well for tropical forests in particular – random processes are more important than deterministic differentiation. (Hubbell 2001, 2005, 2006). Neutral theory has been repeatedly investigated, but is not validated in many empirical studies (McGill and Collins 2003, Chave 2004). However, neutral theory provides a valuable baseline for a comparison with empirical data or other models. If a set of data is not consistent with prediction from a neutral model, some biological factors might be needed to be considered (Rosindell et al. 2011).

Despite recent advances, translating species' niches into geographic distributions remains a complex and difficult task (Holt 2009), and the modelling of spatio-temporal dynamics in multispecies communities remains a significant challenge in ecology (Solé and Bascompte 2006). Modelling large spatial and temporal datasets with multiple interacting species, coupled with spatially (and maybe temporally) varying parameter estimation and non-stationary covariance structures clearly pose considerable challenges for statistical modelling. Indeed, the complexity of multispecies models renders estimating all parameters at the same time impractical (Kissling et al. 2012). The key is therefore to reduce the complexity and there are established methods to achieve this goal.

1.5 Community-level modelling community structure based on quantitative analysis

Quantitative characterisation of community structure involves interpretation of biological survey data from a community perspective. Recognising and analysing community structure can help unravel mechanisms (such as environmental gradients) underpinning community composition and allows the prediction of how composition will change (Wiegand et al. 2003, Schröder and Seppelt 2006, Grimm and Railsback 2012). Many indices have been developed (Stone and Roberts 1990, Atmar and Patterson 1993, Koleff et al. 2003, Podani and Schmera 2011), including nestedness and beta-diversity, which are commonly applied to depict community structure. Nestedness illustrates an ordered decrease in species richness among sites caused by some non-random factors (species loss), while beta-diversity stresses how species are replaced among different sites (species turnover). These two characteristics, species loss and species turnover, are essential for depicting community structure. Therefore, it is often sufficient to use only these two indices to distinguish two communities with distinct structures (Baselga 2010).

1.5.1 Nestedness

The concept of nestedness was firstly proposed by Patterson and Atmar (1986) to explain the peculiar insular fauna structure: island species abundances decrease with distance from the continent; moreover, species on distant islands are a subset of species on proximate ones. A perfect nested matrix indicates that a community is arranged in an absolute mathematical order, in which species within a depauperate site also occur in a species-rich site and thus constitute a sequentially shorter ladder. Therefore in a strongly nested community, species composition is predictable from knowledge of the rare species, since any samples from the community taken at different points in space or time tend to share all of the more common species (Doncaster et al. 2016). This nestedness index took the idea of matrix system disorder or entropy. The relationship between entropy and heat led the nestedness index to quantify matrix randomness in terms of temperature (Norton et al. 2004), which ranges from 0-100°, with T= 0º representing a perfect-nested matrix (minimum entropy) and T= 100º absolute randomness (maximum entropy) (Rodríguez-Gironés and Santamaría 2006). This relationship is like the three phases of water, the liquid phase scaling between phases of maximal structure as ice and maximal disorder as vapour, and low nestedness reflects high turnover of species in space or time whereas high nestedness means ordered accumulation or loss (Doncaster et al. 2016). The concept of nestedness has been widely applied to terrestrial communities (Atmar and Patterson 1993), including small mammals (Cutler 1991), birds (Mac Nally et al. 2002), reptiles (Fischer and Lindenmayer 2005), and invertebrates (Fleishman and Mac Nally 2002). Various hypotheses have been proposed to explain the proximate cause of nestedness, including passive sampling (Fischer and Lindenmayer 2002, Higgins et al. 2006), neutrality (Ulrich and Zalewski 2007), colonization (Patterson 1990, Cook and Quinn 1995, McAbendroth et al. 2005), extinction (Patterson and Atmar 1986, Wright and Reeves 1992, Bruun and Moen 2003, Wethered and Lawes 2005), nested habitat (Brualdi and Sanderson 1999, Fleishman and Mac Nally 2002), and habitat hospitability (Bloch et al. 2007). Among these, the extinction hypothesis, which states that differential local extinction rate among species is the main driving force for nested structure, attracts the most attention (Fleishman and Murphy 1999). Although the mechanisms underlying nestedness patterns are not well elucidated, pattern of nestedness are frequently investigated in studies of species conservation. For example, with a further decrease in habitat size or with a further deterioration in habitat quality, the remnant species are often a non-random subset of the total pool species, and are generally dominated by generalist species that can tolerate more degrees of environmental variance (Cutler 1991, Fleishman and MacNally 2002, Fleishman et al. 2002, Baber et al. 2004, Hylander et al. 2005, Schouten et al. 2007, Hill et al. 2011).

1.6 Applying community orderliness to species richness prediction

Rather than focusing on environmental or intraspecific factors, an alternative approach to predicting community composition is based on intrinsic community pattern. This approach has the advantage that it avoids involving specific processes that may not yet be known. For example, the relationship between species range size and species diversity helps reveal non-random pattern in community structure (Arita et al. 2008). A negative relationship occurs between species diversity and species distribution range of small mammals in North America based on species presence-absence matrix. Species with restricted distribution (rare species) tend to occur in sites with higher species diversity. In other words, “hotspots” that contain more species are mostly the result of presence of more rare species. Another example is that structure indices such as community alpha- (species richness in sites at a local scale) and beta-diversity (species turnover between regional or local species diversity) might potentially be used to predict community composition derived from limited ground survey data (Mokany et al. 2011, Prober et al. 2015, Socolar et al. 2016, Gavish et al. 2017).

From a different perspective, some analyses utilise the property of community structure such as alpha diversity or beta diversity, to predict the pattern underlying the structure. For instance, species accumulation curves use the rate of species accumulation (species number per sampling area, performance of beta-diversity) to predict the overall species richness (the point value of a species pool, one common index of alpha-diversity) based on extrapolation of accumulation curve to an asymptote. Species richness is the basic outline of a community, providing fundamental measurement of community and regional diversity. It is the foundation of many ecological models and conservation strategies. In spite of its importance, ecologists have not always appreciated the influence of abundance and sampling effort on estimated species richness (Gotelli and Colwell 2001). Information on species richness relies mainly on field surveys and requires significant sampling effort. Implementation limitations, such as the difficulty in accessing remote areas and the high costs in surveying a large region, has rendered the determination of the most efficient and reliable sampling scenario a priority in community ecology research. Indeed, several methods have been designed to achieve this goal (Bunge and Fitzpatrick 1993, Colwell andCoddington 1994, Chao and Shen 2004, Gotelli et al. 2009, Chao and Jost 2015a). In recent decades, there have been many attempts to estimate species richness based on species accumulation curves, where species richness is projected based on its relationship with sampling effort or area. With an increase in sampling effort, the number of species initially increases, but typically reaches an asymptote (Soberón and Llorente 1993, Ugland et al. 2003). The steepness of the curve specifies species dissimilarity among sampled areas, which is also useful in estimating species abundance in unsampled areas. In addition, a slight raise in extended tail that typically occurs in species accumulation curves suggests the existence of undetected rare species. An asymptote in the curve indicates that the increase in the species richness of an area is not unlimited, thus making extrapolating total richness from species accumulation curve possible (Morin 1999). Finally, as the asymptote represents the saturation point in species richness, it can help determine the least sampling effort for estimating total richness without all areas being sampled.

One area that has received little attention to date is to investigate how well general insights into community structure can be used to predict species richness, based on the species presence-absence orderliness. For example, nestedness describes community structure by sorting species richness non-randomly from species-rich to species-poor area (Smith and van Belle 1984, Patterson and Atmar 1986, Gotelli 2000, Podani and Schmera 2011, Ricotta and Pavoine 2015). Nestedness might be helpful for improving the predictive power of the species accumulation curve because a nested-order curve includes proportionally more uncommon species in the initial samplings, followed by a decelerating rate of increase in species richness due to fewer uncommon species remaining undetected. Therefore the curve is a constantly upward convex shape with a flattened tail. The first data chapter of my PhD (Chapter 2) focuses on how nestedness structure in a community can improve the predictability of species richness from species accumulation curves.

1.7 The importance of species identity in community composition prediction

Mapping community nested structure gives the ‘big picture’ of how community patterns vary spatially. Although nested structure improves predictability of species richness in an area (see Chapter 2), the species composition in specific sites within the area remains unknown. Information on the consistency in species composition, particularly the repeated co-occurrence of certain species group, is helpful for the prediction of unknown species because the observation of one or more common species can predict the occurrence of other species within each group.

Braun-Blanquet's work (1932) on plant sociology was one of the first attempts to classify communities efficiently (Poore 1955). Fidelity, how frequently a species occurs in a specific community, is critical for assigning species into distinct communities (Braun-Blanquet 1932). A species with higher fidelity means it occurs more frequently in one community and is less likely to be found in the other community. This method sorts the community member based on abundance/dominant scoring (AD score) and the community is represented by a combination of species with the highest score. Subjectivity and arbitrariness in the Braun-Blanquet's scoring system can lead to biased results (Podani 2006). Nevertheless, the idea of species occurrence probability proposed by Braun-Blanquet is still a useful concept for predicting community composition, and lends itself to more quantitative analysis via network analysis.

Network analysis is ideal for analysing complicated relationship among members within communities. Nodes and edges in a network represent members in the community and their interrelations, respectively. Network analysis is based on mathematical algorithms of graph theory and can disentangle complicated relationships even with large amount of data and interactions, so is widely adopted in various academic disciplines such as sociology, information science and biology (Newman 2010). For example, network analysis is adopted in protein research for identifying amino acids co-expressing in the same metabolism pathways (Jeong et al. 2001) and in neural sciences for classifying brain cells and ganglia according to probabilities of participating in the same brain function (Voytek and Voytek 2012). In ecological research, network analysis helps clarify relationship among predators and prey in the food web (Dunne et al. 2002, Krause et al. 2003, Navia et al. 2010, Kéfi et al. 2015) or plant and animal interactions in pollination networks (Olesen et al. 2007, Devoto et al. 2012, Olito andFox 2015, Biella et al. 2017). These studies all involve numerous members (nodes) and complex relationships (edges).

Based on Braun-Blanquet, a community can be defined by simultaneous attendance of the same event (co-occurrence of species in the same location) (Braun-Blanquet 1932, Podani 2006). That is, species with more frequent co-existence should belong to the same community. One type of network analysis termed an affiliation network is capable of sorting large amounts of species data according to this rule. An affiliation network is one kind of two-mode networks which contain two sorts of nodes. Therefore this type of network describes not only the direct linkage between members but also the events these members “attend” (biologically, events attended correspond to the habitat that species occupy) and the relationship among events. Once this member-event relationship is projected into a one-node network that contains only member relationships, the co-membership through attending to the same event will be revealed. When applied to ecological research, such projections can help categorise communities based on species co-occurrence.

1.7.1 Network and network modularity

The first step to define a community network structure is to identify subgroups in a network and there are numerous ways to do this, based on different concepts and algorithms (Guimerà and Amaral 2005, Fortunato 2010) derived from graph theory (Wilson 1996, White and Harary 2001). Nodes with more frequent links indicate cohesion of these nodes. Nodes that link to each other with at least one path are defined as a *component*, which is the most essential part of a subgroup (White and Harary 2001, Moody and White 2003). In a graph component, all nodes are linked to each other in this component by at least one path. A path is the complete linkage route between two nodes. If the path length is larger than one, it means linkages between two nodes are going through other nodes. For example, there are two components in Figure 1, comprising nodes {1, 2 and 3} and {4, 5, 6 and 7} respectively. Component {1, 2 and 3} is termed a strong component because members all directly link to each other. However component {4, 5, 6 and 7} is a weak component because it contains linkages with a path length greater than one. The edge between node 3 and 7 is the only route connecting these two components and is regarded as a bridge. Calculating relationships among all nodes and edges allows sub-groups to be identified.

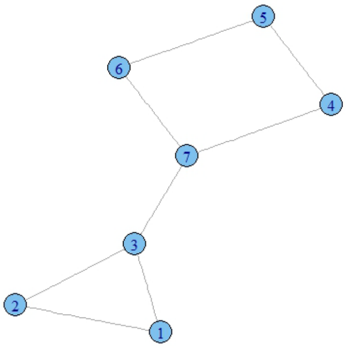


Figure 1. 1 A sample network graph composing two *components*, {1, 2 and 3} and {4, 5, 6 and 7}

One approach for identifying subgroups is to calculate the network modularity (Newman 2004). This concept has been adopted in ecological research to reveal intricate interactions and to detect heterogeneous structure within ecological communities or ecosystems (Olesen et al. 2007, Genini et al. 2012). Modularity not only measures linkages within a module but also considers linkages among modules. A module is a group of nodes, and within-module nodes possess a higher “degree” (the sum of edges link to the node) than other members in the same module but a lower degree to members belonging to other modules (Blondel et al. 2008, Newman 2010). In each module, the node with the highest degree represents the most dominant member (i.e. the member that links to most other members). This node is called a *module hub* and can efficiently indicate the presence of other members in this module (Guimerà and Amaral 2005, Olesen et al. 2007).

1.8 Applying community identification to community composition prediction

Compared to niche modelling that requires environmental attributes, using species information solely to predict species composition is based on the assumption that when a group of species is closely associated with specific ecological requirements, this group of species can be taken as "indicator species" to predict characteristics of unsampled sites (Harms et al. 2001, DeCáceres and Legendre 2009). A method developed by Dufrêne and Legendre (1997), called "indicator species value", stresses species-habitat associations. Dufrêne and Legendre's indicator species are selected based on two characteristics: 1) the degree to which the species is associated with specific habitat. The ideal situation is that a species only occurs in one specific habitat, thus reflecting its ecological preference or requirement; 2) the abundance of the species; that it is abundant enough to be effective as an indicator. Although a rare species can fit the first requirement, it would be hard to observe and therefore not a useful indicator.

Identifying indicator species requires classifying sites (or samplings) into different habitat types that each contains unique species composition. Each species is then assigned an indicator value for each habitat type. For instance, if three habitat types are classified, each species will have three indicator values corresponding to each of the three habitats. Indicator species for each habitat are defined as the species with the highest indicator value. However, the site groups are typically classified based on a clustering method especially non-hierarchical method such as k-means clustering which is a descriptive method whose robustness is hard to validate. It is also an indirect method, relying on distance (difference) between species rather than directly measuring co-occurrence patterns (Milligan 1980, Legendre and Legendre 2012). Clustering is also sensitive to the double-zero issue, which is very common in ecological datasets and may cause group misclassification (Dufrêne and Legendre 1997, Legendre and

Legendre 2012). The double-zero issue happens when comparing two site with the presence or absence of specific species. Presence in both sites of species indicates resemblance of sites, but absence in both sites does not necessarily represent resemblance because the absence of species in two sites may be due to different reasons. Distance metrics calculated in conventional clustering methods cannot reflect this fact. In comparison, a direct measure of species co-occurrence could improve the efficiency of the indicator species method. To improve efficiency of site group classification, site groups based on a modularity algorithm would be a substitute for the clustering groups in the indicator species analysis, with the merit of directly measuring species co-occurrence. Although species co-occurrence has been investigated before, this is the first study to quantify species co-occurrence with network analysis and to apply network-based co-occurrence groups for indicator species analysis. Hence for the second data chapter of my Ph.D (Chapter 3), I apply network analysis and modularity detection techniques for identifying indicator species.

1.9 Utilising orderliness of structure to predict species richness and composition

Knowledge on species richness and species composition is essential for the study of community ecology and is also critical for conservation biology and in conservation policy-making. Community pattern is a defining character of ecological communities and is also useful for predicting potential changes in species composition. Incorporating information on community pattern could improve the effectiveness of predictive community models, thus facilitating research in conservation. My PhD research focuses on how community structure with specific pattern (e.g. the non-random order of nestedness and the structure identified by network analysis) can be applied to improve predictability of conventional mathematic models that focus on community structure, such as species accumulation curve (beta-diversity) and indicator species index (alpha- and beta- diversity). The species accumulation curve is applied to predict overall species richness. In my PhD I show that because nestedness represents orderliness in a community, it is feasible to utilise nestedness analysis to improve the predictability of species richness in unsampled areas (Chapter 2). I show that nested structure in a community could improve predictability of community richness and composition based on a non-random mathematical relationship. In the second part of my thesis, I focus on prediction of species composition. The commonly applied indicator species value requires classification of habitat types that reflect ecological requirements of individual species. I show that network analysis can help produce a better classification of species (Chapter 3). British bird data are utilised to develop the methodologies in chapters 2 and 3 due to the quality of the data available. The last part of my PhD dissertation (Chapter 4) evaluates these methodologies in another island – Taiwan – as a test of the extent to which my findings in Britain can be applicable in a tropical setting. Overall, the methodologies I develop for my PhD study have the potential to increase predictability of the two components (species richness and species composition) that are essential for defining a community and has important applications for conservation research. The two datasets applied in the thesis, British bird data and Taiwanese bird data, include only a single taxon (i.e. birds). Under such a circumstance, using the term '*assemblage*' is more appropriate than community. For the following chapters, assemblage is therefore used in lieu of community.

Chapter 2: The utility of nestedness structure for predicting species richness

2.1 Abstract

A species accumulation curve is frequently applied to predict the overall species richness of a given area when resources for surveying the whole region are unavailable. It typically involves plotting a species accumulation against sampling effort, choosing a mathematical function to fit the curve, and extrapolating final species richness based on the function. Both the order of inputting in sampling data and the selected function affect the shape of the curve and the predicted species richness. Conventionally a randomisation process is used to minimise the prediction error caused by the order of data input. However, this randomisation process means that useful additional information on community structure is lost that may increase the predictive power of species accumulation curves. In this study, the degree to which incorporating nested structure (that is, inputting data from the most species rich to species poor sites, rather than randomly) improves the predictive power of species accumulation functions was assessed, with British avian data as an example. Two methods of ordering input data (randomly and based on nested structure) each for three levels of sampling efforts (10%, 20%, 40%) were fitted to three species accumulation curves functions (negative exponential, logarithmic, and Clench) under two spatial scales (the whole British island and 100 km square). While inputting data based on nested order had similar performance as random order under British island scale, inputs based on nestedness order detectably increase the predictability of total species richness at the 100 km square scale, especially coupled with the logarithmic function. This is probably because there is only one possible curve when ranking sites by nested order (starting with the most species-rich sites) whereas there are many possible curves when ranking sites based on random order. This study demonstrated that sampling according to nested order is generally more favourable than when based on random order. More studies considering different species and study site characteristics are needed to test for the general applicability of using nested order in species accumulation curves.

2.2 Introduction

2.2.1 Estimating species richness and species accumulation curves

Species richness provides a basic outline of a community and is an essential piece of information for species conservation. Information on species richness relies mainly on field surveys and requires significant sampling effort. Implementation limitations, such as the difficulty in accessing remote areas and the high costs of surveying a large region, have rendered the determination of the most efficient and reliable sampling scenario a priority in community ecology research. Indeed, several methods have been developed to estimate species richness based on individual abundance (individual based) or presence-absence data (sample-based) (Bunge and Fitzpatrick 1993, Colwell and Coddington 1994, Gotelli and Colwell 2001). These have mainly included species accumulation curve-fitting (Soberón and Llorente 1993, Díaz-Francés and Soberón 2005), parametric model fitting (Hortal et al. 2004, Connolly et al. 2009, Engen et al. 2011), nonparametric extrapolation from species accumulation curves (Colwell et al. 2012, Chao et al. 2014), or asymptotic estimators (Colwell and Coddington 1994, Chao and Shen 2004, Cardoso et al. 2014). Among these four methods, when only presence-absence data are available, the accumulation curve-fitting method is the most commonly applied approach.

There have been many attempts to estimate species richness based on species accumulation curves. Species accumulation curves take advantage of the relationship between species richness and sampling effort or area. As sampling effort is increased, the number of species initially increase, but typically reaches an asymptote (Soberón and Llorente 1993, Ugland et al. 2003). The steepness of the curve can be used to quantify how rapidly species dissimilarity changes between sampled areas, as well as to estimate species richness in non-sampled areas.

In addition, a slight raise in extended tail that typically occurs in species accumulation curves suggests the existence of undetected rare species. As the species richness of an area will not increase infinitely with increasing sampling effort, the species accumulation curve eventually reaches an asymptote and once this is reached, the total species richness of the area can be predicted (Morin 1999). Finally, as the asymptote represents the saturation point in species richness, it can help determine the least sampling effort for estimating total richness in other, ecologically similar regions.

A species accumulation curve can be constructed based on one-time sampling data or mean value of multiple samplings; the latter is called a rarefaction curve (Gotelli and Colwell 2001). The one-time accumulation curve uses original values collected randomly from a series of samples or along a gradient. However, the shape of the curve is largely determined by the order of adding additional samples (Ugland et al. 2003, Gray et al. 2004). For example, samples entered in the order of ascending richness leads to a smoothly growing curve while samples entered in the order of descending richness results in a steeply upward curve. Therefore, repeatedly randomly selecting the order in which samples are added, and building a curve based on the mean value of these randomly ordered sampling curves (that is, rarefaction) is necessary to eliminate the arbitrariness of one-time sampling (Colwell and Coddington 1994).

Species accumulation (or rarefaction) curves allow for the prediction of undetected species through the generation of a prediction function based on curve fitting and extrapolation. Again, the shape of the extrapolated species accumulation curve affects final predicted results and relies on choosing the appropriate equation for extrapolation. Soberón and Llorente (1993) proposed three prediction functions - exponential, logarithmic and Clench functions. The exponential (specifically negative exponential) (Miller and Wiegert 1989) and logarithmic (Gleason 1922) functions have long been adopted for fitting species accumulation curves

(Tjørve 2003, Dengler 2009). The Clench function originates from the Michaelis-Menten kinetics function in biochemistry, and was later applied to species accumulation curves by Clench (1979). The shape of the species accumulation curve determines which prediction function is most appropriate. The logarithmic curve is a continuously growing, concave downward curve without an upper asymptote while the negative exponential and Clench curves are saturation curves that decrease at an increasing rate until finally reaching an asymptote. The latter two curves vary in their rate of decrease and thus different turning angles, which is where the tangent slope along the curve has the greatest shift.

It is critical but difficult to select an appropriate fitting function for extrapolation because there is no universal recipe for deciding on which function to use (Soberón and Llorente 1993). A thorough understanding of the biological process governing a specific community is helpful for choosing the most appropriate function (Tjørve 2003). For instance, Thompson *et al.* (2003) suggested that a species accumulation curve was influenced by the proportion of common to rare species in a community. A community with more common species increases more rapidly in the early sampling stage, leading to an initially steep accumulation curve and then saturates (Thompson and Withers 2003). In contrast, a community with more rare species results in an accumulation curve with flatter slope and prolonged period of slow increase until reaching an asymptote. Therefore, studying community structure can potentially improve the predictive power of species accumulation functions as it provides additional information on orderliness and dynamics of communities that simple species accumulation curves fail to capture.

2.2.2 Community structure and nestedness subset

Quantifying community structure can be defined as considering biological survey data from a community perspective. For example, indices such as beta-diversity and nestedness are commonly adopted to describe the dissimilarity along sites. Recognising and analysing community structure can help reveal mechanisms (such as environmental gradients) underpinning community composition and allow the prediction of how species composition will vary based on these mechanisms. Many indices for quantifying community structure have been developed (Stone and Roberts 1990, Atmar and Patterson 1993, Koleff et al. 2003, Podani and Schmera 2011), including nestedness. The concept of nestedness was first proposed by Patterson and Atmar (1986) to explain a common phenomenon in the structure of insular fauna: species richness on islands decreases with distance from the continent; moreover, species on distant islands are a subset of those on proximate ones. Nestedness takes account of both species richness and species occurrence and is calculated through arranging community data into a matrix of study sites and species presence-absence data for each site. By arranging species and sites in a matrix, species composition can be compared among sites or the frequency of occurrence can be compared among species (Ulrich et al. 2009). A perfect nested matrix indicates that a community is arranged in an absolute mathematical order, in which species within a depauperate site also occur in a species rich site and thus constitute a sequentially shorter ladder (Figure 2. 1). Owing to such a unique order, a nestedness index can potentially improve predictability of species accumulation curves, because no other indices (ex. beta-diversity) provide similar information, but such a novel idea has never been applied for enhancing species richness predictability. Also by using this index, it should be possible to stabilise the shape of accumulation curves in a way other than using the conventional randomisation method, which can possibly lose biological information inherent in community structure. Because the uncertainty in predicting species number based on species accumulation curve largely comes from the randomisation processes (Gray et al. 2004) and the initial stage of curve shaping (Bebber et al. 2007), incorporating nested structure should decrease the uncertainty of species accumulation curve and increase the accuracy of predictive results.

Community nestedness can be represented with several indices but all involve counting the ‘surprises’ in the study matrix against a perfectly-nested matrix (Ulrich et al. 2009). Here the ‘surprise’ represents ‘unexpected absence’ and ‘unexpected presence’ - species that would be present/absent from the site in a perfectly nested assemblage. For example, the index “nestedness temperature” is calculated by identifying unexpected absences in species-rich site or unexpected presences in a species-poor site. By summing these ‘surprises’ in each community sub-unit, the extent to which the community is perfectly nested can be determined so that a higher temperature represents more ‘surprises’ and less orderliness in a community. In other words, a higher temperature indicates lower nestedness.

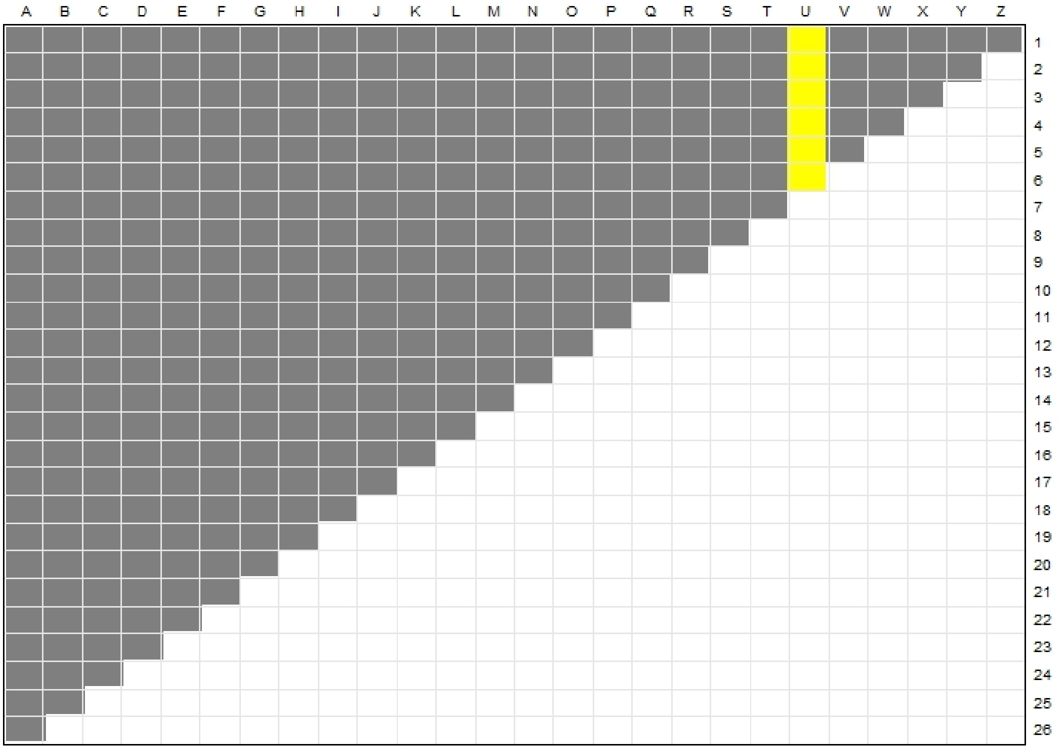


Figure 2. 1 A sample matrix representing a theoretical community of species A to Z in sites 1 to 26. Grey cells indicate which species are contained in each site. This virtual community provides an example of a perfectly nested community. For example, species U occurs only in the 6 species richest sites (sites 1-6).

2.2.3 Utilising nestedness to predict species richness

As mentioned earlier, the idea presented here is that the predictive power of species accumulation curve can be enhanced by a better understanding of the community structure. As a measure of community structure (orderliness of the community), nestedness could be helpful for improving the predictive power of the species accumulation curve when there is a moderate degree of nestedness in the community. This is because in a perfectly nested community, the site with the greatest species richness includes all species and each of the other sites is a non-random subset of the most speciose site (Wright 1998). Figure 2.1 gives a theoretical example. The species accumulation curve of a perfectly nested assemblage is very steep, but then level off very quickly because no more new species can be added after the first site, as the first site contains all species in the community. Such a species-rich to species-poor order will consistently result in an asymptotic curve with little variation. In other words, incorporating nested orderliness results in a more consistent species accumulation curve than the other commonly used rarefaction curve, which is derived from randomised multi-sampling data (Soberón and Llorente 1993, Scheiner 2003). For example, when based on nestedness order, sampling the example matrix (Figure 2. 1) only by sites of even numbers produces a curve similar to when sampling all sites. If species accumulation curves based on nested order do indeed give a more consistent curve shape than rarefaction curves, this could mean that such nestedness-based species accumulation curves also have higher predictive accuracy than the standard rarefaction curve approach.

2.2.4 Aim of this study

This study aims to evaluate whether building species accumulation curves based on ordering samples by their nestedness structure improves their utility for predicting species richness. British avian assemblage data (Gibbons et al. 1993) were examined because it contains a complete avian species list for Britain which provide a solid base for model validation (Ugland et al. 2003). In this study, differences in the predictive accuracy of three species accumulation curve functions (exponential, logarithmic, and Clench functions; Soberón & Llorente 1993) based on random vs. nested order in the addition of the sampling sites were compared. Three levels of sampling efforts (10%, 20%, or 40% of all sampling sites) were selected as the cut-off points for a comparison of accuracy in predicting species richness. All analyses were run at two spatial scales (whole of the British mainland vs. 100 km squares, see Methods) to evaluate whether predictive power is scale-dependent.

2.3 Methods

2.3.1 Avian data

The British Trust for Ornithology (BTO) 1988-1991 survey data (Gibbons et al. 1993) cover Britain and Ireland. In this study, only the British data were used, which were divided into 55 squares (Figure 2. 2) and each 100 km square was further divided into at most one hundred sub-squares (10 km x10 km). The entire BTO dataset contains 273 avian species in 2830 sub-squares. Presence-absence data for every species are available at the sub-square level and were inputted in a species-site matrix. The sub-square is the basic unit for the analyses of community structure of the British avian assemblage.

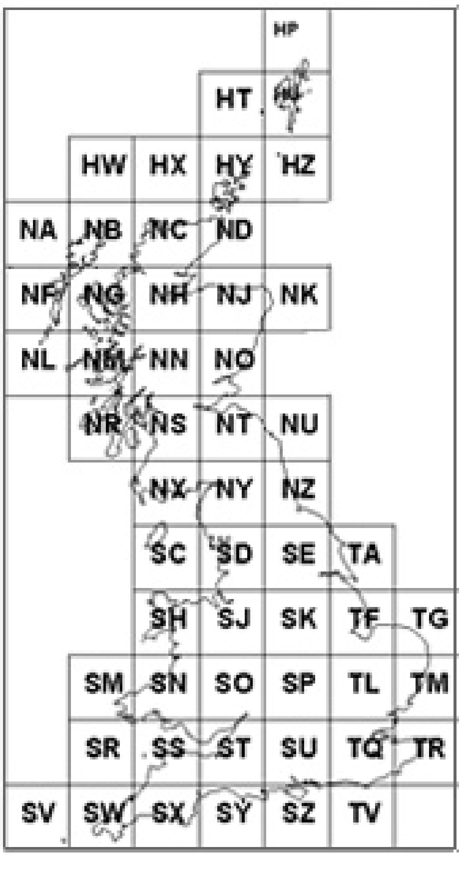


Figure 2. 2 Arrangement of British 100 km squares (http://www.bto.org/volunteer-surveys/birdatlas/taking-part/correct-grid-references/know-your-place)

2.3.2 Calculating nestedness

Nestedness temperature, a measure of community nestedness, takes unexpected presences and unexpected absences into account (Ulrich and Gotelli 2007, Ulrich et al. 2009). It is represented by a normalised sum of squared relative distance of absences above and presences below a hypothetical isocline that separates occupied from unoccupied areas in a perfect nested matrix (Atmar and Patterson 1993). According to Atmar and Patterson (1993), the calculation of nestedness is based on the unexpectedness, which is measured by the distance an unexpected presence or absence of a species lies in the matrix as compared to the perfectly nested matrix. Unexpectedness runs diagonally along a line running parallel to the skew diagonal. For a matrix with *n* species and *m* sites, the local unexpectedness (u_{ij}) is calculated as:

$$u_{ij} = \left(\frac{d_{ij}}{D_{ij}}\right)^2$$

D_{ij} is the length of the full line running through the *j*th species on the *i*th site and d_{ij} is the specific length along that line. The total unexpectedness (*U*) is then represented by:

$$U = \frac{1}{mn \sum_i \sum_j u_{ij}}$$

A perfectly ordered matrix should have no unexpectedness, thus *U*=0, and a matrix with maximum unexpectedness always has $U_{max} \approx 0.04145$. The nestedness temperature T is then defined as:

$$T = kU$$

$$k = \frac{100}{U_{max}}$$

T ranges from 0-100°, with T=0° representing a perfect-nested matrix and T= 100° absolute randomness.

Ordering a presence-absence matrix in nested order therefore means re-arranging it into a new matrix where sites are organized from the most species-rich to the most species-poor. The nestedness of the avian assemblage of all of Britain and of each of the fifty-five 100 km squares was calculated as a first step in understanding whether using nested order can help improve the predictions of species accumulation curves.

2.3.3 Species accumulation curve and extrapolation

2.3.3.1 Orders of sampling sub-squares for species accumulation curves

British avian assemblage data were arranged in three different types of order in terms of adding in sampling sub-squares: 1) random order with the mean of 100 permutations (rarefaction curves; sensu Gotelli & Colwell 2001); 2) nested structure – that is, ordering sampling sub-squares from the most species-rich to the most species poor. 3) reverse nested structure – that is, ordering sampling sub-squares from the most species-poor to the most species-rich.

2.3.3.2 Functions for fitting species accumulation curve

Three functions for fitting species accumulation curves were compared for all three types of orders of sample sub-squares (random, nested, reversed nested) to evaluate their accuracy in terms of predicting overall species richness of Britain and of each 100 km square. These three functions are based on Soberón & Llorente (1993):

Negative exponential function: $\mathbf{y} = \mathbf{a} \times \left(\mathbf{1} - \boldsymbol{e}^{-\frac{\boldsymbol{x}}{\boldsymbol{b}}}\right)$ (Equation 1)

Logarithmic function: $\mathbf{y} = \mathbf{a} + \mathbf{b} \times \mathbf{log}(\boldsymbol{x})$ (Equation 2)

Clench function: $\mathbf{y} = \frac{\boldsymbol{ax}}{\mathbf{1}+\boldsymbol{bx}}$ (Equation 3)

In which y is the species number and x the number of sub-squares.

2.3.3.3 Calculating the predictive power of species accumulation curve

The three functions mentioned above were used for extrapolating species richness in non-sampled areas. The ability of the three species accumulation functions to predict total species richness across Britain and within each 100 km square was compared based on three levels of sampling effort: 10%, 20%, and 40% of sub-squares from each 100 km square (termed 10%, 20% and 40% sampling data henceforth). For example, for the 20% sampling effort, 20% of sub-squares in each 100 km square were randomly selected to form 20% sampling data. This selected sampling data was randomly shuffled 100 times to calculate a mean value (random order) or rearranged according to a nestedness order (nestedness order). This procedure was conducted in two spatial scales to see whether predictability is scale dependent: the whole British mainland and within each of the fifty-five 100 km squares in Britain. At the scale of Britain as a whole, randomly selected data from each of the fifty-five 100 km squares were pooled together. For example, for 20% sampling data, 20% sub-squares from each 100 km square were selected and combined to form the total 20% sampling data of Britain. For the 100 km square scale, 20% of each sub-square were selected and predictability then assessed separately for each of the fifty-five 100 km squares. The three percentages were chosen based on the minimum sub-squares needed for estimating 90% and 100% of all species (approximately 20% for 90% of species and 80% for 100% of species, detail see Appendix B). However as 80% sampling is unrealistic in most instances, comparisons were conducted starting at 10% sampling effort (half of 20%).

To summarise, each of the datasets (two orders of sampling each for three levels of sampling efforts) was fitted to three curve functions (i.e. negative exponential, logarithmic and Clench) under two spatial scales (whole British mainland and 100 km square). The procedure of randomly selecting samples or re-arranging according to nestedness order and then extrapolation from fitting curves was repeated 50 times (therefore 50 random samples) and each combination of two orderliness and three levels of sampling efforts for three functions, in two different geographic scales), to access a mean predicting value for species richness. The inaccuracy in the prediction of total species richness, that is, error rate, is defined as:

$$1-(\text{predicted species richness}/\text{observed species richness})$$

A positive value indicates an underestimation while a negative value indicates an overestimation of results. The error rate was correlated with nestedness temperature (at 100 km square scale) using Spearman's correlation test to assess any association between nestedness temperature and error rate. Error rates (absolute value) with nestedness versus random order data were compared using t-tests (compare 50 repeats in whole British mainland) and paired t-test (compare mean results in each 100 km square). The calculation and statistical procedures are conducted in R (R Development Core Team 2013).

2.4 Results

2.4.1 Nestedness of British avian assemblage

The nestedness temperature of Britain as a whole for avifauna is 21.24° while the temperature for each of the 55 squares (100 km x 100 km) ranges from 1.58º to 26.95º, with >40% squares with a temperature in the range of 20º-25º (Figure 2. 3). These indicate a moderate level of nestedness (against the 0º to 100º scale) for Britain and most of the 100 km squares (Wright 1998). Figure 2. 4 illustrates an overview of the presence-absence matrix for the entire British avian assemblage (Appendix A provides detailed axes information for the presence-absence matrix and gives the site and species lists respectively in nested order) and one 100 km square from southern England with a typical nestedness temperature (23.85º). Species generally cluster in the upper-left corner but are more scattered in the lower-right side, indicating a trend of nestedness, with species in more species rich sub-squares (on the top) encompassing those in depauperate sub-squares (on the bottom). The isocline delineates the hypothetically perfect nested structure (Figure 2. 4). The white cells above the isocline are the “unexpected absences”, while the red cells below the isocline are the “unexpected presences”; both reduce the nested structure and increase the nestedness temperature.

The 10 most species rich sub-squares are: ND27, TL87, TF74, TF84, TF62, TQ77, TM47, SZ19, TR06, and SU70, with each sub-square containing >120 avian species. Except for ND27 located in north-eastern Scotland, and SZ19 and SU70 located in southern England, all other sites are in south-eastern England, which is close to the European Continent. In contrast, the 10 most species depauperate sub-squares are: NZ39, NL79, NF61, NA81, NM05, HY35, SY38, SW65, SS11, and TR46. Roughly half of these sites are located at in the far north and the other half in south-western England except for TR46 (Figure 2. 5).

The 10 most commonly occuring avian species among the 2830 sub-squares are: wren (*Troglodytes troglodytes*), skylark (*Alauda arvensis*), pied/white wagtail (*Motacilla alba*), blackbird (*Turdus merula*), robin (*Erithacus rubecula*), starling (*Sturnus vulgaris*), swallow (*Hirundo rustica*), song thrush (*Turdus philomelos*), chaffinch (*Fringilla coelebs*), and willow warbler (*Phylloscopus trochilus*). The 10 most limited occurrence species are: emperor goose (*Chen canagica*), swan goose (*Anser cygnoides*), king eider (*Somateria spectabilis*), helmeted guineafowl (*Numida meleagris*), red-footed falcon (*Falco vespertinus*), little egret (*Egretta garzetta*), black swan (*Cygnus atratus*), northern bobwhite (*Colinus virginianus*), white-winged black tern (*Chlidonias leucopterus*), and smew (*Mergellus albellus*).

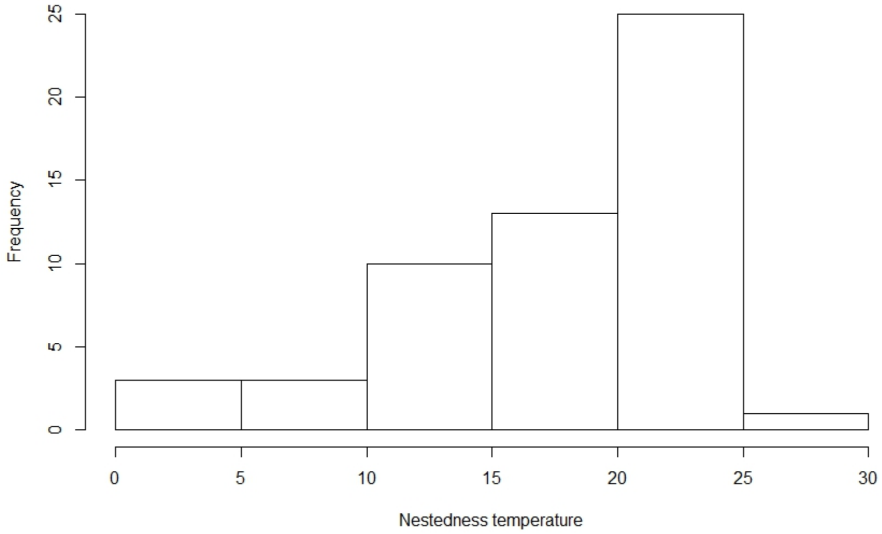


Figure 2. 3 Frequency distribution of nestedness temperature for the fifty five 100 km square of Britain.

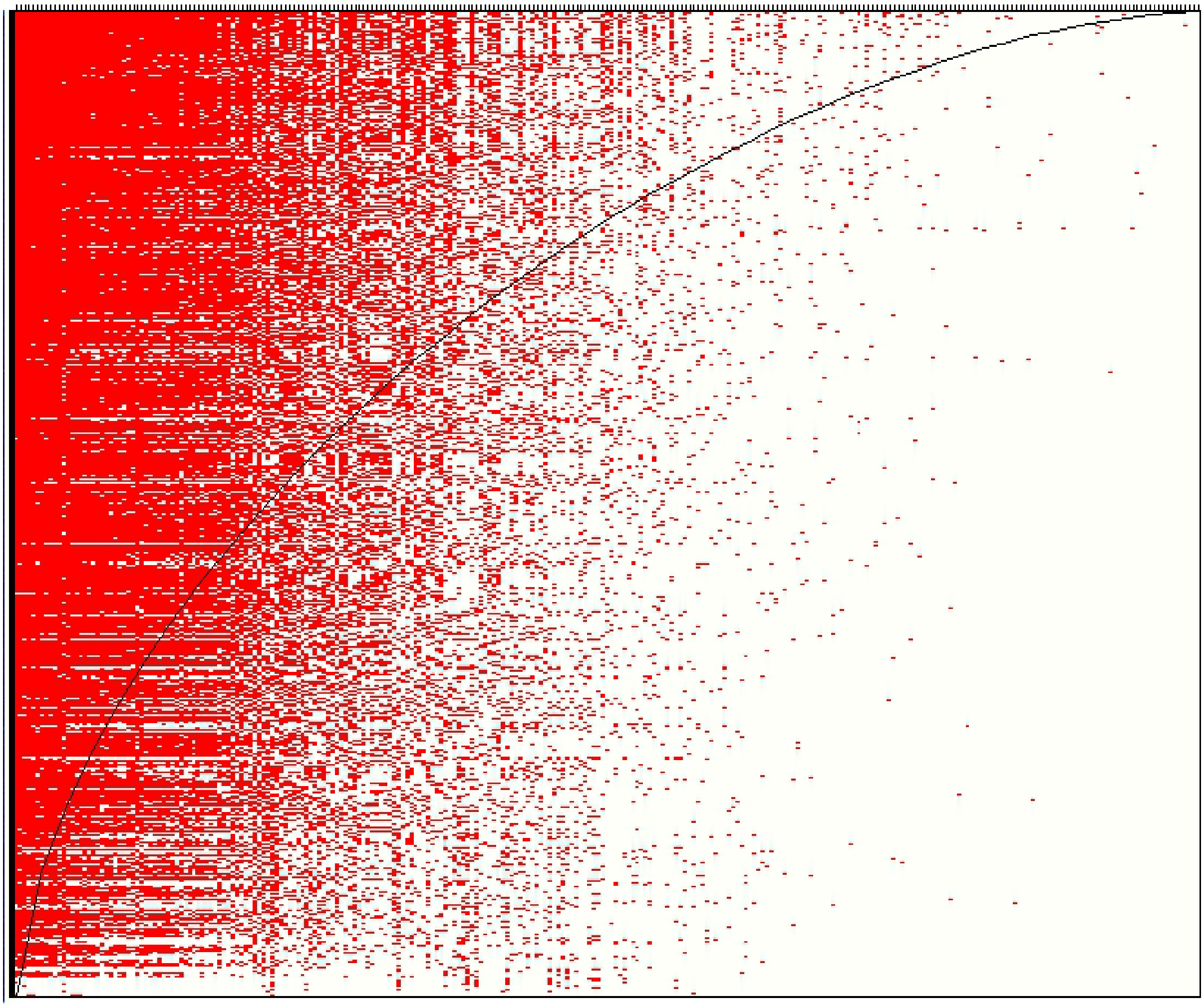


Figure 2. 4 Presence-absence matrix of British avian assemblage. The x-axis represent sites (sub-squares) and the y-axis represent species. No site or species names are shown here as there are not legible in this figure; Appendix A provides full site and species lists in nested oreder for this presence-absence matrix. Red cells indicate species occurrence and white ones, species absence. The black concave diagonal curve represents the isocline that delineates the hypothetically perfect nested matrix. The white cells locating at the left hand side of the diagonal curve represent the unexpected absence and the red ones at the right hand side, unexpected presence.

Figure 2. 5 Species richness at 10 km scale. Darker area is with higher species richness.

2.5 Species accumulation curve and data orderliness

The avian assemblage of entire British sub-squares were arranged according to three orders: nested structure, mean of random order after 100 permutations, and reverse nested structure. Species accumulation curve varied with orders of adding in sub-squares (Figure 2. 6). Curves based on nested structure quickly saturated while curves based on reverse nested structure increased slowly and barely saturated. Random sampling curve was situated between the curves based on nested and reversed nested structure which represented the upper and lower limit, respectively (Figure 2. 6).

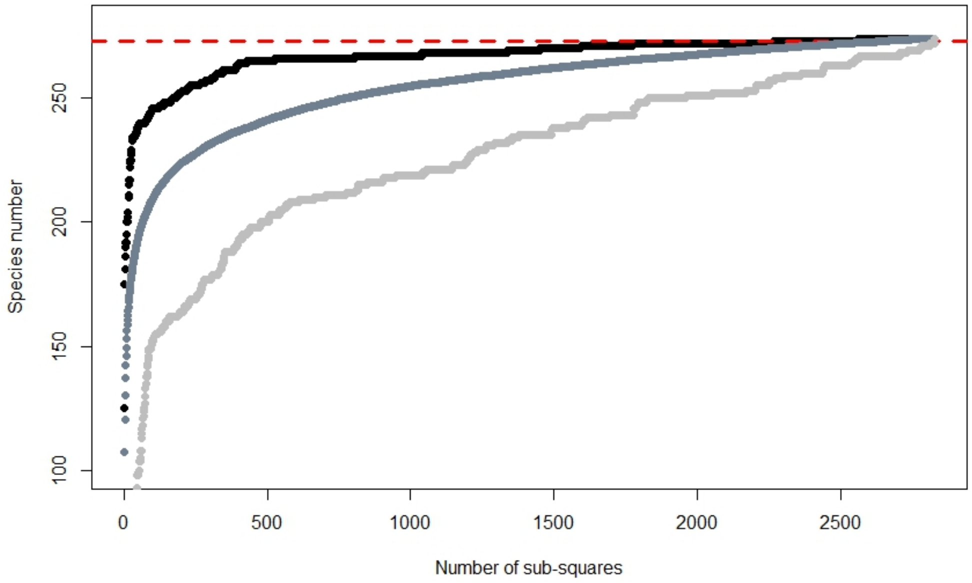


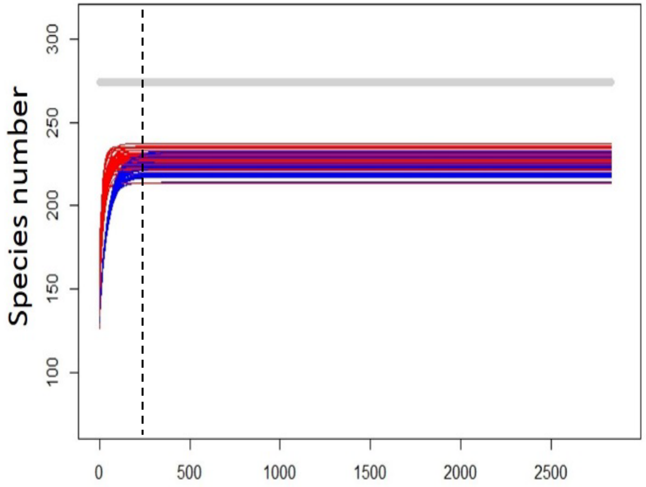
Figure 2. 6 Species accumulation curves based on three orders of data arrangement. Black dots (●): data arranged by nested order, species richness from high to low; dark grey dots (●): mean of 100 randomly arranged data; light grey dots (●): data arranged by reverse nested order, species richness from low to high. The red dash line indicates the real total number of species (273 species).

2.6 Species richness extrapolation for Britain

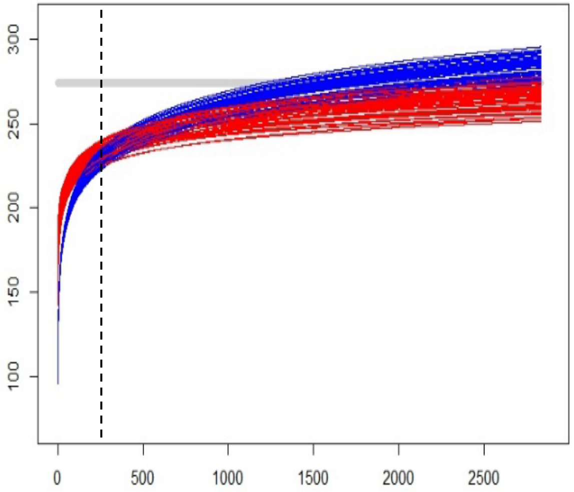
Because predictability based on reverse nestedness for all three species accumulation curve functions was 20-30% less efficient than random and nestedness orders, only results for random and nestedness orders are presented in the following section. Total species richness for Britain was predicted based on three species accumulation curve functions (exponential, logarithmic, and Clench) with three level of sampling efforts (10%, 20%, and 40% of total sites) under two data arrangement orders (random and nestedness) each with 50 replicates. Figure 2. 7 depicts species accumulation curves and Figure 2. 8 summarises the predicted total species richness. For all three levels of sampling effort, projections based on the logarithmic function was more accurate than those based on exponential and Clench functions irrespective of data arrangement orders. However, predictions with random data order tended to overestimate total species richness whereas those with nested order were more likely to underestimate total species richness when based on the logarithmic function (Figure 2. 7 and 2. 8). There was significant difference in error rate (absolute value, $p < 0.05$) between random and nestedness orders for exponential or Clench prediction function but not for logarithmic function ($p > 0.05$). Overall, the method based on nestedness order had lower or equal error rate in the prediction of species richness than that based on random order.

Error rate decreased with sampling effort for both nestedness and random order (Figure 2. 8). Among the three functions, accuracy based on logarithmic curves increased the least when sampling efforts doubled from 10% to 20% or from 20% to 40%, with < 2% of decrease in error rates for both random and nestedness order. This was mostly due to the already high accuracy under the 10% sampling efforts (3.5% error rate, absolute values) and the 20% sampling efforts (<2.7% error rate, absolute values). In comparison, although doubling sampling efforts comparatively improved predictability of exponential and Clench functions, error rates were higher than 7% even under the 40% sampling effort.

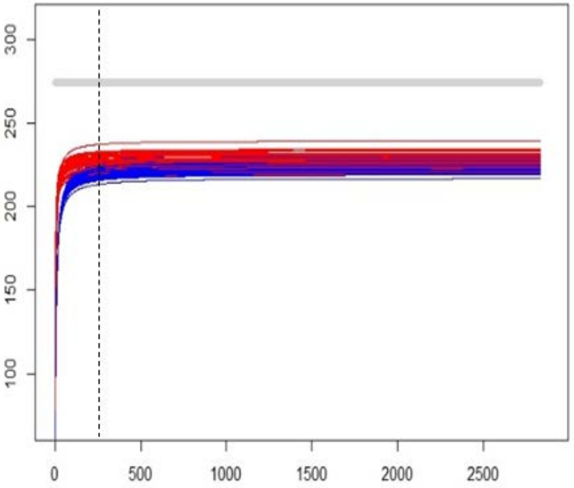
(a.1)



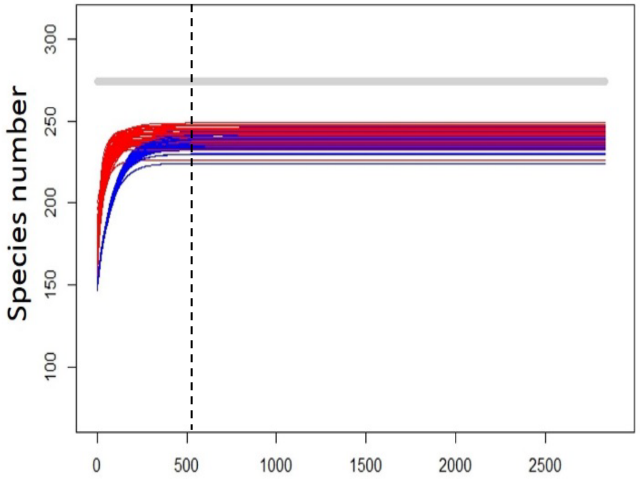
(b.1)



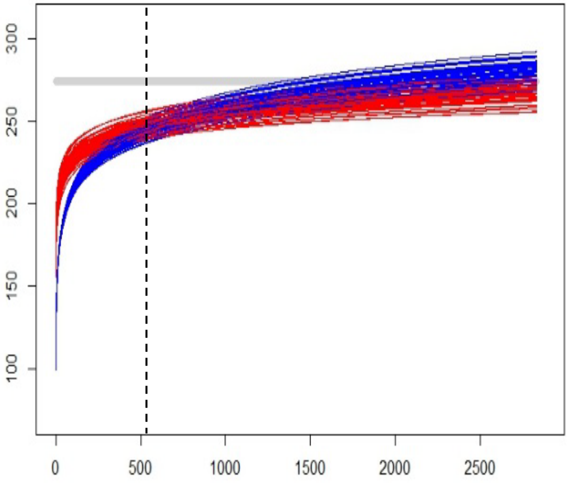
(c.1)



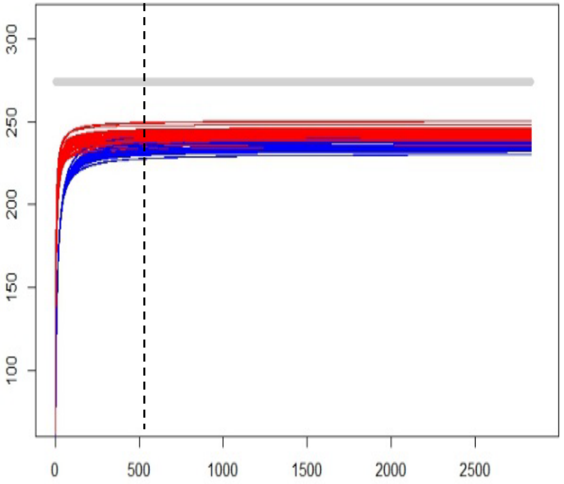
(a.2)

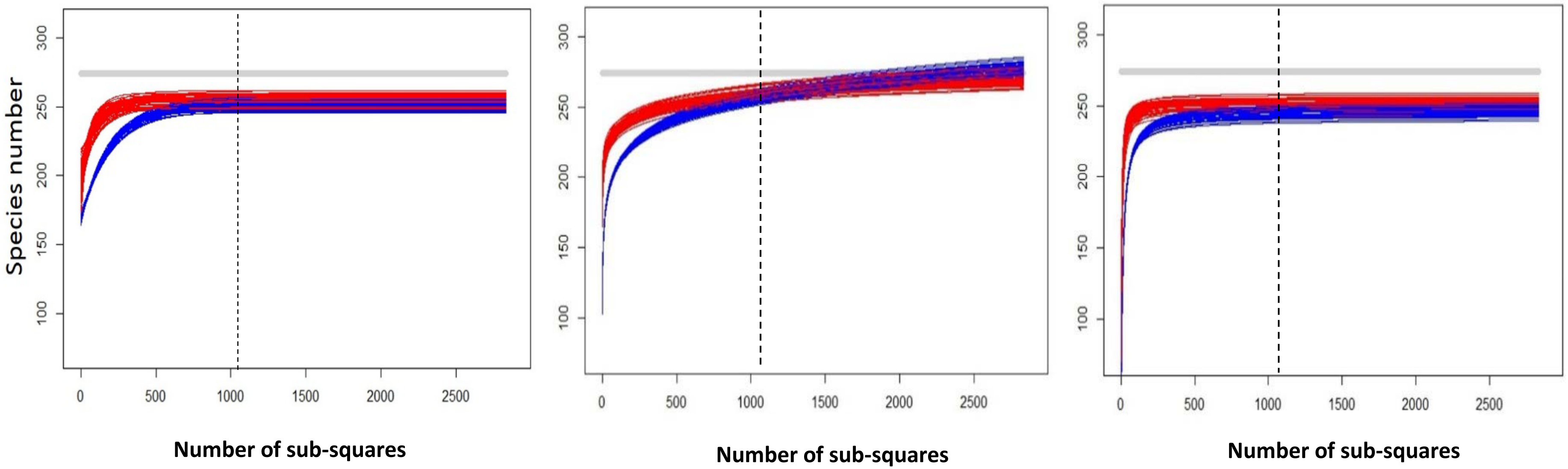


(b.2)



(c.2)





(a.3) (b.3) (c.3)

Figure 2. 7 Prediction of whole British avian species based on three species accumulation curve functions (a: exponential; b: logarithmic: c: Clench) with three levels of sampling effort (1:10%; 2: 20%; 3: 40%, of total sites) under two data arrangement orders (blue: random; red: nested). Each prediction was repeated 50 times. The grey herizontal line indicates the real total number of species (273 species). The vertical dash line indicates the number of sampling sites.

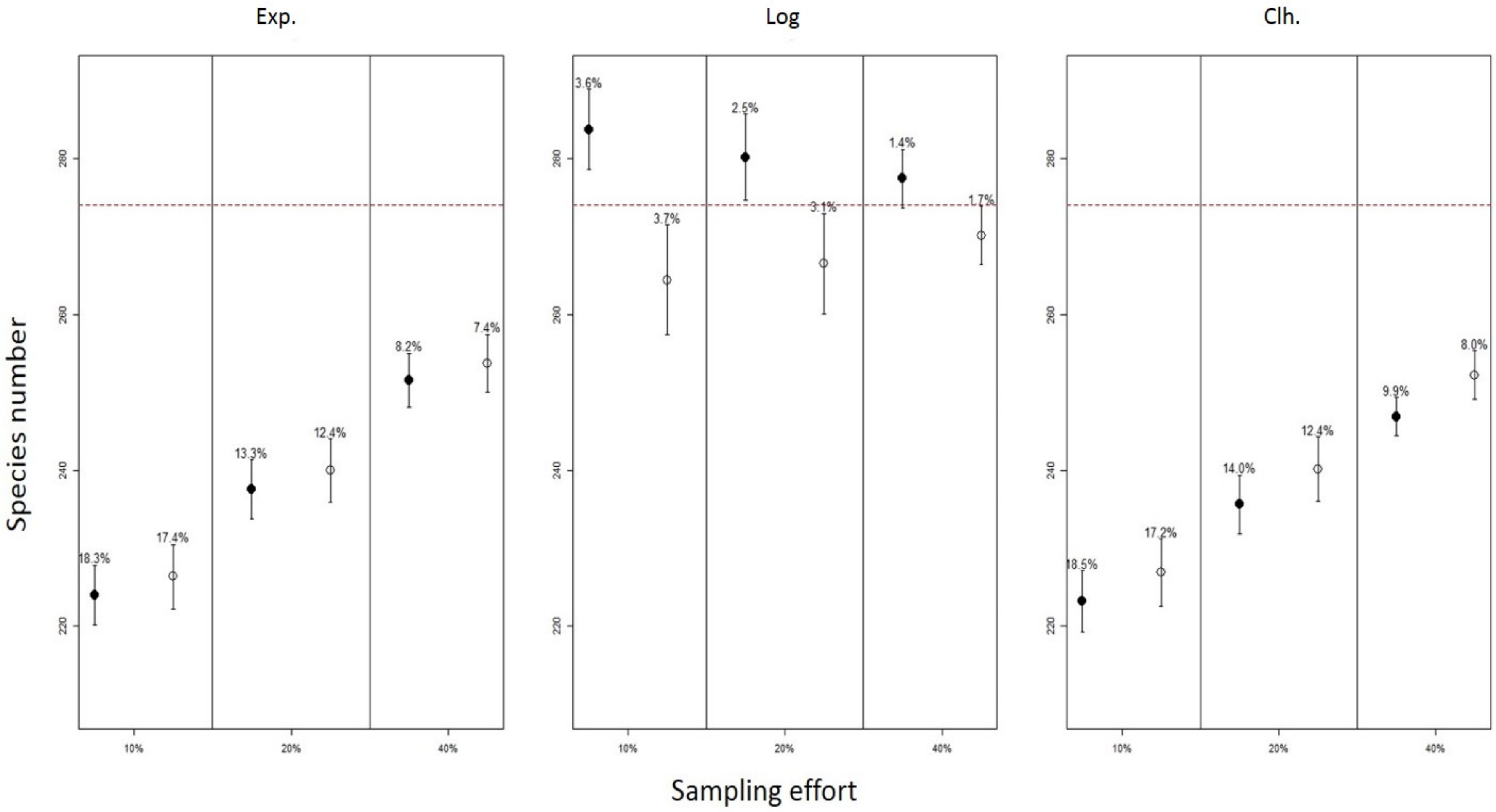


Figure 2. 8 Predicted total species richness (±sd) for a combination of two dataset orders (solid circle: random order; open circle: nested order), three prediction functions (*Exp.* = negative exponential function; *Log* = logarithmic function; and *Clh.* = Clench function), and three sampling

efforts (10%, 20%, and 40%). The percentage above the bars represented the mean error rate. Red dash line represented the observed species richness (273 species).

2.7 Species richness extrapolation for 100 km squares

The total species richness of each of the fifty-five 100 km squares was also predicted based on three species accumulation curve functions with three level of sampling efforts under two data arrangement orders each with 50 replicates.

Under nestedness order, logarithmic function had lower error rates (4.6±5.3%, 3.4±3.5%, 1.4±2.7% for 10%, 20%, 40% sampling efforts, respectively) than exponential (20.9±2.9%, 13.7±11.7%, 9.6±2.2%) and Clench functions (19.0±2.3%, 14.8±3.1%, 9.0±1.9%) for all three levels of sampling efforts. Error rates did not differ with geography (e.g. northern vs. southern, coast vs. inland), but squares containing full sub-squares (100 sub-squares) tended to be more accurate than squares in the periphery of Britain that encompassed < 100 sub-squares (Figures 2. 9, 2.10, 2. 11). There was no correlation (Spearman's correlation coefficient r= -0.69 to 0.04, $p>0.05$) between nestedness temperature and error rates among the 55 squares; however, the error rate increased with lower nestedness when it approaches random order.

Similarly, under random order, logarithmic function had lower error rates (-12.0±6.8%, -7.0±5.3%, 3.9±3.2%) than exponential (20.5±2.2%, 9.1±30.8%, 10.4±1.8%) and Clench functions (14.9±2.5%, 12.5±4.8%, 9.6±2.3%) when the comparisons were based on absolute values. There was no correlation (r= -0.66 to 0.38, $p>0.05$) between nestedness temperature and error rates among the 55 squares; however, the error rate increased with lower nestedness when it approaches random order.

Error rates varied considerably among different combinations of arrangement orders, curve functions and sampling efforts (Table 2.1). However, error rates were generally lower when based on nestedness order than when based on random order ($p<0.05$), especially under logarithmic function that had much lower error rates than exponential and Clench functions (Table 2. 1).

Table 2. 1 Comparison of error rate between nestedness order (N) v.s random order (R) under three functions (Exp. = negative exponential function; Log = logarithmic function; and Clh. = Clench function) and three levels of sampling efforts (10%, 20%, and 40%) at 100km x 100km scale. Bold type indicates significant differences at p=0.05 level (with absolute value).

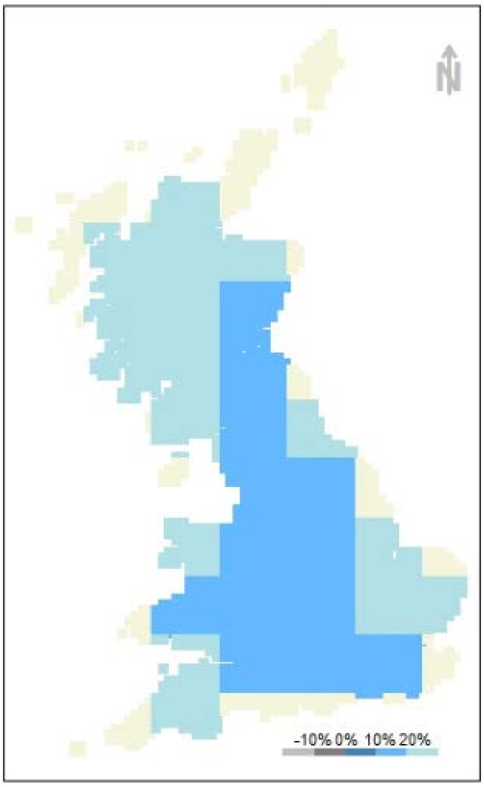
Functions	*Sampling effort*	*Error rate** *N*	*R*	*t***	*df*	*p-value*
Exp.	10%	20.9±2.9%	20.5±2.2%	1.05	28	0.30
	20%	13.7±11.7%	9.1±30.8%	-1.29	41	0.20
	40%	9.6±2.2%	10.4±1.8%	-2.45	42	**0.02**
Log	10%	4.6±5.3%	-12.0±6.8%	-4.53	35	**<0.01**
	20%	3.4±3.5%	-7.0±5.3%	-3.85	39	**<0.01**
	40%	1.4±2.7%	3.9±3.2%	-4.05	42	**<0.01**
Clh.	10%	19.0±2.3%	14.9±2.5%	7.52	28	**<0.01**
	20%	14.8±3.1%	12.5±4.8%	3.84	35	**<0.01**
	40%	9.0±1.9%	9.6±2.3%	-3.43	42	**<0.01**

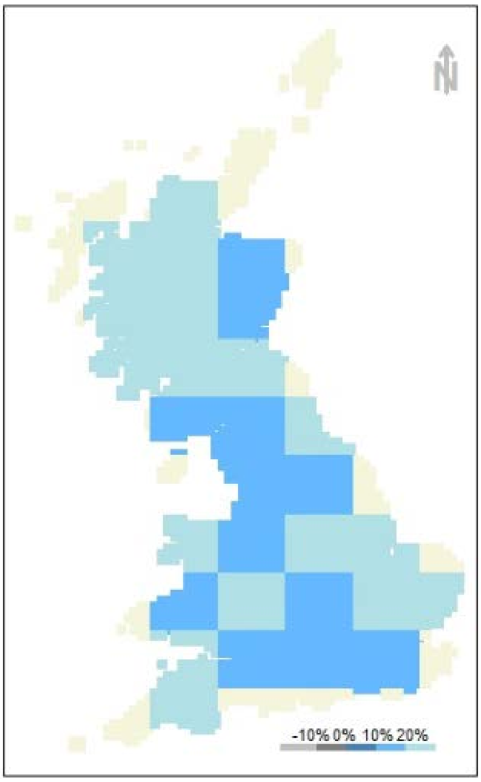
*positive value indicates an underestimation while negative an overestimation of species richness

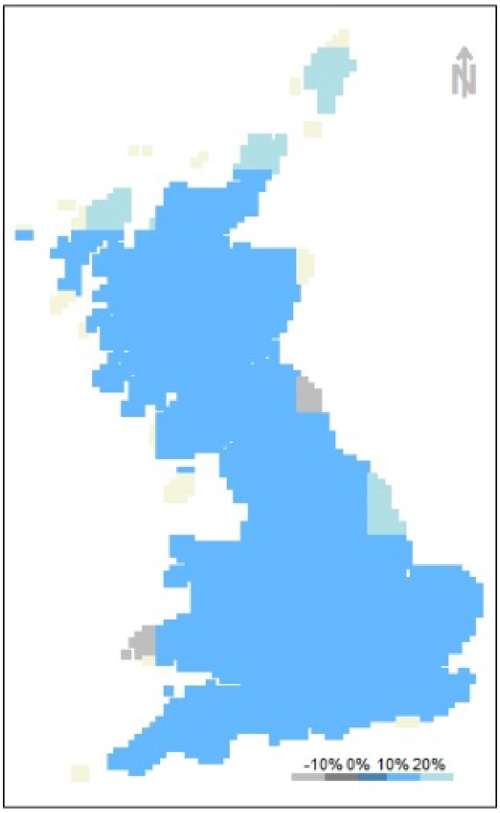
**positive indicates higher error rate for nested order and vice versa (comparison based on absolute value)

Random

Nested







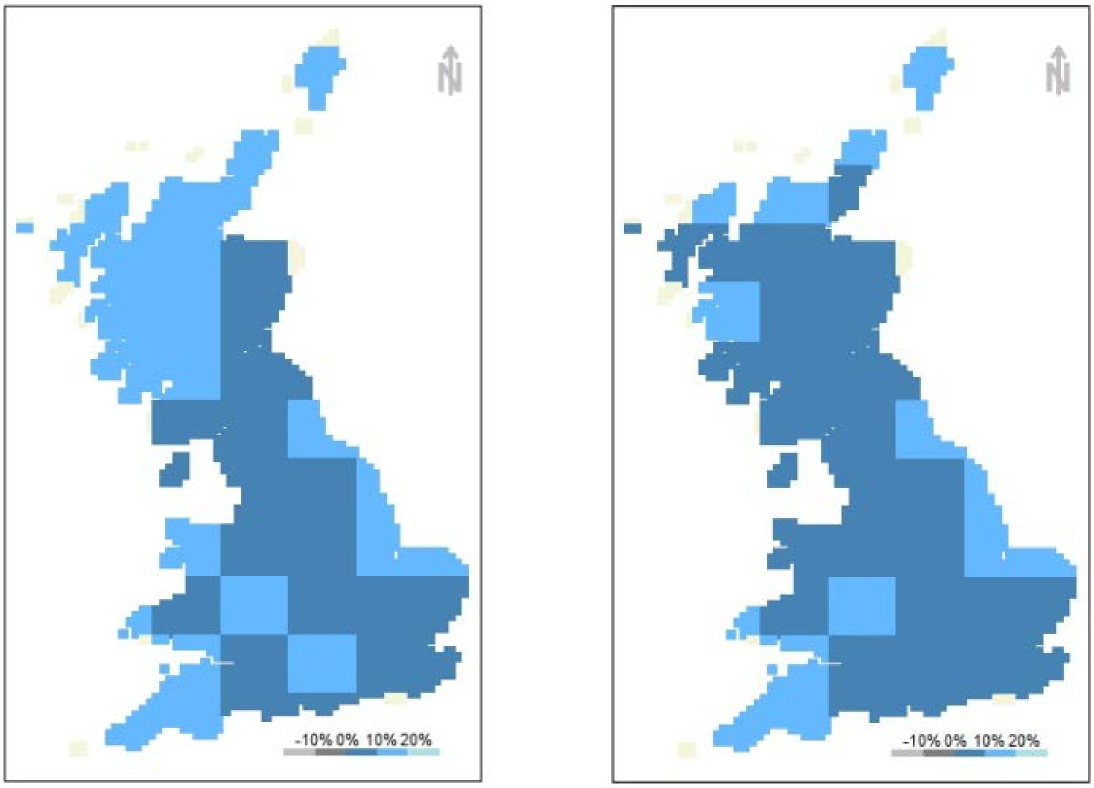
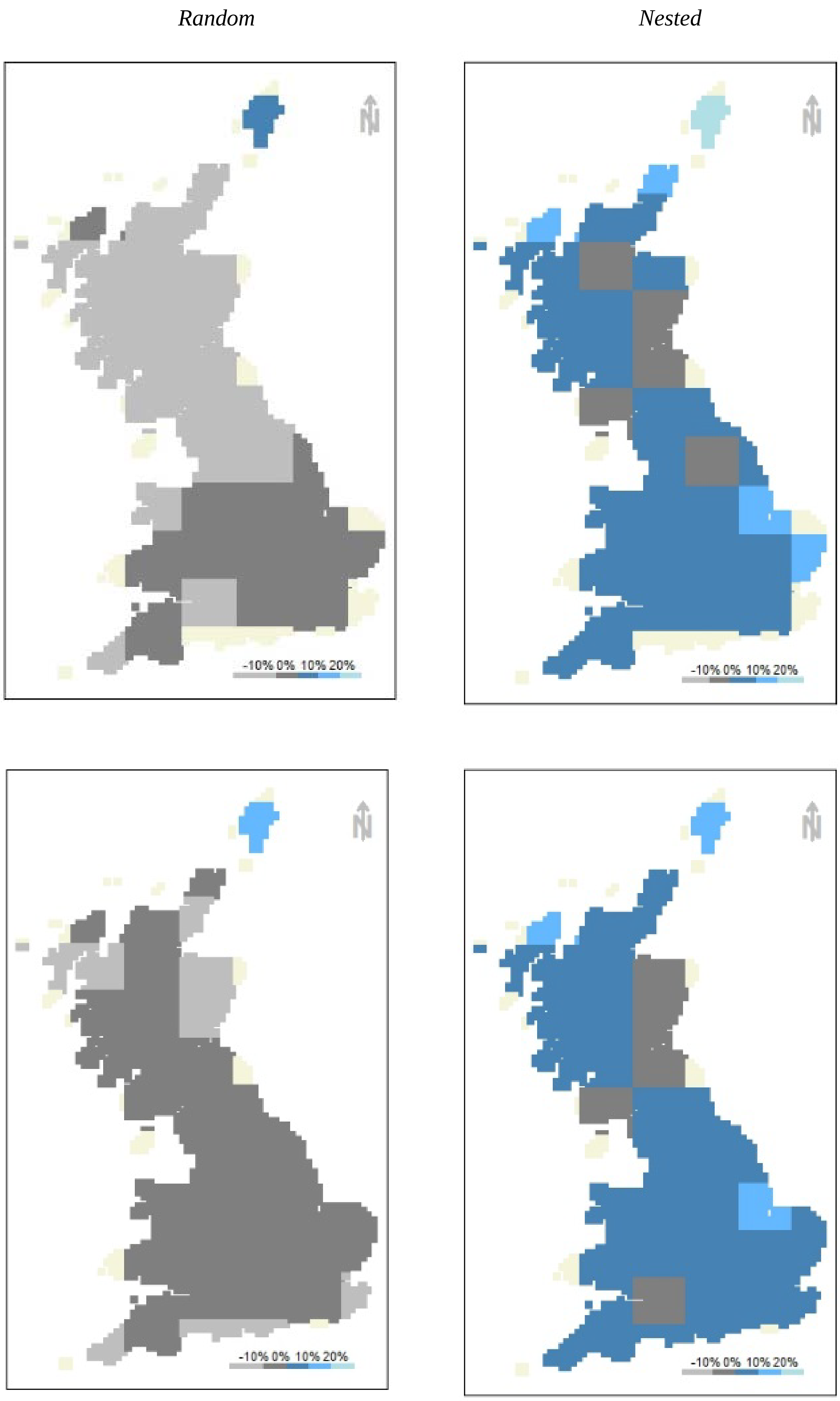


Figure 2. 9 Error rate (mean of 50 repeats) based on exponential accumulation curve function for fifty five 100 km squares with three level of sampling efforts (1st row: 10%; 2nd: 20%; 3rd: 40% of total sites). The left column shows results by random order and right column, nestedness order. Blue areas indicate underestimation and grey areas overestimation of prediction results. Darker colours represent better predictive results (lower error rates). Beige areas indicate sub-squares without enough data for extrapolation.



Random *Nested*

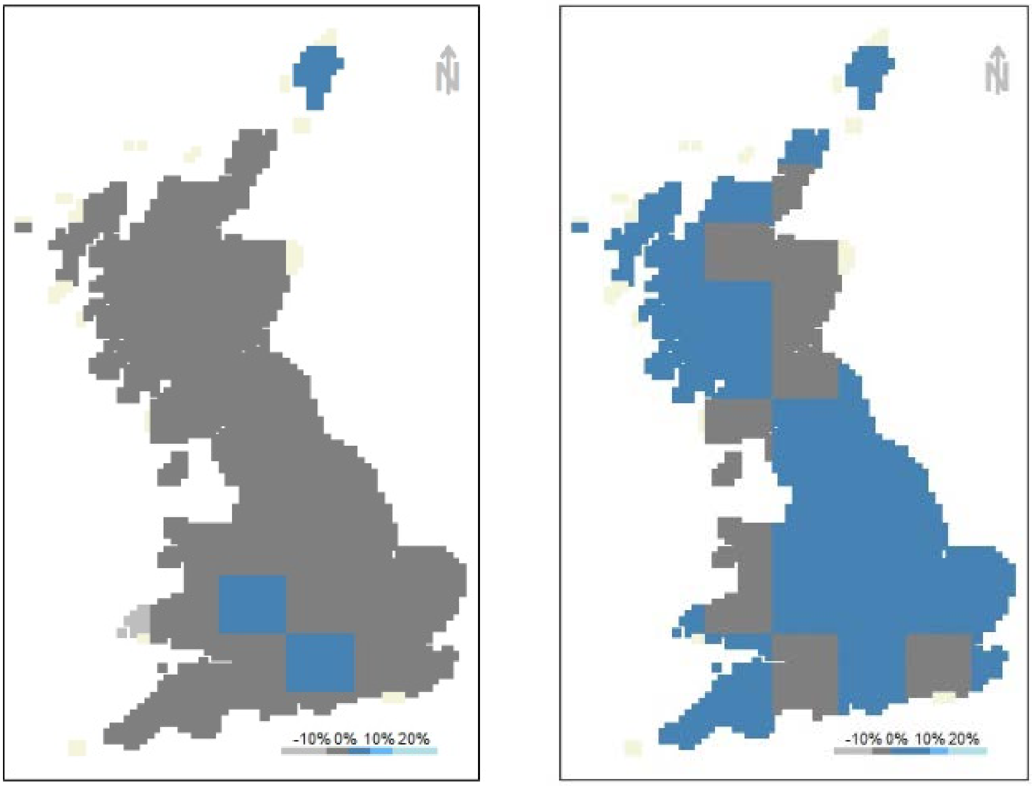
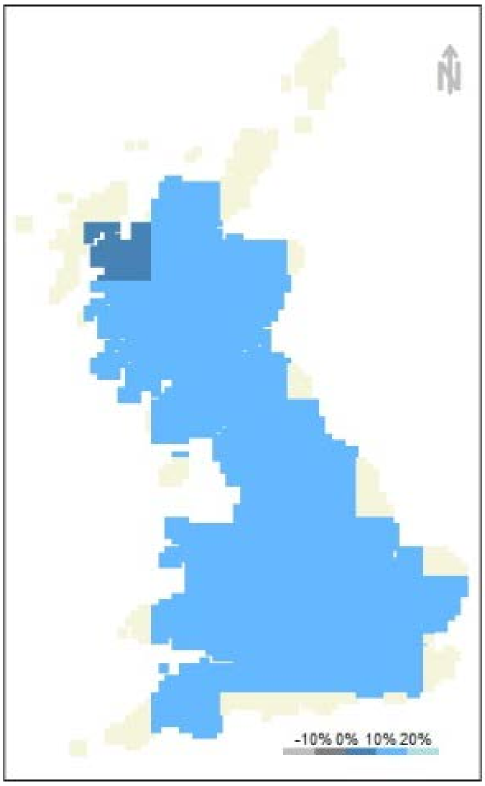
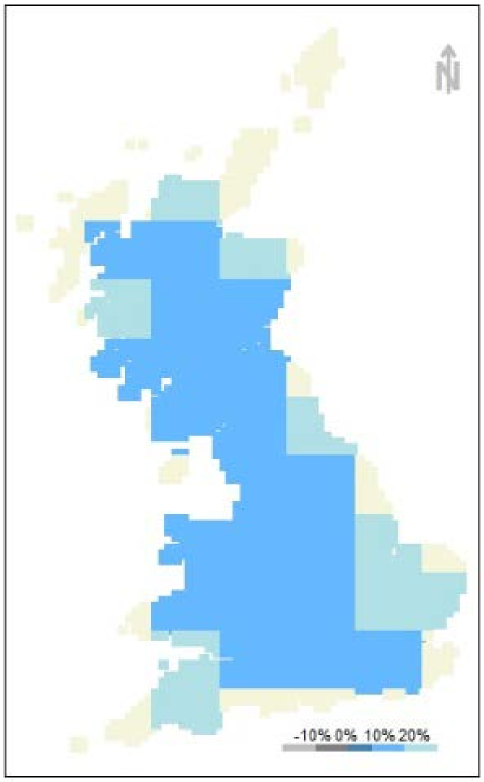
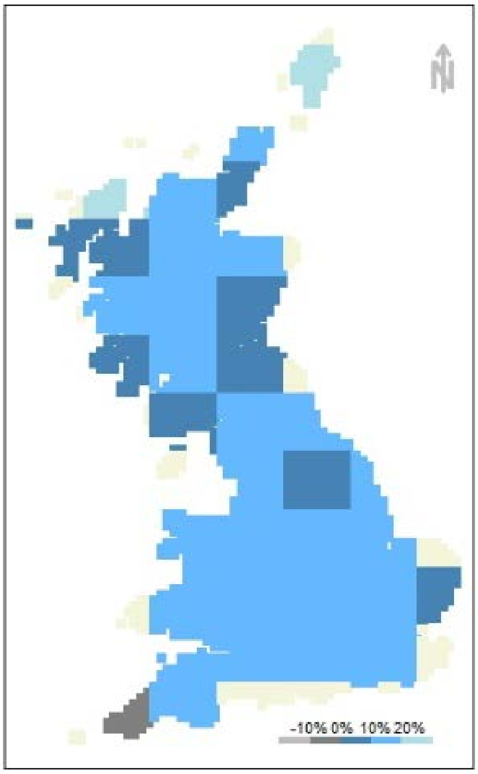


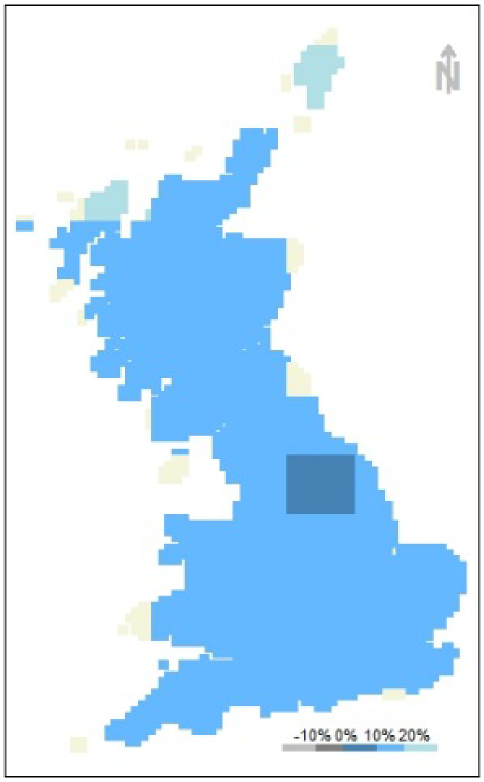
Figure 2. 10 Error rate (mean of 50 repeats) based on logarithmic accumulation curve function for fifty five 100 km squares with three level of sampling efforts (row 1st: 10%; 2nd: 20%; 3rd: 40% of total sites). Left column showed results by random order and right column, nestedness order. Blue areas indicate underestimation and grey areas overestimation of prediction results. Darker colours represent better predictive results (lower error rates). Beige areas indicate sub-squares without enough data for extrapolation.

Random *Nested*









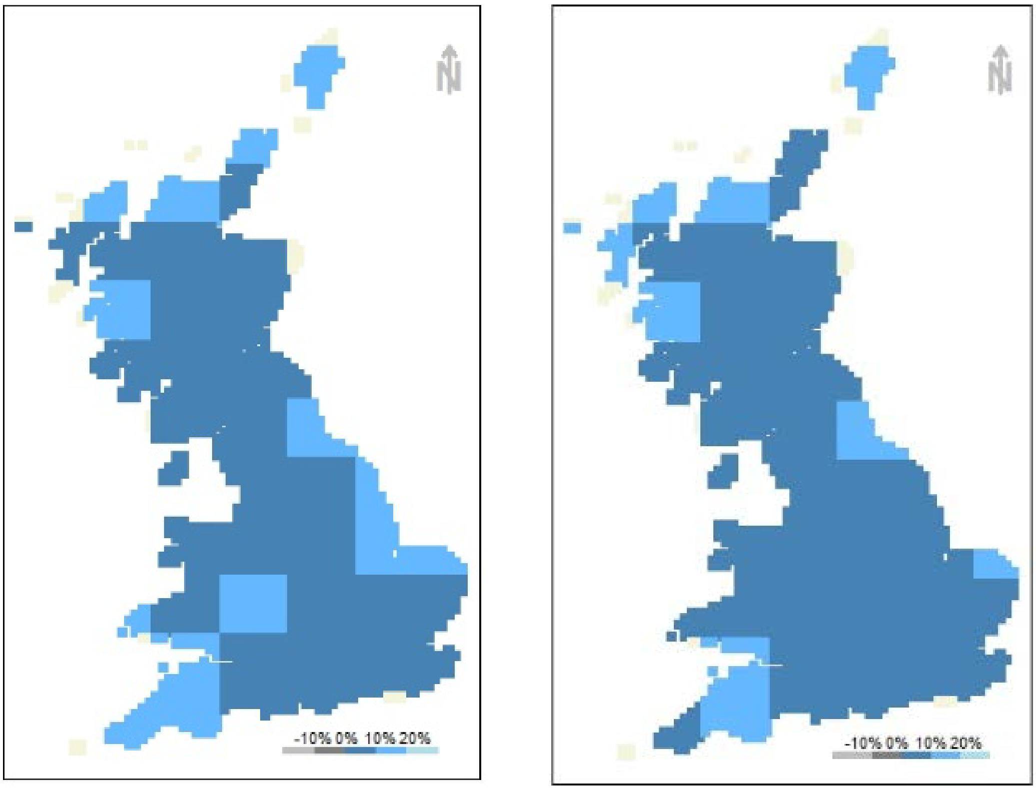


Figure 2. 11 Error rate (mean of 50 repeats) based on Clench accumulation curve function for fifty five 100 km squares with three level of sampling efforts (row 1st: 10%; 2nd: 20%; 3rd: 40% of total sites). Left column showed results by random order and right column, nestedness order. Blue areas indicate underestimation and grey areas overestimation of prediction results. Darker colours represent better predictive results (lower error rates). Beige areas indicate sub-squares without enough data for extrapolation.

2.8 Discussion

In the current study of the British avian assemblage, it was found that accuracy in the projections of species richness increased with sampling effort, and error rates were much lower when fitted with logarithmic function than with exponential or Clench function for both British and 100 km square spatial scales (Figure 2.8, Table 2.1). More importantly, when fitted with the best performing logarithmic function, prediction was at least as good or better when data were arranged based on nestedness order than based on random order (Figure 2.8, Table 2.1).

2.8.1 Species accumulation curve with nestedness structure

Extrapolation from species accumulation curves is a commonly adopted and relatively simple solution to estimate species richness with limited information. Approaches based on asymptotic estimators (Walther and Moore 2005, Magnussen 2014, Béguinot 2015, Chao and Jost 2015, Gwinn et al. 2015) or nonparametric extrapolation from species accumulation curves (Colwell et al. 2012, Chao et al. 2014) require further information on species abundance (instead of presence-absence data only) and are considered to be more accurate in estimating species richness (Brose et al. 2003, Cayuela et al. 2015) than curve fitting method that could be estimated based solely on presence-absence data. The biggest challenge in curve fitting method comes from selecting the optimal extrapolation function (Soberón and Llorente 1993, Dengler 2009b) or randomisation process (Ugland et al. 2003, Gray et al. 2004), which might be improved after the incorporation of information on community structure. In this study, we have demonstrated that ordering data based on nestedness could potentially improve predictability. Most recent research on nestedness has focused on describing community structure (Picazo et al. 2012, Herczeg and Horvath 2015, Chaves and Ariez 2016) or how to improve the algorithms of calculating nestedness indexes (Almeida-Neto et al. 2008, Podani et al. 2014). This study is a new attempt to incorporate nestedness index to improve the predictive power of the species accumulation curve fitting method. Compared with random order, inputs based on nestedness order would enhance the ability of species accumulation curves to predict avian species richness in Britain at the 100 km square scale, especially coupled with logarithmic function. Accuracy in the prediction of total species richness for both the entire British island and 100 km squares could be >95% for 10% sampling effort when logarithmic species accumulation curve function are fitted with data based on nestedness order. In comparison, the accuracy was less than 90% when data were arranged randomly for 10% sampling effort under logarithmic function at spatial scale of 100 km square.

2.8.2 Nestedness application in sampling design

Owing to the limited resources available for field surveys, it is worthwhile to improve the efficiency of sampling efforts (Ashcroft et al. 2010). Both mathematical pattern and biological mechanism can potentially improve the efficacy of sampling effort. For example, Pearman and Weber (2007) found that including widely distributed species can better predict overall richness than when species of limited distribution were contained. The nestedness-order dataset applied in this study can similarly lead to a better prediction when combined with the best fit extrapolation functions. For instance, error rate was <5% based on nestedness order compared with 12% based on random method when logarithmic functions were applied for the estimation of British avian species richness in 100 km squares with 10% sampling effort.

Conventionally, intensive randomisation procedures are needed for the species accumulation curve to achieve higher predictive power (Ugland *et al.* 2003; Chao & Shen 2004; Gray *et al.* 2004). The randomisation procedure reduces the probability of including extreme values and attempts to simulate real conditions by considering a large number of replications. Alternatively, sampling based on nestedness order is a novel attempt for capitalising on community structure for the prediction of total species richness. Results of this study have demonstrated that overall, sampling according to nested order is more favourable than based on random order. Melo *et al.* (2003) compared predicted species richness from species accumulation curve in different taxa and indicated at least 40% of sampling effort is needed to acquire above 90% accuracy. The nestedness method applied in this study captured information about assemblage composition and used it to achieve better accuracy with lower sampling effort. Thus taking advantage of the nested structure can potentially help predict assemblage composition by focusing on information in the most species-rich sub-squares. However, these findings may be related to the fact that the British avian assemblage demonstrates a moderate level of nestedness at two spatial scales: the entire British island and 100 km squares. More studies considering different species and study site characteristics, as well as different levels of nestedness, are needed to test for the general applicability of nested order.

2.8.3 Species accumulation curve functions for British avian species

Selecting appropriate species accumulation curve function to predict species richness is difficult, as there is no universal principle for the selection of the optimal function. Ugland *et al.* (2003) suggested that the exponential model is suitable for small scale data, while Díaz-Francés and Soberón (2005) proposed that the power model and logistic model are ideal for median and larger scale data or when information on taxa is scarce. This study indicates that predictability does vary remarkably with functions and selecting the optimal function would greatly improve the predictability. Soberón and Llorente (1993) compared predictability of negative exponential, logarithmic, and Clench functions (but not incorporating information on nestedness) and also concluded that functions vary in their performance. For instance, negative exponential function tends to approach asymptote too soon and typically underestimates overall species richness; however, the same function would also avoid species overestimation under homogenous environment. Indeed, negative exponential function fitted Mexican butterfly data better than logarithmic and Clench functions and this was likely due to the relatively homogenous environment in the Mexican study (Díaz-Francés and Soberón 2005).

The fact that negative exponential function, along with Clench function, usually perform better under homogenous environment is due to both functions belonging to the saturation model (Dengler 2009). That is, there is an upper asymptote and the accumulation curve eventually approaches a saturation point. Compared with negative exponential function, Clench function fits better when homogenous area contains more rare species (Soberón and Llorente 1993, Moreno and Halffter 2000) although both functions tend to underestimate overall species richness (Soberón and Llorente 1993, Cardoso et al. 2008). In comparison, logarithmic function creates a unbounded curve without mathematical asymptote and is more suitable for heterogeneous environment despite this function tending to overestimate species number due to an infinite increase (Tjørve 2003, Dengler 2009).

In this study, logarithmic function performed the best for both the whole British island and 100 km square spatial scales. This agrees with the findings of Lennon *et al.* (2001) in which they assessed the British avian assemblage and found that the logarithmic function, on average, predicted total species richness better than the power function for 10 km to 90 km spatial scales. Because the whole British island and most of the 100 km squares also cover coastal areas (Figure 2.2) that contain water birds, both spatial scales are relatively heterogeneous in the species composition (i.e. they contain both water and land birds), hence the better predictability of logarithmic function than negative exponential or Clench function. In this study, the basic unit of assemblage is the grid square, which is an artificial unit, but the constant size of each grid square avoids the confounding effects of area on species richness and allows a comparison of input order in the species accumulation curve. Besides, if nestedness order can enhance predictability in artificial square data (which may contain more than two real communities), it might perform even better in real communities. In other words, the advantage of nestedness order might be underestimated when based on the artificial square grid data here.

Logarithmic function tends to overestimate species richness (Tjørve 2003, Dengler 2009). Species overestimation does occur in the current study when the prediction is based on logarithmic function with conventionally applied random order. However, replacing random order with nested order instead underestimates total species richness. This is probably because when compared with random order, data input based on nested order include proportionally more uncommon species in the initial samples (Figure 2.1), followed by a decelerating rate of increase in species richness (Figure 2.6, 2.7) because fewer uncommon species remain undetected. Such deceleration might lead to an underestimation of total species richness. On the contrary, fewer species are included in the initial samples when based on random order (Figure 2.6, 2.7). This is later compensated by an accelerating rate of increase, causing an overshoot in the prediction of total species richness. Owing to similar but opposite levels of error rates with random versus nestedness order for logarithmic function under British scale (Figure 2.8), it is interesting to test whether combining results from both random and nested order leads to better predictions. Indeed, a test showed that predicted species richness can be as low as 0.07% using logarithmic function in 40% sampling at British scale due to the error rates cancelling each other out when using the mean of both estimates. Further studies should test whether incorporating both random and nestedness orders further increases success in the prediction of species richness.

Relative to random order, data input based on nestedness order have better predictive power at small spatial scale (100 km square) when both are fitted with logarithmic function (the most accurate function). Such superiority in prediction does not occur under large spatial scale (e.g. British island), where both ordering approaches lead to similar predictability. Comparatively, new species are more likely to be encountered in larger areas (Ney-Nifle and Mangel 1999, Tjørve 2003, Ulrich and Buszko 2007), thus increasing the uncertainty of including new species and decreasing the prediction accuracy. Under such circumstance, input based on nestedness order might not be more favourable than when based on random order. In comparison, the chance of including new species is lower in smaller areas, especially when the species-rich sites are included in the first samples (i.e. nestedness order), thus the higher prediction accuracy of nestedness ordered data input. However, whether the superiority of nested order is scale–dependent deserves further investigation.

Chapter 3: Incorporating network analysis in the identification of indicator species of the British avian assemblage

3.1 Abstract

Predicting species identity instead of species richness alone is of significant value particularly with species of conservation concerns. The indicator species approach is commonly applied to predict species composition in unsampled sites, but the conventional clustering method can often lead to inconsistent and misleading results due to that this method measuring the relative dissimilarities among species. In this study, we applied network analysis that directly measures co-occurrence among species to improve the predictive power of indicator species, with British birds as an example. Five indicator species groups identified among British avian species across all sites differed between two approaches of site group detection, clustering method and network analysis, with the latter showing clear geographical demarcation. Selected indicator species based on 40% study sites also differed between clustering method and network analysis; however, species in unsampled sites were better predicted by network analysis than by clustering method. Our study demonstrates that incorporating network analysis can improve predictability of indicator species and this novel method can be of broad applicability to other study systems.

3.2 Introduction

Results from the previous chapter showed that incorporating information on assembly structure, such as nestedness, can improve predictability of species richness in comparison with capitalising simply on mathematical equations that lack any biological underpinning. Based on the aforementioned approach, only species richness but not species composition can be predicted; nevertheless, information on species identity is frequently more important than species richness, particularly with species of conservation concerns (Webb 1989, Cousins 1991). For example, two sites with similar species richness merit different management plans when one harbours species that are endangered or have a limited distribution while the other is composed primarily of exotic species. Furthermore, conservation strategy focusing solely on areas with the highest species richness (i.e. hotspots) sometimes neglects the requirement of rare species that happen not to occur in the hotspot; for example, rare liverworts and aquatic plants have environmental requirements distinctive from other terrestrial flora (Prendergast et al. 1993). Evidently, information merely on species number might be misleading, particularly when concerning conservation or policy legislation (Prendergast et al. 1993, Prendergast 1997, Grundel et al. 2014).

In this chapter, I take a step further and focus on species identity and composition instead of species richness only. Community structure, including species co-existence information, is analysed to see if simply having limited information on species identity but no environmental attributes can predict overall species composition. Compared to niche modelling that requires environmental attributes, using species information solely to predict species composition is based on the assumption that when a group of species is closely associated with specific ecological requirements, this group of species can be taken as “indicator species” to predict characteristics of unsampled sites (Harms et al. 2001, DeCáceres and Legendre 2009). This method is called “indicator species value” and was developed by Dufrêne and Legendre (1997).

An interest in associations between habitat and species assemblage can be traced back to Two Way Indicator Species Analysis (TWINSPAN) (Hill 1979), which is the earliest numerical method on habitat association of species assemblage (Dufrêne and Legendre 1997). However, TWINSPAN has two main weaknesses. Firstly, it is based on correspondence analysis (CA) or detrended correspondence analysis (DCA) that projects sampled species onto an assumed linear-gradient axis, which might not in fact be linear. Furthermore, this analysis retrieves information only from the first axis of CA or DCA and neglects the second axis which might also affect the ordination result. Secondly, TWINSPAN is a top-down division process, clustering and arbitrarily dividing the first axis of CA (or DCA) into two even groups. This process might misclassify species near the centre of the axis. Moreover, the two-group method may not always be suitable for classification, for example in categorizing a xeric to hydric gradient into either group because in-between there exists mesophytic plants. Therefore, classification by TWINSPAN may be inappropriate and lead to low predictability when associations between species and habitats are weak. To solve these problems, a method termed “indicator species value” (Dufrêne and Legendre 1997) was devised to provide a more efficient way for identifying indicator species.

The indicator species concept has roots in the idea of representative diversity (Webb 1989, Cousins 1991), which argues that species assemblages are associated with specific habitat types. Such species assemblages help characterise habitats based on Species–Environment Relationships (SER) analysis, which categorizes species assemblage according to spatial and environmental factors (Borcard 1992). The SER analysis also involves measuring environmental attributes and relies on multivariate methods for modelling species distribution.

Instead, the indicator species value method stresses measuring association between species and habitats, using hierarchical or nonhierarchical classification method (e.g. clustering) (Dufrêne and Legendre 1997). In contrast to TWINSPAN, indicator species are not classified along a gradient but are selected based on two characteristics: 1) the degree to which the species is associated with specific habitat. The ideal situation is that a species only occurs in one specific habitat, thus reflecting its ecological preference or requirement. This character is usually designated as A in indicator species value calculation; 2) the abundance of the species; that is, is it abundant enough to be effective as an indicator. Although a rare species can fit the first requirement, its lower numbers mean a lower detection rate. This is usually designated as B in the calculation. These two indices can be applied and interpreted individually or jointly. According to Dufrêne and Legendre (1997), the indicator value is the product of A and B and a higher value suggests more feasibility as an indicator species. In contrast to species distribution modelling that relies on environmental variables to model species richness or composition, the indicator species method depends solely on information inherent in species groups (DeCáceres and Legendre 2009). For example, research on carabid spatial distribution shows that this taxon has strong habitat preference and clear distributional boundaries. Based on the indicator species value, a new habitat type was identified and indicator species were found to possess species indicator values (Magura 2002).

Identifying indicator species requires classifying sites (or samples) into different habitat types that each contains unique species composition. Each species will then be assigned an indicator value for each habitat type. For instance, if three habitat types are classified, each species has three indicator values corresponding to each of the three habitats. The indicator species for each habitat are defined as the species with the highest indicator value. However, the habitats are typically classified based on clustering methods (especially non-hierarchical methods such as k-means clustering), which is a descriptive method with its robustness hard to validate and it is also an indirect method, relying on distance (difference) between species rather than directly measuring co-occurrence patterns (Milligan 1980, Legendre and Legendre 2012). It is also sensitive to the double-zero issue, which is very common in ecological datasets and may cause group misclassification (Dufrêne and Legendre 1997, Legendre and Legendre 2012). A double-zero issue happens when comparing two sites with the presence or absence of specific species. Presence in both sites of species indicates resemblance of sites, but absence in both sites does not necessarily represent resemblance, for the absence of species in two sites may due to different reasons. However the distance calculations in conventional clustering methods cannot reflect this fact. In comparison, a direct measure of species co-occurrence could improve the efficiency of indicator species method.

Numerous field observations have revealed that species co-occurrence is not a random process but instead one species tends to occur more frequently with certain species (Waddle et al. 2010). Such co-occurrence may reflect similar environmental requirements such as food and habitats, thus forming functional groups (Woodcock et al. 2010, Ehouman et al. 2012). It is also likely to be due to cooperation or mutualism such as the formation of feeding flocks (Farine et al. 2012). If such a relationship is stable through space and time, whole groups of co-occurring species could potentially be predicted when one of the members is observed. Until now, most studies have focused on measuring negative species co-occurrence, such as the checkerboard score (Gotelli 2000). The checkerboard scores emphasise how two species do not co-occur due to competition although predictability of species is likely to be enhanced when instead based on how and how often species co-occur (Gotelli 2000, Gotelli and McCabe 2002). Under such circumstance, network analysis may serve as a valuable tool, because of its efficiency in disentangling complex relationships among members in an assemblage or community.

Network analysis has now been widely used in social, biological, and many other fields to help illustrate and disentangle complicated interactions and relationships among members (Whitehead 1995, Lusseau 2003, Newman and Girvan 2004). It is particularly useful for managing large amounts of data (Blondel et al. 2008) and can usually be visualised to help reveal entangled relationships (Moody et al. 2005). A network contains two essential elements: nodes and edges. Each node is an individual member in a network, and the edge represents relation among them. For example, food webs are one kind of network that is composed of plants and animals (nodes) with each interaction representing an edge. Such complicated interactions, common in ecological studies, are usually difficult to analyse by conventional methods. Network analysis, instead, can tackle the problems even with a community comprising thousands of interactions (Montoya andSol 2002, Estrada 2007, Gilbert 2009).

When information on sites is incorporated into analyses of interactions among species, a specific network type called multi-mode network should be used (here, a two-mode network containing species and sites). Multi-mode networks encompass more than one type of node (Newman 2010). For example, a two-mode network can represent the relationship among species and their resident sites, and there are two types of nodes representing species and sites harbouring the species respectively. The community data can then be arranged into an “incidence matrix” to reveal whether some groups of species coexist more frequently on particular sites than others. Such network analysis reveals relationships among species that co-occur in various sites. When the analysis focuses on only one aspect of a relationship (e.g. relationship among species in the same sites or sites containing similar species group), a two-mode network can be projected into a one-mode network to facilitate analysis (Prell 2011). For example, when the network is projected along the site axis, it emphasises how sites containing similar species can aggregate to form site groups (i.e. subgroup in network analysis) that reflect similar ecological requirements. For instance, sites containing puffins and petrel species form a subgroup that represents rocky-coastal habitat. If the associations between site groups and species are robust and repeatable, the habitat type or species composition of an unsampled site can potentially be predicted based on observed species or habitat. This is in accordance with the concept of indicator species value (Dufrêne and Legendre 1997). Furthermore, network analysis has the potential to improve the predictability of the latter method by its superior analytical capability in (site) group detection.

Based on different concepts and algorithms (Guimerà and Amaral 2005, Fortunato 2010) derived from graph theory (Wilson 1996, White and Harary 2001), there are various ways to identify subgroups in a network. One approach is to calculate the network modularity (Newman 2004). This concept has been adopted in ecological research to reveal complicated interactions and to detect heterogeneous structure within ecological communities or ecosystems (Olesen et al. 2007, Genini et al. 2012). For example, network analysis has been used to analyse the European fauna structure (Araújo et al. 2011) and successfully reveal the robustness of each fauna group by geographical distribution and resistance to climate change. Modularity measures not only linkages within a module but also the linkages among modules. Nodes possess a higher “degree” (the sum of edges link to the node) than other members in the same module, but a lower degree with members in other modules (Blondel et al. 2008, Newman 2010).

The aim of this chapter is to investigate the possibility of substituting clustering methods with network modularity in indicator species analysis, in the hope of improving predictability of species composition or habitat type in unsampled sites. In this research, the sites were sorted into groups according to their linkage levels (i.e. the number of sharing species). Site groups based on the modularity algorithm would be a substitute for the clustering groups in the indicator species analysis. This new method was compared with conventional clustering-based indicator species to investigate its merits in directly measuring species co-occurrence.

3.3 Methods

3.3.1 Species data

The species dataset is the same as Chapter 2, which is termed the BTO dataset hereafter. In this study, the analytic unit is the 10 km square (a site), with species represented by presence (designated as 1) or absence (designated as 0). No species abundance data were used in the current study.

3.3.2 Network analysis data organisation

The incidence matrix was arranged by site vs. species, with marginal sums as the total number of species in a site and species prevalence as occurrence of specific species across all sites. Based on the matrix, a two-mode network (two types of nodes, with each representing site and species), describing which species resided in what site (linkage), was then plotted. To better understand how sites clustered according to similarity in species composition, the network was projected into a one-mode network containing only nodes that represented sites. Three possibilities of linkages can occur in a pair of sites: no linkage meaning no shared species, one linkage meaning only one species in common, and multi-linkages meaning sharing more than one species. A weighted network illustrating relationship among sites and based on the extent of species in common can then be constructed (Newman 2010).

3.3.3 Site group detection with network modularity

The modularity detection method termed randomtrap (Pons and Latapy 2005) is based on random walk via the closest distance among neighbouring nodes (Fortunato 2010). In network analysis, the closest neighbour node is the node with only one-step linkage to the target node.

The nodes are then organized into linked subgroup (that is, a module) that contain closer neighbours. The modularity detected by the randomtrap method may be visualized as a person walking randomly among nodes along the shortest distance; with limited steps, the walker would readily be trapped in a closely linked subgroup (module). When this process is repeated, the nodes are gradually assigned into different modules and this can be recorded by plotting a dendrogram of random walks. This random-walk method aggregate nodes in an agglomerative direction and measures the similarity between nodes to detect modules. It is an efficient and timesaving method, ideal for analysing large and complicated networks (Pons and Latapy 2005, Fortunato 2010).

3.3.4 Site group detection with k-means clustering versus network modularity

The efficiency of network modularity versus clustering in site group detection was compared within the same avian assemblage dataset. Dufrêne and Legendre's (1997) method was followed for determination of the optimal number of clusters for the indicator species value analysis. This method firstly calculates the indicator value (IndVal) of each avian species for each clustering level (e.g. IndVal for species 1 to species 273 at clustering level 2, 3, 4 etc.). The difference in IndVal between consecutive cluster levels for each species is then calculated (e.g. the difference in IndVal between cluster 2 and cluster 3 for species a) and plotted. Because larger IndVal represents a better predictor, the optimal clustering level is when the difference in IndVal (cluster n+1 minus cluster n) is the largest and is most positive (Appendix C). For comparison purposes, the *k* value (the number of groups to be split) was set to equal to the number of modules in modularity detection. In this case (Appendix C, Figure A-1), because the value (difference in IndVal, y axis) is the highest between cluster 2 and cluster 3, and is of similar high value between cluster 4 and cluster 5 (both also have smaller negative value unlike between cluster 7 and cluster 8), it is most optimal to divide the avian assembly into three or five clusters. However, either three or five clusters has lower predictability than the method based on network modularity, as demonstrated briefly in Figure A-2 (below, for three clusters) and comprehensively in Chapter 3 (for five clusters).

In this stage, the site map partitioned by network modularity and k-means clustering method was compared, and was termed population modularity and population clustering, respectively. The resulting site groups (based on the whole dataset) were compared with site groups determined by partial sampling of whole dataset (see the following paragraph) to investigate the reliability of these two partitioning methods.

3.3.5 Sampling site and sampling site groups

For deciding the sampling size used in this analysis, a preliminary sensitivity analysis was done to estimate the sampling size for effectively representing population modularity. The analysis showed when using 10% and 20% sampling, the sub-squares belong to a small coastal site group would usually be missed and not be included in the modularity result, but in 40% sampling these sub-squares would be selected more often and had more representativeness of the population modularity (for details see Appendix D). Also, in the previous chapter, a 40% sampling effort can accurately predict species richness, therefore the 40% sampling effort was applied here. To assess the predictability of indicator species, 40% random samples from the whole BTO dataset were partitioned based on network modularity and k-means clustering and were called sampling modularity and sampling clustering, respectively. Assigned modularity based on 40% sampling was compared with assigned modularity based on the whole population using the *Fowlkes-Mallows index* (*FM index*) (Fowlkes and Mallows 1983) to see whether the sampling sites were classified in the same module as when the assignment was based on the whole population. The same process was applied to compare sampling clustering and population clustering. The *FM index* compares two clusters and evaluates the extent of similarity or dissimilarity between the clusters. Assuming that two clusters X_1 and X_2 both with *n* objects and *k* clusters. X_1 ($k=1\ldots i$) and X_2 ($k=1\ldots j$) are compared and a matching matrix (m_{ij}) can be created as (when *k*=2):

$m_{ij}=$

		X_2	
		1	2
X_1	1	m_{11}	m_{1j}
	2	m_{i1}	m_{ij}

The *FM index* is defined as:

$$\text{FM\_Index} = \sqrt{\frac{T_k}{P_k \cdot Q_k}}$$

Where $T_k = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{ij}{}^2 - n$, $P_k = \sum_{i=1}^{k} m_{i\cdot}{}^2 - n$, $Q_k = \sum_{j=1}^{k} m_{\cdot j}{}^2 - n$. The *FM index* ranges from zero to one, with higher value indicating more similarity between two clusters. Significance of the *FM Index* is evaluated against $\mathrm{E}(FM) \pm 2\left(var(FM)\right)^{1/2}$, where *E(FM)* is the mean and *var(FM)* is the variance under the assumption of no relation between X_1 and X_2.

3.3.6 Selecting the indicator species

Based on 40% sampling effort, the indicator species value for each species in each site (termed IndVal hereafter) is determined by two indices, A and B, which represent specificity (the *positive predictive value*) and frequency (the *sensitivity*) of species to the site, respectively (sensu Dufrêne and Legendre 1997; De Cáceres *et al.* 2012). A_{ij} is the proportion of species i present in specific site group j ($\mathrm{Npresence}_{ij}$) relative to presence of species i in all site groups ($\mathrm{Npresence}_{i}$), and the higher A_{ij}, the more species i is specific to site group j. Similarly, B_{ij} is the frequency of sites in site group j where species i is present (Nsites_{ij}) among all sites in site group j (Nsites_{j}).

$$A_{ij} = \mathrm{Npresence}_{ij} / \mathrm{Npresence}_{i}$$

$$B_{ij} = \mathrm{Nsites}_{ij} / \mathrm{Nsites}_{j}$$

$$\mathrm{IndVal}_{ij} = A_{ij} \times B_{ij} \times 100$$

$$\mathrm{IndVal}_{i} = max\left[\mathrm{IndVal}_{ij}\right]$$

Significance of IndVal_i was evaluated by bootstrapping: 1000 randomly resampled observed sites were used to generate an approximate distribution for the observed data and confidence interval for A, B and IndVal.

For selecting the indicator species, the square root of indicator value ($sqrtIV^g$) was used to measure the association between species and sites (*sensu* De Cáceres and Legendre 2009). The *g* denotes group equalisation, to correct for differences in group size among indicator species when summing A_{ij} values in J areas.

$$A_{pa}^{g} = \frac{\mathrm{Npresence}_{ij} / \mathrm{Npresence}_{i}}{\sum_{j=1}^{J} \mathrm{Npresence}_{ij} / \mathrm{Npresence}_{i}}$$

A threshold of A=0.6 (the positive predictive value) and B=0.2 (the sensitivity) was set to select effective indicators (DeCáceres et al. 2012). Because in comparison with single species, a combination of species could be more efficient in being indicator species (DeCáceres et al. 2012), a combination of four-species was selected as candidate species. It is necessary to set a limit for the number of indicator species, otherwise sensitivity of indicator will decrease and uncertainty of positive predictive value will increase with too many indicator species. Indicator species based on two grouping methods, network modularity and k-means clustering, were then used to predict unsampled sites. The A value of indicators also represented the likelihood of an unsampled site to be predicted as a site group including the selected indicators. For example, if an unsampled site contained indicator species *i* indicating site group *I* with an A value of 0.7, then this unsampled site had 0.7 likelihood of belonging to site group *I* by the presence of species *i*. The aim is to search for the maximum value that indicates the most likely species-site association. Significance of the species-site association can be tested by the percentile bootstrap method, which resampled the observed data to generate an approximate distribution of the indicator values (DeCáceres and Legendre 2009, DeCáceres et al. 2012).

All analyses were conducted in the R environment (R Development Core Team 2013) using the R packages “igraph” (Csardi and Nepusz 2006) (for the network analysis), “indicspecies” (DeCáceres and Legendre 2009) (for the indicator values), and “dendextend” (Fowlkes and Mallows 1983) (for the Fowlkes-Mallows index).

3.4 Results

The network of the BTO dataset contained 273 bird species in 2830 nodes (sites). The population network included 3,957,073 edges, with each edge representing a linkage among sites with the same species.

3.4.1 Subgroups detected by population modularity and k-means clustering

3.4.1.1 Modularity

The Randomtrap method identified five modules in the BTO dataset (population modularity) (Figure 3. 1 left panel), with each module containing 133, 475, 11, 1345 and 866 sites, respectively. Sites in the same module had similar species composition (species list in Table 3. 1 left). There was a clear latitudinal segregation among these five modules (Figure 3. 1 left). One small module was located mainly at the northern coast and a few islets (red module in Figure 3. 1 left). Three modules (blue, orange, and purple modules in Figure 3. 1 left) were distributed primarily in highland Scotland, middle to west coast area, and south-eastern area, respectively. There was also a small module (with 11 sites) that was scattered at some coastal areas (green module). The avian species contained in each module are listed in Appendix E, left.

3.4.1.2 Clustering

Because the difference in IndVal is the highest between cluster 2 and cluster 3, and is of similar high value between cluster 4 and cluster 5 (both also have smaller negative value unlike between cluster 7 and cluster 8, Figure A-1 in Appendix C), it is optimal to divide the avian assembly into three or five clusters. However, since the network modularity identified five modules, for comparison purposes, the number of groups to be split was set to five (results for three clusters are briefly demonstrated in Figure A-2 of Appendix C, and also had lower predictability than the method based on network modularity).

The five clusters determined by k-means clustering method each contained 513, 436, 775, 748 and 358 sites (Figure 3. 1 right). Likewise, sites in the same cluster had species with shorter distance (less dissimilar) among themselves than with those outside the cluster. The five clusters were also located approximately from north to south, but the geographic demarcation was less clear-cut than those modules determined by randomtrap method. The northern coast and islands were assigned to the same cluster (blue cluster in Figure 3. 1 right), but the Highland area was partitioned into two clusters (green and orange clusters). The green cluster represented most coastal areas except the northern coast, however the range was larger than the green coastal module in the modularity result. The southern urban area was also divided into two clusters (purple and red clusters) but cannot be easily segregated. Avian species contained in each cluster are listed in Appendix E, right.

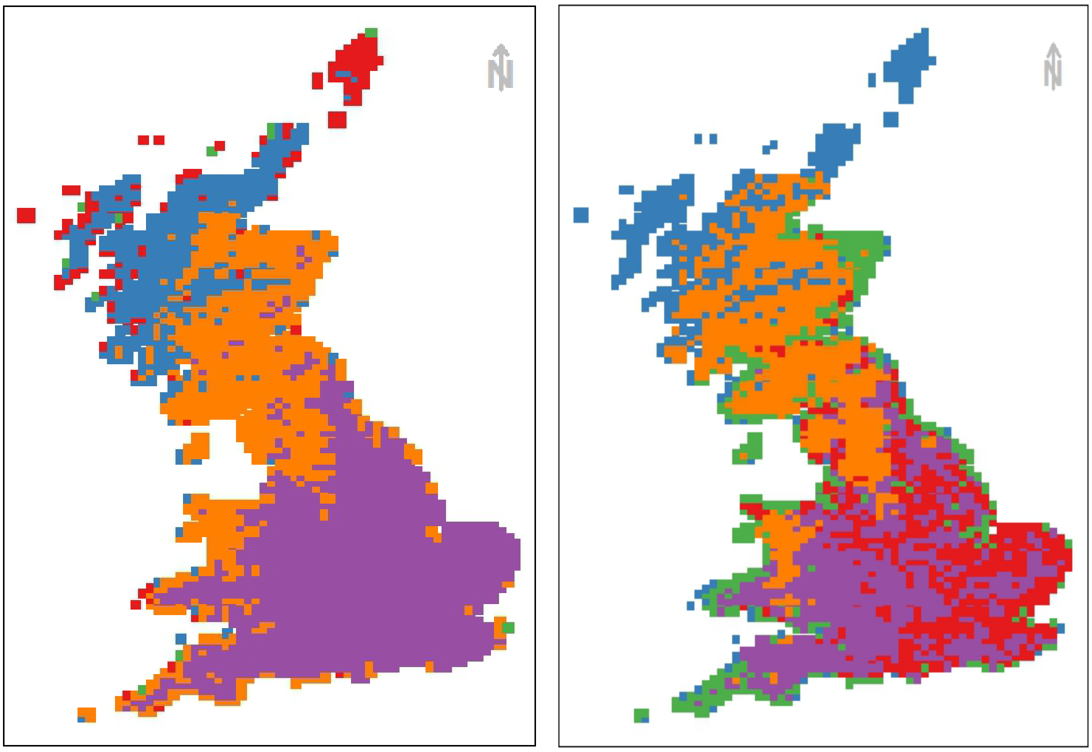


Figure 3. 1 Results of modularity detection (left) and k-means clustering (right) from the entire BTO dataset. Different colours represented different modules or clusters. Colours and order of module (or cluster) are arbitrary. Each square is a 10 km sub-square.

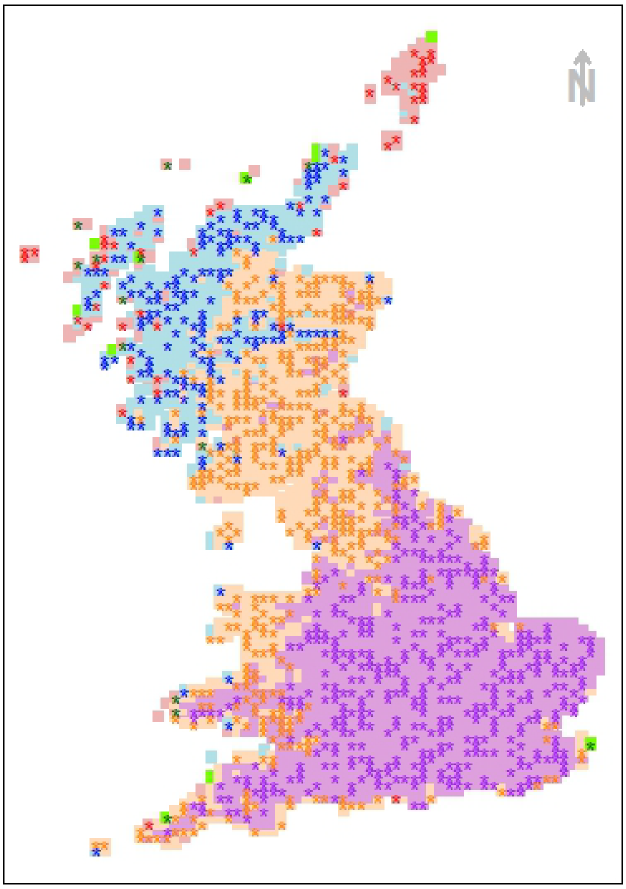


Figure 3. 2 Assigned module of sites based on 40% sampling (colour asterisk) and whole sampling (background colour) of BTO dataset with randomtrap modularity algorithm. Different colours represented different modules. Each square is a 10 km sub-square.

3.4.2 Subgroups detected by sampling modularity and k-means clustering

The Randomtrap method also identified five modules in random sampling of 40% BTO dataset (Figure 3. 2). Sampled sites were largely assigned to the same modules (based on 40% sampling) as those based on the whole dataset (i.e. 100% sampling) (Figure 3. 2), with a FM index of 0.83 and an expected value of 0.34 and variance less than 0.01. In comparison, there was less overlap in the assigned clusters between those sites based on 40% vs. 100% sampling using k-means clustering method (Figure 3. 3), with a lower FM index of 0.62 and an expected value of 0.28 and variance less than 0.01.

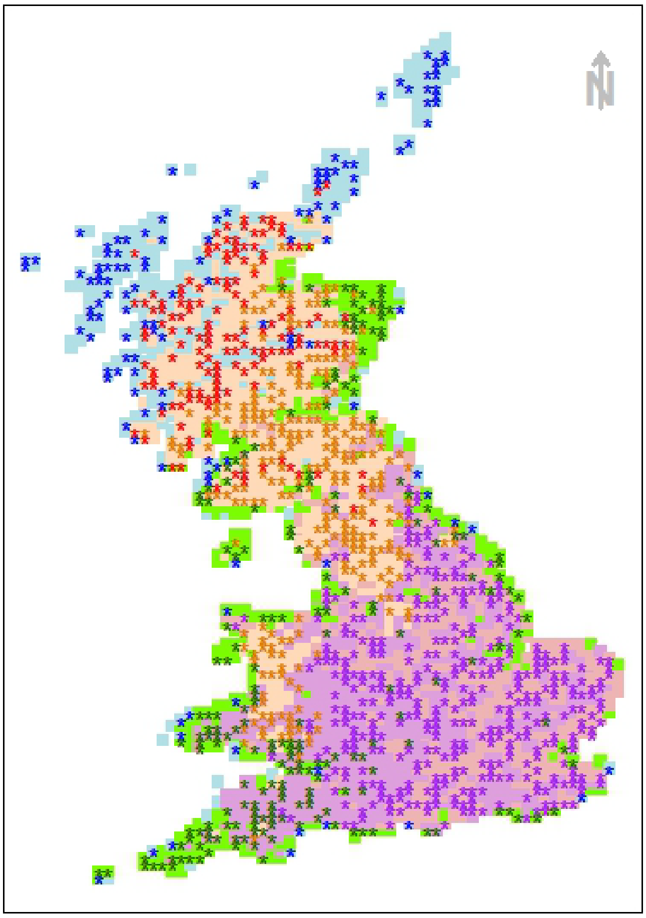


Figure 3. 3 Assigned cluster of sites based on 40% sampling (colour asterisk) and whole sampling (background colour) of BTO dataset with k-means clustering method. Different colours represented different modules. Each square is a 10 km sub-square.

3.4.3 Indicator species selected by sampling modularity and k-means clustering

Each of the five sampling modules and sampling clusters contained groups of indicator species (up to four avian species in each species group, Table 3. 1) that all had significant indicator values (p <0.05, except the green group), but modularity and clustering gave different indicator species combinations (Table 3. 1). For instance, the blue module and the blue cluster were located in approximately the same area but have different indicator species groups with different positive predictive values (Table 3. 1). When all indicator species of a specific site group are observed in an unsampled site, the chance that this site belongs to the site group is higher. For example, in the red sampling modularity, the indicator group comprising fulmar (*Fulmarus glacialis*), meadow pipit (*Anthus pratensis*), shag (*Phalacrocorax aristotelis*) and black guillemot (*Cepphus grille*) has the highest positive predictive value of 0.69. This is followed by fulmar, great black-backed gull (*Larus marinus*), meadow pipit and black Guillemot with a positive predictive value of 0.68, and herring gull (*Larus argentatus*), fulmar, meadow pipit, and shag with a value of 0.64. When an unsampled site contains one of the three indicator groups, the site probably belongs to the northern coast area or coastal islands and based on the indicator values, one can evaluate how confident the classification is. Nonetheless, a very small module (green) does not meet the threshold of selecting valid indicator group. The fulmar is the most suitable indicator species when no threshold is set, with a positive predictive value of 0.51 (Table 3. 1 left).

The positive predictive values (A values) of sampling modularity range from 0.64 to 0.88 (except for the green group which is without an indicator group) and the sensitivity values (B values) range from 0.70 to 0.86 (Table 3. 1 left). Aside from the green group, the rest of indicator groups have high $sqrtIV^g$ values (>0.7) so that site types can be confidently identified. In comparison, the positive predictive values (A values) of sampling clustering range from 0.66

to 0.87, with the mean higher than those of sampling modularity, but the sensitivity (*B* values) varies significantly (from 0.31 to 0.87) (Table 3. 1 right). The value $sqrtIV^g$ also varied remarkably (from 0.47 to 0.87) (Table 3. 1 right).

3.4.4 Using indicator species to predict site groups in unsampled sites

3.4.4.1 Modularity

Indicator species determined by 40% sampling were used to predict assigned modules in unsampled sites and predictability was evaluated by comparing with the assigned module based on population modularity (that is, the whole dataset).

Figure 3. 4 shows how well prediction based on 40% sampling matched those assigned with population modularity. Except for the green and, to a lesser extent, red modules that both contain fewer sites than the other three modules, prediction based on 40% sampling generally matches assigned population modularity. For example, when predicted sites are known to belong to the blue population modularity (Figure 3. 4b), the likelihood of unsampled sites also to be assigned blue modularity is 0.8 (Figure 3. 4 b2), while the likelihood of unsampled sites being assigned to the other four modules is 0 (Figure 3. 4 b1, 3, 4, 5). In terms of the red population modularity (Figure 3. 4 a), the red group indicators successfully predicted the red group to be red (Figure 3. 4 a1) but falsely predicted the blue group to be red (Figure 3. 4 a2). For the small green population modularity, the predictability was low (Figure 3. 4 d) because of no suitable indicator species (Table 3. 2 left).

3.4.4.2 Clustering

Similarly, Figure 3. 5 illustrates frequency of occurrence of different clusters when predicted based on 40% sampling under each of the five population clustering. However, prediction based on 40% sampling generally did not match assigned population clustering except for the blue and green groups with mixed successes. While some predicted sites were assigned the same clustering as the population clustering, there were also a large number of sites that should be assigned blue or green clustering but failed to do so (Figure 3.5 b2, c3). Also, the orange indicator group could predict unsampled sites correctly with lower levels of false negatives (Figure 3. 5 e5), but still with chances of assigning unsampled sites to the wrong colour groups.

Although the sampling clustering indicator group had almost the same level of square root value (indicator species value) as the modularity approach, the power to predict unsampled sites by sampling clustering indicator groups was weaker compared to the sampling modularity indicator groups. Only the orange group was able to successfully predict the site groups without confusion with other groups (Figure 3. 5e). This may due to the inconsistency of partitioning results between the population clustering and sampling clustering (Figure 3. 3) and low level of dissimilarity between species combination in clusters of sampling clustering result (Table 3. 1).

Table 3. 1 The indicator groups selected by sampling modularity (left) and sampling clustering (right) of British Trust of Ornithology (BTO) dataset. *SP*= species name codes; *A* = positive predictive value to a site group of each indicator group; *B* = sensitivity of each species group to a site group; $sqrtIV^g$ = grouped equalised square root value of *A* X *B*.

*: The indicator values of the green module do not reach the threshold of A=0.6 and B=0.2. Under the circumstances no indicator species can be selected. The result here is the result of when no threshold sets.

Modules	Indicator species group			
	SP	*A*	*B*	$sqrtIV^g$
Red	*Fulmarus glacialis* *Anthus pratensis* *Phalacrocorax aristotelis* *Cepphus grylle*	0.69	0.82	0.75
	Fulmarus glacialis *Larus marinus* *Anthus pratensis* *Cepphus grylle*	0.68	0.82	0.75
	Larus argentatus *Anthus pratensis* *Phalacrocorax aristotelis* *Cepphus grille*	0.64	0.86	0.74
Blue	*Corvus cornix* *Motacilla alba* *Oenanthe oenanthe* *Phylloscopus trochilus*	0.84	0.73	0.78

Clusters	Indicator species group			
	SP	*A*	*B*	$sqrtIV^g$
Red	*Fringilla coelebs* *Actitis hypoleucos* *Corvus cornix* *Erithacus rubecula*	0.77	0.76	0.77
	Fringilla coelebs *Aquila chrysaetos* *Oenanthe oenanthe* *Phylloscopus trochilus*	0.84	0.69	0.77
	Fringilla coelebs *Aquila chrysaetos* *Anthus pratensis* *Phylloscopus trochilus*	0.83	0.70	0.76
Blue	*Fulmarus glacialis* *Larus marinus* *Phalacrocorax aristotelis*	0.61	0.68	0.64

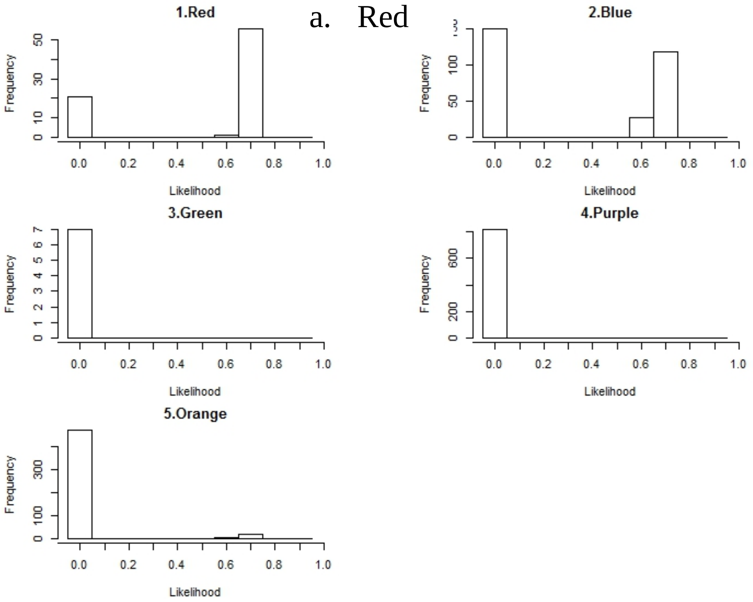
	Actitis hypoleucos *Corvus cornix* *Oenanthe oenanthe* *Phylloscopus trochilus*	0.88	0.70	0.78
	Corvus cornix *Anthus pratensis* *Motacilla alba* *Phylloscopus trochilus*	0.83	0.74	0.78
Green*	*Fulmarus glacialis*	0.34	0.79	0.51
	Phalacrocorax aristotelis	0.26	0.57	0.39
	Fulmarus glacialis *Larus marinus* *Phalacrocorax aristotelis*	0.27	0.50	0.37
Purple	*Garrulus glandarius Sylvia curruca* *Gallinula chloropus* *Phasianus colchicus*	0.85	0.84	0.8
	Turdus merula *Garrulus glandarius* *Sylvia curruca* *Gallinula chloropus*	0.84	0.85	0.84
	Sylvia atricapilla *Garrulus glandarius* *Sylvia curruca* *Gallinula chloropus*	0.84	0.85	0.84

	Fulmarus glacialis *Cepphus grylle*	0.67	0.61	0.64
	Fulmarus glacialis *Larus marinus* *Larus argentatus* *Phalacrocorax aristotelis*	0.60	0.67	0.64
Green	*Corvus corone* *Fulmarus glacialis* *Carduelis carduelis* *Passer domesticus*	0.67	0.37	0.48
	Corvus corone *Fulmarus glacialis* *Carduelis carduelis* *Erithacus rubecula*	0.67	0.37	0.48
	Corvus corone *Fulmarus glacialis* *Carduelis chloris* *Pica pica*	0.73	0.31	0.48
Purple	*Corvus corone* *Phylloscopus collybita* *Fulica atra* *Streptopelia turtur*	0.95	0.85	0.90
	Phylloscopus collybita *Fulica atra* *Carduelis carduelis* *Streptopelia turtur*	0.95	0.85	0.90
	Corvus corone *Fulica atra* *Corvus monedula* *Streptopelia turtur*	0.95	0.84	0.90

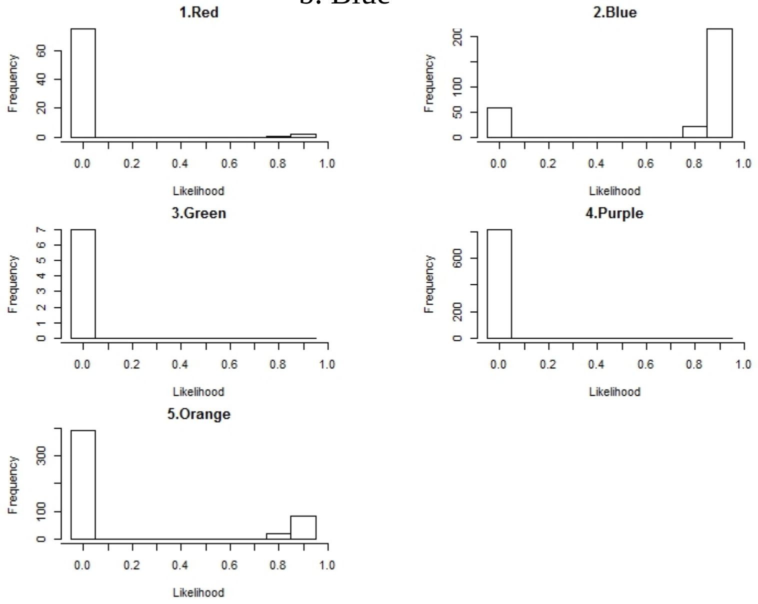
Orange	*Corvus corone* *Numenius arquata* *Apus apus* *Oenanthe oenanthe*	0.74	0.70	0.72
	Numenius arquata *Corvus monedula* *Apus apus* *Oenanthe oenanthe*	0.73	0.70	0.71
	Corvus monedula *Alauda arvensis* *Apus apus* *Oenanthe oenanthe*	0.69	0.74	0.72

Orange	*Actitis hypoleucos* *Cinclus cinclus* *Carduelis carduelis* *Carduelis cannabina*	0.78	0.83	0.80
	Actitis hypoleucos *Cinclus cinclus* *Carduelis carduelis* *Muscicapa striata*	0.75	0.86	0.80
	Actitis hypoleucos *Cinclus cinclus* *Corvus monedula* *Carduelis cannabina*	0.77	0.83	0.80

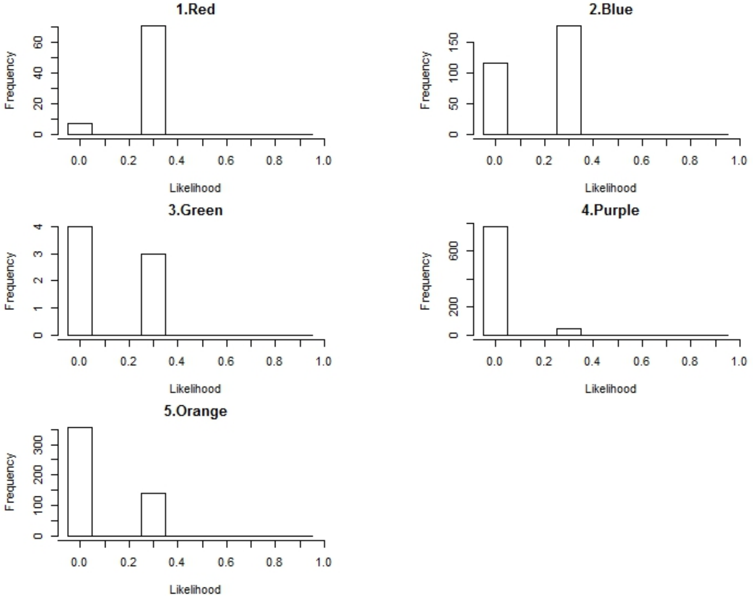
a. Red



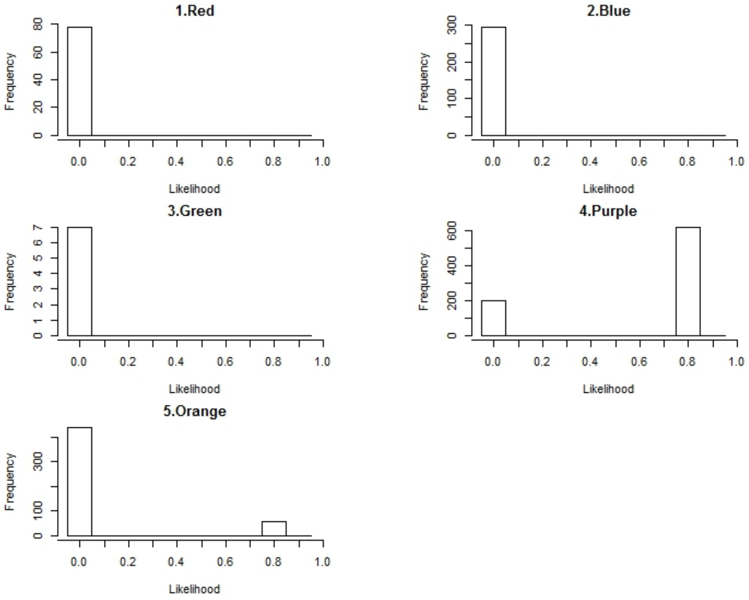
b. Blue



c. Green



d. Purple



e. Orange

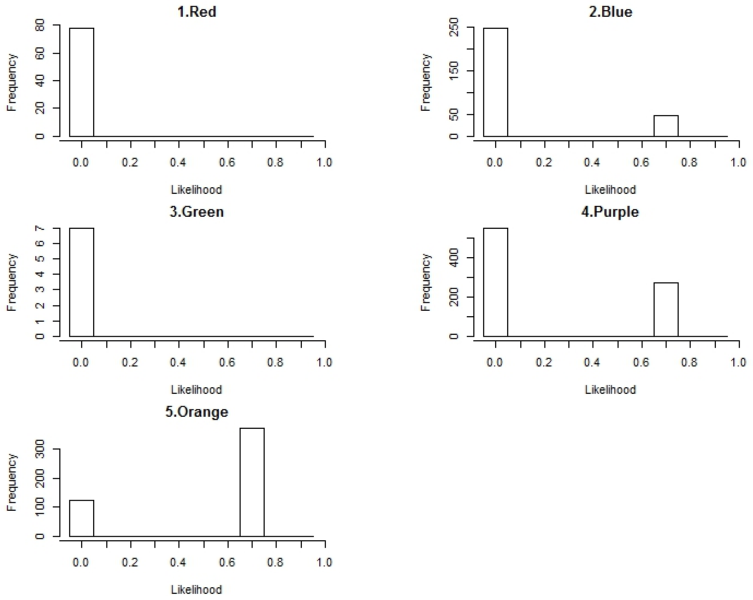
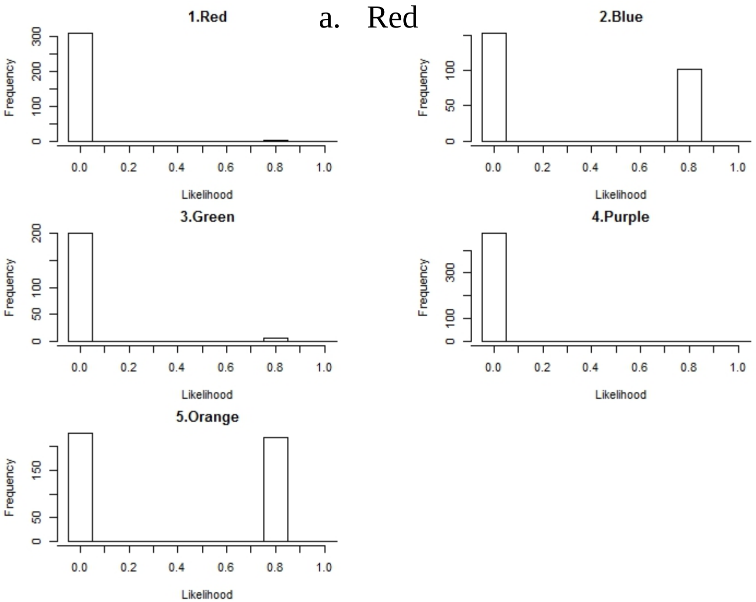
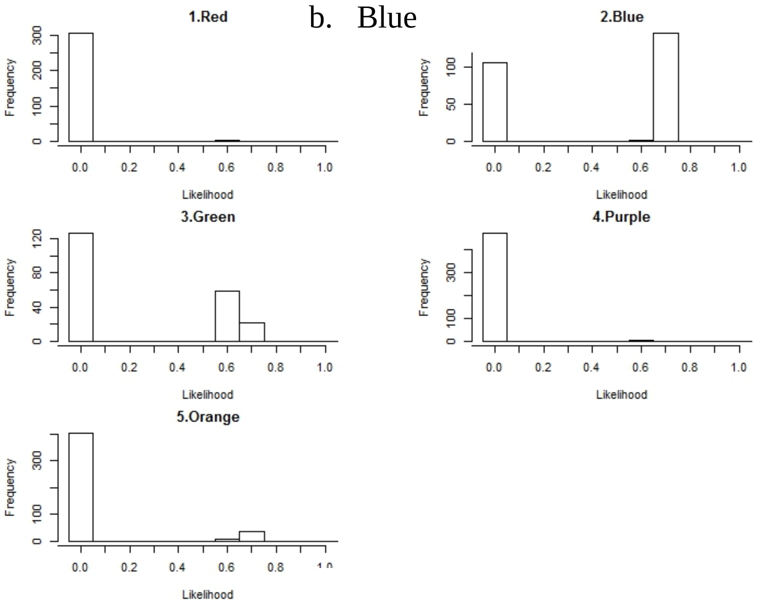
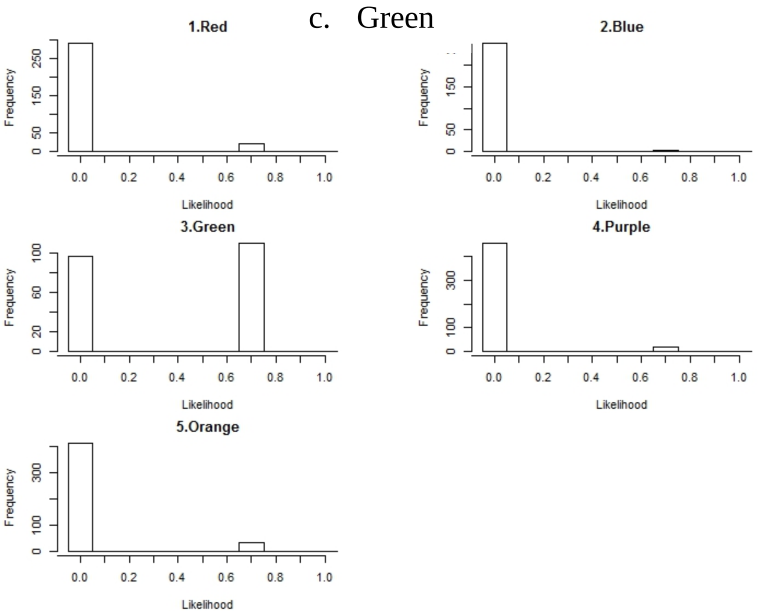
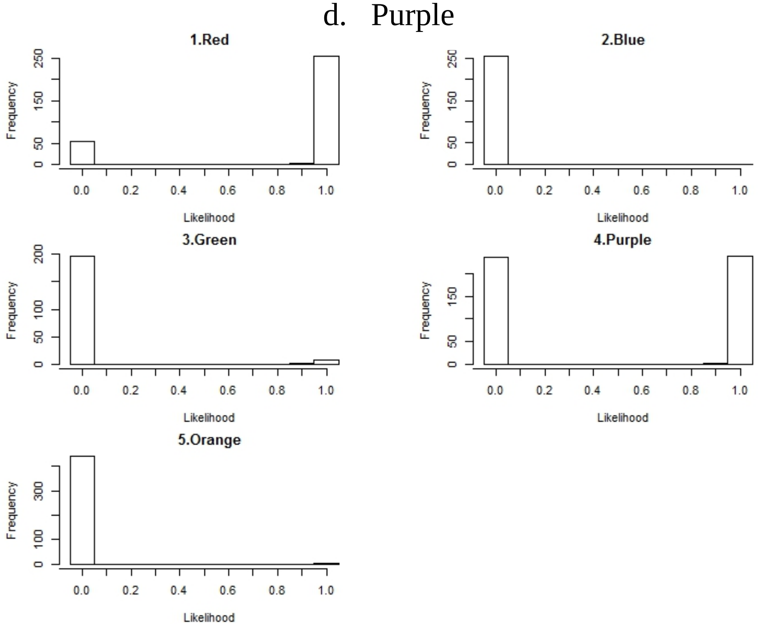


Figure 3. 4 The predictive ability of the site group to the unsampled sites when predicted based on 40% sampling and under each of the five sampling modularity indicator species group. The title (a-e) indicates each of the five assigned population modularity and the x-axis is the likelihood of assigning an unsampled site to the target site group, comparing with the original site memberships (i.e. the site group assigned by the population modularity). For example, figure (a) is the indicators derived from the red group and figure (a.1) is the likelihood to assign an unsampled site, which is originally assigned to the red group in population modularity, to the red site group by the red indicators.


a. Red


b. Blue


c. Green


d. Purple

e. Orange

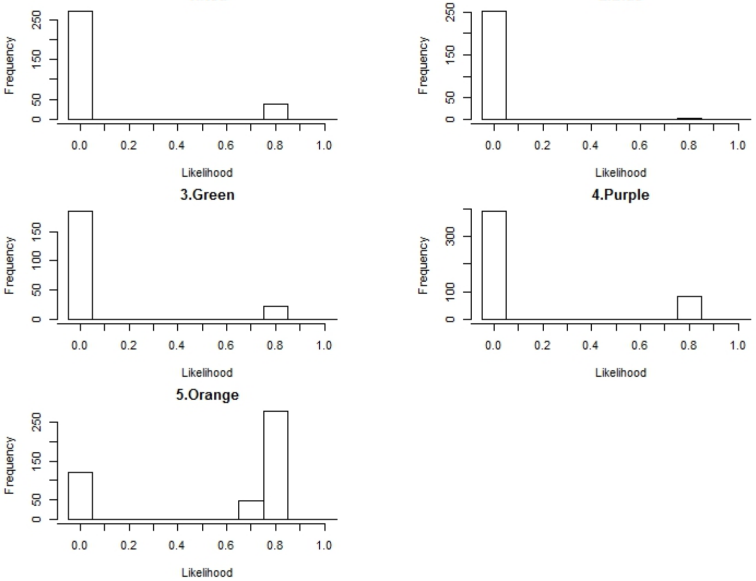


Figure 3. 5 The predictive ability of the site group to the unsampled sites when predicted based on 40% sampling and under each of five sampling clustering indicator species group. The title (a-e) indicates each of the five assigned population clustering and the x-axis is the likelihood of assigning an unsampled site to the target site group, comparing with the original site memberships (i.e. the site group assigned by the population clustering). For example, figure (a) is the indicators derived from the red group and figure (a.1) is the likelihood to assign an unsampled site, which is originally assigned to the red group in population clustering, to the red site group by the red indicators.

3.5 Discussion

3.5.1 Consistency in site-species association in the indicator species analysis

Compared with the k-means clustering method, indicator species identified by network modularity detection can more successfully assign unsampled sites to correct species groups. Indicator species analysis is an efficient tool for recognising representative species and has been applied on various research topics such as plant species composition (Flinn et al. 2008, Willerslev et al. 2014), vector ecology (Obsomer et al. 2013), mycological community (Taylor et al. 2014, Varela-Cervero et al. 2015) and intestinal microflora ecology (Seedorf et al. 2014, Planer et al. 2016). The accuracy of prediction of site types relies on robust classification of sampled species and sites (Dufrêne and Legendre 1997). The results of clustering, however, can be easily influenced by several factors. Firstly, because cluster analysis is based on dissimilarity distance among members, the clustering result depends on which sites are sampled. Different sampled sites lead to different clustering outcomes, turning into inconsistent species-site associations. The second concern is that subgroups partitioned by clustering methods are arbitrarily delineated by user-defined cut off thresholds and may create artefact groups (Legendre and Legendre 2012). Although k-means clustering is recommended to be utilised in indicator species analysis (Dufrêne and Legendre 1997), this method can be easily affected by to which group the original node is assigned (Celebi et al. 2013). Even though Dufrêne and Legendre (1997) suggest repeating the partitioning process until the indicator values of all species are decreasing to find the appropriate number of clusters, there is still no reliable criterion to decide the number of clusters.

In comparison, network analysis measures species co-occurrence directly and partitions the subgroups more objectively and is therefore an efficient substitute for group partitioning in the indicator species analysis. Moreover, a notable difference between network analysis and clustering is that network analysis measures network distance while clustering measures Euclidean distance. Network distance is the number of direct paths between two nodes (the distance from one node to its nearest neighbour is one path away and nodes without direct connection will have more than one path). The Euclidean distance is not identical to the network distance and might not be suitable for measuring the relationship in a network containing geographical information (which is a spatial network) (Yiu and Mamoulis 2004) given that it neglects the relationships among nodes but focuses only on how different these nodes are. It is thus unsurprising that clustering and network analysis lead to different partitioning results and compared with clustering, network analysis is more consistent in assigning species groups.

3.5.2 Geographical distribution of site groups

The spatial network has been frequently applied in fields such as transportation and mobility networks, internet, mobile phone networks, power grids, social and contact networks (Barthélemy 2011). For example, a spatial network of chain restaurants and their proximities to the nearest road systems can help identify areas that are suitable to start a new branch (Yiu and Mamoulis 2004). In social science, user distribution based on users' social connection such as check-in locations (Joseph et al. 2012) or online photo tag co-occurrence (Zhang et al. 2012) can be revealed with geo-social networks. Similarly, network analyses have been adopted by biologists and many biological studies have applied spatial networks to analyse co-occurrence pattern of microbes in human bodies (Levy and Borenstein 2013). In ecological research, incorporating ecological processes such as dispersal, competition, or food webs into network analysis to uncover geographic pattern remains challenging (e.g. Cumming et al. 2010) and most of the research focuses on single species or individual (e.g. Saura and Rubio 2010, Pereira et al. 2011).

The current study shows that network analysis can efficiently classify subgroups of co-occurring avian species in Britain according to species co-occurrence. For example, the complete bird assemblage is sorted into subgroups according to occurrence in coastal or inland areas, along a north-south gradient. The sites are divided into several groups and sites within a module are usually geographically connected compared to the geographical mosaic pattern of clustering results. This suggests the ability to combine species (co-occurrence) and spatial information by network analysis. In the indicator species analysis, it also provides more ecological meaning and consistency to species-site associations. However, this ability has not attracted enough attention and more research is required. Newly developed methods such as constrained clustering (Yuan et al. 2015, Cheruvelil et al. 2017) have attempted to balance spatial continuity with landscape homogeneity. In this method, it is not only similarities in characteristics among sites that are considered; adjacency of sites is also stressed, which helps delineate site groups. Incorporating this network modularity algorithm into geographical site arrangement might be a novel direction for analysing geographical networks.

3.5.3 Indicator species value as surrogate in plant and animal research

This research tests a simple idea - utilising assemblage structure to predict species composition. In most studies, predictions of species richness or community structure usually involve environment attributes. However, environment attributes could be affected by so many parameters that identifying the most important and disentangling their relative influence could become difficult (Sutherland 2006, Arita et al. 2008). Capitalising simply on assemblage structure has the advantage of avoiding confounding effects arising from various environment factors by focusing solely on the relationships among species. For indicator species analysis, where only species information is considered, the question is how these indicator species can be effective surrogate for whole flora or fauna. The indicator species value concept originates from research on plant ecology (Dufrêne and Legendre 1997), and is extensively adopted to characterise plant communities (Ricotta et al. 2015), but indicator species value is not limited to studies on plant communities (Tichý and Chytrý 2006). However, whether a similar concept can be applied to mobile animals remains little studied. Sattler et al. (2014) applied indicator species analysis (with traditional clustering methods) to six taxa, including plants, insects, and birds and compared the suitability for reflecting biodiversity measures (such as biodiversity index, species richness, etc.). Overall, indicator suitability of plants was better than for birds. Our study demonstrates that when incorporating network analysis, the indicator species value concept can also be applied to mobile organisms.

The indicator species value aims to construct site typology by comparing species difference instead of predicting existence of specific species. Research that commonly applies this value, such as vegetation classification, thus stresses more the dominant plant composition and community type (Chytrý et al. 2002, Tichý and Chytrý 2006). However, it will be challenging to extend this method to predict the appearance of undetected species in an unsampled site, for this value still lacks information on associations among detected species and undetected species. More information, such as environmental attributes or non-random numerical relationship like nestedness, is required to achieve the goal of predicting individual species.

3.5.4 Identifying indicator species and umbrella species

In this research, network analysis was applied to identify “key-members” among large amounts of data and uncover complicated relationships. Conceptually, this is related to other studies that have attempted to identify umbrella or focal species objectively. For example, Fleishman et al. (2001) used percentage of co-occurrence as an umbrella species index. The number of species co-occurring with the umbrella species was treated as the dependent variable to investigate the efficiency of an umbrella species index. Umbrella species identified by this index, however, did not cover more species than when umbrella species were randomly selected. Simply using species richness reveals very limited information on species composition which is essential for identifying umbrella species. Randomly selected species may include more species but fail to identify whether they are common species with little conservation value. In the present study, the concept was advanced with network analysis by not only estimating the number of species each member is linked to but also the identity of the linked species.

Chapter 4: Prediction of Taiwan avian assemblage with nestedness and network analyses

4.1 Abstract

Results of the previous studies on British avian species illustrate that incorporating assemblage structure (nestedness) in species accumulation curve and network analysis of indicator species value can improve predictability of avian species richness and species groups, respectively. However, the general applicability of these methodologies remains unexplored. Here, I apply the same methods developed for Britain to the birds of Taiwan based on the Taiwanese Breeding Bird Survey. Taiwan is a small island (ca. 36,000 km^2) but with elevation ranging from sea level to nearly 4,000 meters. Moreover, past studies in Taiwan have focused on environmental associates of avian species richness without considering species status (e.g. common, exotic, or rare species), meaning that the environmental determinants of the Taiwanese avian assemblage are not well understood. For the species accumulation curve result, the Arrhenius power function was found to perform better than negative exponential, logarithmic, or Clench function in the prediction of Taiwan avian species richness. This result is different from the British avian assemblage in which the logarithmic function provides the best prediction for species richness. In addition, a very high prediction rate (error rate = 1.55%) was achieved when species richness estimation based on two curves, random and nestedness ordered, was averaged; this was when only about 1% area of Taiwan has been surveyed (410 1x1 km plots out of Taiwan's 36,000 km^2 area). For the indicator species analysis, selecting indicator species by network modularity had better performance than by conventional k-means clustering method. The resultant species groups can be best differentiated by elevation and

Normalised Difference Vegetation Index. This study demonstrates that capitalising on nestedness structure and network analysis can enhance predictability in Taiwan avian assemblage. Moreover, in contrast to the conventional usage of species richness as a dependent variable, associating environmental attributes with network-based indicator species groups is informative in the recognition of factors critical for delineating community assemblage and is valuable in the preservation of intact community structure and function.

4.2 Introduction

In previous chapters, I have discussed how species accumulation curves and indicator species analysis can improve predictability of species richness and species composition, respectively, by incorporating information on assemblage structure. Species accumulation curves are commonly applied to predict species richness in unsampled areas, but conventional species accumulation curves usually overestimate species richness, especially when the sampling area is unknown or when the sample size is very small (Soberón and Llorente 1993, Ugland et al. 2003). My previous analysis of British Trust for Ornithology (BTO) 1988-1991 survey data demonstrated that species accumulation curves that incorporate nestedness structure can predict species richness more precisely than conventional curves and without overestimation. However, it remains unclear how well the nested species accumulation curve performs in areas with different environmental characteristics from Britain. In this chapter, I investigate the performance of nested species accumulation curves in a very different setting - Taiwan.

Predicted species richness will depend on the extrapolation models, so choosing a suitable formula is important for a precise prediction (Soberón and Llorente 1993). The suitability of specific extrapolation models depends on the characteristics of the sampled area. For example, the exponential function which approaches an asymptote more swiftly might be more suitable for small or homogeneous areas or when there are few rare species (Soberón and Llorente 1993, Tjørve 2003, Díaz-Francés and Soberón 2005). By contrast, Clench or logarithmic models, which grow more rapidly, are better for sampling areas with many unknown taxa or large, heterogeneous areas with many rare species. The logarithmic model is also suitable for areas comprising fauna or fauna that experience severe annual fluctuations (e.g. many tropical butterfly species) (Soberón and Llorente 1993). Lastly, the commonly used power model usually overestimates species richness (Thompson et al. 2003a) and is best suited to intermediate to large sampling areas or islands (Preston 1962a, 1962b, He and Legendre 1996, He et al. 1996, Rosenzweig and Ziv 1999, Tjørve 2003, Scheiner 2003). In previous studies (Lennon et al., 2001 and my previous chapter), the logarithmic model was found to be best in predicting British avian assemblage richness. However, the logarithmic model might not be appropriate where environmental characteristics or coverage of the sampling area is different from Britain.

On the other hand, some extrapolation models, such as the logarithmic or power model, lack an asymptote, so their application is confined to medium-sized areas and is unsuitable for very small or very large areas (Tjørve 2003). Applying functions without an asymptote to predict species richness is risky (Ugland et al. 2003), as estimating species richness depends on an user-defined boundary. Based on my previous BTO 1988-1991 survey study (Chapter 2), when data are input based on nested order, proportionally more uncommon species accumulate in the initial stage of sampling, as the most species rich site has the most uncommon species, followed by a decelerating rate of increase. Thus accumulation leads to a curve slowly approaching an asymptote, making it possible to estimate the complete species pool from a sample of the population. Comparatively fewer uncommon species are included in the initial samples when data are input based on a random order. Compensation by an accelerating rate of increase in the later stage of accumulation can cause an overshoot in the predicted species richness. However, it remains unclear whether incorporating nested order can help stabilise species accumulation curves and improve predictability in regions other than Britain.

Determining key parameters for modelling is typically the most important but difficult issue in species distribution modelling (Johnson and Omland 2004, Guisan and Thuiller 2005) and each model selection method has its own requirement and limitation (see reviews in Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Elith and Leathwick 2009). Indicator species analysis can help identify groups of species that implicitly have similar ecological requirement or have similar response to change in the environment (Dufrêne and Legendre 1997, McGeoch 1998, De Cáceres and Legendre 2009, De Cáceres et al. 2010). These species groups can then be associated with environmental attributes. This is in contrast to the conventional method where researchers select specific environmental characteristics (e.g. primary forests) first, followed by listing species commonly observed in selected habitats, which is usually a subjective process. However, the classification of site group, which is the basis for selecting indicator species, has to be coherent and robust. My previous study on the BTO dataset (Chapter 3) has demonstrated that incorporating the concept of ecological networks can improve the performance of conventional indicator species analysis. However, the applicability of this approach to regions apart from Britain needs to be validated.

Taiwan is an island with different environmental characteristics from Britain. It is located along the south-eastern coast of mainland Asia and was connected with the main continent several times through land bridges during the glacial periods (Voris 2000). Fauna and flora in Taiwan is therefore partially inherited from the mainland Asia (Lee et al. 2004). On the other hand, the oceanic environment near the coast allows waders and seabirds to thrive. This is similar to the avian assemblage source and composition of Britain (Taiwan has 64 of the 76 families in Britain, with 208 species in common). However, in contrast to Britain, Taiwan is a subtropical island (latitude: 22 – 25 degree north) with an altitudinal range of nearly four thousand meters in a small area of 36,000 km^2 (Ding et al. 2005). In this research, the robustness of nested species accumulation curves and network based indicator species analysis was evaluated with the Breeding Bird Survey (BBS) data from Taiwan. Moreover, I looked at environmental determinants of species composition in Taiwan, as previous studies on the Taiwanese avian assemblage have focused on uncovering the environmental determinants of species richness (Lee et al. 2004, Ding et al. 2005, Koh et al. 2006) instead of species composition. Among the environmental attributes considered, including mean annual temperature, elevation, primary productivity (using a Normalised Difference Vegetation Index (NDVI) proxy) and urbanization, avian species richness was found to be associated with primary productivity in southern (Ding et al. 2005) and northern Taiwan (Koh et al. 2006) but elevation and urbanization played a more important role in explaining variation in avian species richness in the whole island of Taiwan (Lee et al. 2004). However, sites with similar species richness might have different species composition. Furthermore, sites with the highest species richness might contain only common species (e.g. in human disturbed areas) instead of any rare or threatened species. Therefore, focusing solely on species number ignores important information on species composition, with the latter usually more critical for conservation.

Neglecting information on species composition also confounds association of species richness with environmental attributes. For example, assume that there are two sites of the same avian species richness with distinct environmental characteristics, such as one under intensive anthropogenic disturbance while the other is pristine. Applying species richness to these two sites will rate them as equal and could lead to the conclusion that human disturbance is unimportant. By contrast, incorporating information on species composition is very likely to uncover the significance of human disturbance to the avian assemblage since composition usually varies with levels of human disturbance. Because the indicator species can identify groups of species that have similar environmental requirements, the other goal of the study was to investigate environmental associates of avian species groups in Taiwan after they were identified with network analysis.

4.3 Methods

4.3.1 Study site

Taiwan is a moderately sized island (35,873 km^2) located in Southeast Asia. Elevation ranges from 0 to 3952 metres (a.s.l.) with the central mountain dividing Taiwan into eastern and western parts. Eastern Taiwan is dominated by mountains while western Taiwan is characterised by flat plains and intensive human disturbance. The Tropic of Cancer runs through southern Taiwan and the island is covered with tropical and subtropical vegetation. Small associated islands (Kin-men, Ma-tou, Pescadores, Little Liuchiu, Green and Orchid islands) were not included in this study, as I only consider mainland Britain in my other chapters.

4.3.2 Research data

The Taiwan BBS is a volunteer-based bird survey across the main island of Taiwan that started in 2009. Surveys are conducted in a 1x1 km grid-square system, using the point count method. Up to 2016, a total of 410 sites (*ca.* 1.1% of total area of Taiwan main island) have been surveyed and 283 species of breeding birds recorded, however there is a total of 626 species in Taiwan according to the most recent checklist (Ding et al. 2014). These species count data were transformed into an incidence matrix that contains 410 rows (sites) x 283 columns (species). The Taiwan island is divided into four regions: north, west, east and mid-high elevation (areas above 1,000 meters) for their distinct climate and geographic characteristics (Ko et al. 2013). In this study, all 410 squares served as the training data to predict avian species richness in main island of Taiwan by using species accumulation models. The basic geographical unit used in this analysis is the 1 x 1 km square.

4.3.3 Calculating assemblage nestedness structure and network analysis

The methods (from *calculating nestedness* to *selecting the indicator species*) in this chapter are identical to the corresponded processes in Chapter 2 and 3. The section *environmental associates of site groups* is only implemented in this chapter.

4.3.4 Calculating nestedness

Nestedness temperature, a measure of community nestedness, takes unexpected presences and unexpected absences into account (Ulrich and Gotelli 2007, Ulrich et al. 2009). It is represented by a normalised sum of squared relative distance of absences above and presences below a hypothetical isocline that separates occupied from unoccupied areas in a perfect nested matrix (Atmar and Patterson 1993). According to Atmar and Patterson (1993), the calculation of nestedness is based on the unexpectedness, which is measured by the distance an unexpected presence or absence of a species lies in the matrix as compared to the perfectly nested matrix. Unexpectedness runs diagonally along a line running parallel to the skew diagonal. For a matrix with n species and m sites, the local unexpectedness (u_{ij}) is calculated as:

$$u_{ij} = \left(\frac{d_{ij}}{D_{ij}}\right)^2$$

D_{ij} is the length of the full line running through the jth species on the ith site and d_{ij} is the specific length along that line. The total unexpectedness (U) is then represented by:

$$U = \frac{1}{mn\sum_i \sum_j u_{ij}}$$

A perfectly ordered matrix should have no unexpectedness, thus *U*=0, and one with maximum unexpectedness always has $U_{max} \approx 0.04145$. The nestedness temperature T is then defined as:

$$T = kU$$

$$k = \frac{100}{U_{max}}$$

T ranges from 0-100°, with T=0° representing a perfect-nested matrix and T= 100° absolute randomness.

Ordering a presence-absence matrix in nested order therefore means re-arranging it into a new matrix where sites are organised from the most species-rich to the most species-poor. Nestedness of the Taiwan avian assemblage was estimated as a first step toward understanding whether using nested order can help improve the predictability of species accumulation curves.

4.3.5 Species accumulation curve and extrapolation

4.3.5.1 Ordering of sampled sub-squares for species accumulation curves

Taiwan avian assemblage data were arranged in two different orders in terms of adding in sampling sub-squares: 1) random order with the mean of 100 permutations (rarefaction curves;

sensu Gotelli & Colwell 2001); and 2) nested structure, that is, ordering sampling sub-squares from the most species-rich to the most species poor.

4.3.5.2 Functions for fitting species accumulation curve

Four functions for fitting species accumulation curves were compared for two kinds of sampling orders (random and nested) to evaluate their accuracy in terms of predicting overall species richness of Taiwan. Three of the four functions are based on Soberón & Llorente (1993) and the other from Ugland et al. (2003); one function, the power function (Equation 4), has been applied by Lennon et al. (2001) to describe the pattern of the British avifauna:

Negative exponential function: $\mathbf{y} = \mathbf{a} \times \left(\mathbf{1} - \boldsymbol{e}^{-\frac{\boldsymbol{x}}{\boldsymbol{b}}}\right)$ (Equation 4)

Logarithmic function: $\mathbf{y} = \mathbf{a} + \mathbf{b} \times \mathbf{log}(\boldsymbol{x})$ (Equation 5)

Clench function: $\mathbf{y} = \frac{\boldsymbol{ax}}{\mathbf{1}+\boldsymbol{bx}}$ (Equation 6)

Arrhenius function: $\mathrm{y} = \mathrm{a} \times x^{z}$ **(Equation 4)**

in which y is the species number and x the number of squares.

4.3.5.3 Calculating the predictive power of species accumulation curves

The four functions mentioned above were used for extrapolating and were compared for their ability to predict overall avian species richness across Taiwan. The datasets were fitted to four curve functions (i.e. negative exponential, logarithmic, Clench, and Arrhenius). The inaccuracy in the prediction of total species richness, that is, error rate, is defined as:

1-(predicted species richness/recorded species richness)

There are 626 avian species in Taiwan according to the most recent checklist of birds of Taiwan (Ding et al. 2014), so the recorded species richness in the calculation of error rate is set as 626. A positive error rate indicates an underestimation while negative value an overestimation. The calculation and statistical procedures were conducted in R (R Development Core Team 2013).

4.3.6 Network analysis data organisation

The incidence matrix was arranged by site vs. species, with marginal sums as the total number of species in a site and species prevalence as occurrence of specific species across all sites. Based on the matrix, a two-mode network (two types of nodes, with each representing site and species), describing which species resided in what site (linkage), was then plotted. To understand how sites clustered according to similarity in species composition, the network was projected into a one-mode network containing only nodes that represented sites. Three possibilities of linkages can occur in a pair of sites: no linkage, meaning no shared species; one linkage, meaning only one species in common; and multi-linkages, meaning more than one species is shared. A weighted network illustrating relationships among sites and based on the extent of species in common was constructed (Newman 2010).

4.3.7 Site group detection with network modularity

The modularity detection method termed randomtrap (Pons and Latapy 2005) is based on a random walk via the closest distance among neighbouring nodes (Fortunato 2010). In network analysis, the closest neighbour node is the node with only one-step linkage to the target node. The nodes then are organised into linked subgroup (that is, a module) that contain close neighbours. The modularity detected by the randomtrap method is similar to a person that randomly walks among nodes along the shortest distance; with limited steps, the walker would be readily trapped in a closely linked subgroup (module). When this process is repeated, the module position can be recorded by plotting the dendrogram of random walks, which is an efficient and timesaving method for identifying groupings and is ideal for analysing large and complicated networks (Pons and Latapy 2005, Fortunato 2010).

4.3.8 Site group detection with k-means clustering versus network modularity

The efficiency of network modularity versus clustering in site group detection was compared using the same avian assemblage dataset. The non-hierarchical clustering k-means method is commonly applied in indicator species analysis (Dufrêne and Legendre 1997). For comparison purposes, the *k* value (the number of groups to be split) was set to be equal to the number of modules in modularity detection.

In this stage, the site map partitioned by network modularity and k-means clustering method was compared, and was termed population modularity and population clustering, respectively. The resulting site groups (based on whole dataset) were compared with site groups determined by partial sampling of the whole dataset (see the following paragraph) to investigate the reliability of these two partitioning methods.

4.3.9 Selecting the indicator species

The indicator species value for each species in each site (termed IndVal hereafter) was determined by two indices, A and B, which represents specificity (the *positive predictive value*) and frequency (the *sensitivity*) of species to the site, respectively (sensu Dufrêne and Legendre 1997; De Cáceres *et al.* 2012). A_{ij} is the proportion of species i present in specific site group j ($Npresence_{ij}$) relative to presence of species i in all site groups ($Npresence_{i}$), and the higher

A_{ij}, the more species i is specific to site group j. Similarly, B_{ij} is the frequency of sites in site group j where species i is present (Nsites_{ij}) among all sites in site group j (Nsites_j).

$$A_{ij} = \text{Npresence}_{ij} / \text{Npresence}_i$$

$$B_{ij} = \text{Nsites}_{ij} / \text{Nsites}_j$$

$$\text{IndVal}_{ij} = A_{ij} \times B_{ij} \times 100$$

$$\text{IndVal}_i = max\left[\text{IndVal}_{ij}\right]$$

Significance of IndVal_i was evaluated by permutation tests that randomly reallocated sites among site groups and was calculated by the difference between the observed value and the mean results of permutations, weighted by standard deviation of the values obtained by permutations. Each IndVal value was run for 999 times, which is more than recommended (>200 times; Dufrêne and Legendre 1997).

For selecting the indicator species, the square root of indicator value ($sqrtIV^g$) was used to measure the association between species and sites (*sensu* De Cáceres and Legendre 2009). The *g* denotes group equalisation, to correct for differences in group size among indicator species when summing A_{ij} values in J areas.

$$A_{pa}^{g} = \frac{\text{Npresence}_{ij} / \text{Npresence}_i}{\sum_{j=1}^{J} \text{Npresence}_{ij} / \text{Npresence}_i}$$

An effective indicator is defined as A≥0.6 (the positive predictive value) and B≥0.2 (the sensitivity) (DeCáceres et al. 2012). Since in comparison with a single species, a group of species could better serve as indicator species (DeCáceres et al. 2012), up to a four-species combination was selected as candidate species. It is necessary to set a limit for the number of indicator species, otherwise sensitivity decreases and uncertainty of positive predictive value increases. The A value of indicators also represents the likelihood of an unsampled site to be predicted as a site group when it contains a specific indicator. For example, if an unsampled site contains indicator species *i* that can indicate presence of site group *I* and is with an A value of 0.7, this unsampled site has 70% likelihood of belonging to site group *I*, with the presence of species *i*. The aim is to search for the maximum value that indicates the most likely species-site association. Significance of the species-site association was tested by a percentile bootstrap method that resampled the observed data to generate an approximate distribution of the indicator values (DeCáceres and Legendre 2009, DeCáceres et al. 2012).

4.3.10 Environmental associates of site groups

Redundancy analysis (RDA) was applied to identify visually key environmental attributes that can discern site groups categorised by network modularity. The RDA method is mainly used for linear responses (comparing to canonical correspondence analysis: CCA, for unimodal ones). Because the chi-square distance used in CCA is calculated by Euclidean distance that considers relative abundances and is weighted by inverse of the species sums; that is, a common species contributes less to the calculated distance than a rare species, rare species may have a disproportionately large influence on the analysis result (Legendre and Gallagher 2001, Legendre and Legendre 2012). Legendre and Gallagher (2001) have demonstrated that the problems CCA commonly raises can be prevented by applying a transformation that allows data having non-linear response to be analysed by RDA. In this study, the dependent variable was the site x species matrix for Taiwanese BBS data, after Hellinger transformation (square root of relative species abundance) to prevent uneven contribution of common versus rare species and double-zero bias (Legendre and Gallagher 2001). The environmental attributes in this research included 13 variables: mean annual temperature (°C, as the mean of 12 monthly mean temperatures), range of monthly temperature (°C, as the difference between maximum and minimum monthly mean temperatures within an year), mean annual precipitation (mm, as the average of 12 monthly total precipitation), range between maximum and minimum annual precipitation (mm, as the difference between maximum and minimum monthly total precipitation within an year), NDVI (as the monthly average of every 8 days' maximum value at a resolution of 250 meter), range of elevation (m), mean elevation (m), maximum elevation (m), minimum elevation (m), number of landcover types (number of different landcover types), population per county (persons/county), population density (person/km^2), and nearest distance to road (m, the nearest distance to a national or provincial road). Climatic data were extracted from Taiwan Central Weather Bureau (http://www.cwb.gov.tw), NDVI based on MODIS images (Moderate Resolution Image Spectroradiometer, NASA), and elevation derived from a 40-m digital elevation model from Aerial Survey Office (Taiwan Forestry Bureau). Landcover types were based on those defined by National Land Surveying and Mapping Center (Ministry of the Interior, Taiwan), population data derived from Ministry of the Interior, Taiwan, and distance to road was retrieved from Directorate General of Highways (Ministry of Transportation and Communications, Taiwan). These variables were generated at a spatial resolution of 1 km. Temperature, precipitation and NDVI were calculated over the period of 2009 to 2016 while population calculated over the 2009-2010 period. Multinomial logistic regression (MLR), with response variable a categorical variable (Kempen et al. 2009), was applied to identify environmental attributes important for categorisation of avian indicator species groups. MLR is an extension of logistic regression, which analyses binomial categorical dependents. When the dependent variables have more than two levels and they are not ranked data then the MLR should be more appropriate than the ordinal logistic regression model (Agresti and Kateri 2011). Firstly, among highly correlated variables (Spearman's rank correlation > 0.7, the same criteria to avoid multicollinearity, Dormann et al. 2013), only the variable with the lowest Akaike information criterion (AIC) value in MLR (with only the selected independent variable in the model) was retained for further analysis. Variables not highly associated with other variables were also included in the final MLR analyses. AIC was then used to identify the combination of environmental variables that can best discern species groups.

All analyses were conducted in the R environment (R Development Core Team 2013) using the R packages “vegan” (Oksanen et al. 2013) “igraph” (Csardi and Nepusz 2006), “indicspecies” (DeCáceres and Legendre 2009), “nnet” (Venables and Ripley 2002), and “dendextend” (Fowlkes and Mallows 1983).

4.4 Results

4.4.1 Nestedness of Taiwan Breeding Bird Survey data

The nestedness temperature for birds in Taiwan was 10.92° (Figure 4. 1).

4.4.2 Nestedness and species accumulation curve

Among the four extrapolation functions, the Arrhenius model had the lowest error rates, with 507.8 species (18.88% error rate) and 763.6 species (-21.98% error rate, overestimated) predicted based on nestedness order and random order, respectively. The average of these two data inputting orders is 635.7 species (-1.55% error rate). In comparison, the other three extrapolation functions underestimated overall species richness when based on either nestedness or random order: logarithmic model, 427.3 (31.74% error rate, nestedness order) and 480.3 species (23.27% error rate, random order); negative exponential model, 282.1 (54.94%) and 278.4 (55.53%); Clench model, 285.0 (54.47%) and 285.7 (54.36%).

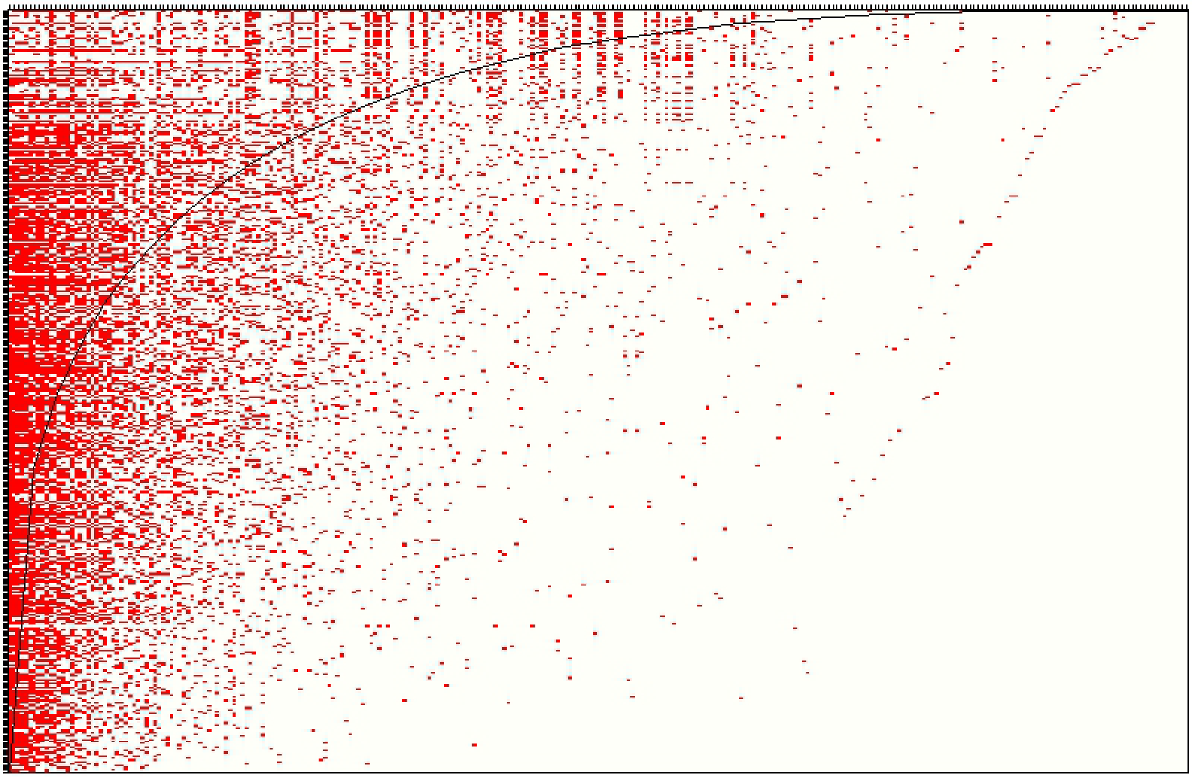
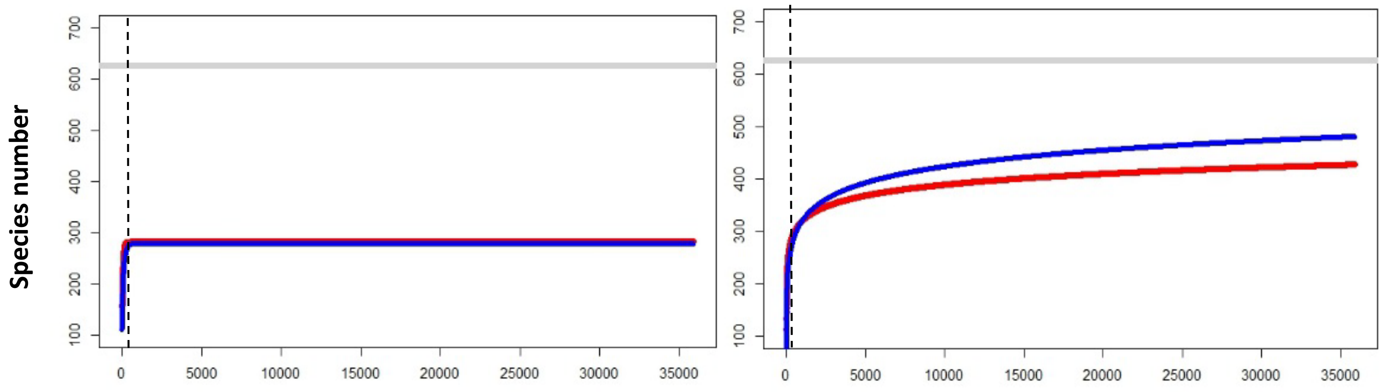
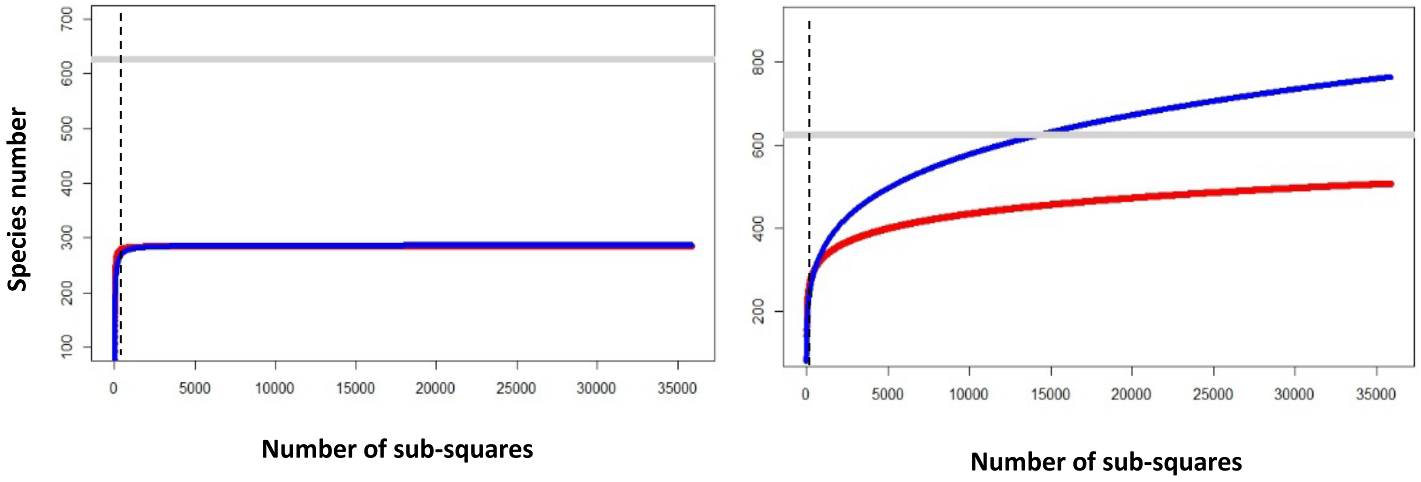


Figure 4. 1 The nestedness matrix of Taiwan Breeding Bird Survey data. The x-axis represent sites (sub-squares) and the y-axis represent species. No site or species names are shown here as there are not legible in this figure; Appendix A provides full site and species lists in nested oreder for this presence-absence matrix. Red cells indicate species occurrence and white ones, species absence. The black concave diagonal curve represents the isocline that delineates the hypothetically perfect nested matrix. The white cells locating at the left hand side of the diagonal curve represent the unexpected absence and the red ones at the right hand side, unexpected presence.



(a) negative exponential

(b) logarithmic



(c) Clench

(d) Arrhenius

Figure 4. 2 Prediction of Taiwan avian species based on four species accumulation curve functions (a: exponential; b: logarithmic; c: Clench; d: Arrhenius) with 410 sampling sites under two data arrangement orders (blue: random; red: nested). The

grey horizontal line indicates the total number of species (626 species). The vertical dash line indicates the number of sampling sites.

4.4.3 Subgroups detected by modularity versus clustering algorithms

The network of Taiwan BBS data contained 283 avian species residing in 410 nodes (sites), with 5658 edges representing linkages among sites that harboured the same species. Four network modules were identified, each containing 6, 35, 131, or 238 sites (represented by the purple, red, blue and green group respectively in Figure 4. 3 left). In comparison, the four clusters classified with the k-means clustering method contained 53, 141, 92, 124 sites (represented by the purple, red, blue and green group respectively in Figure 4. 3 right) (For species lists of each site group, see Appendix F).

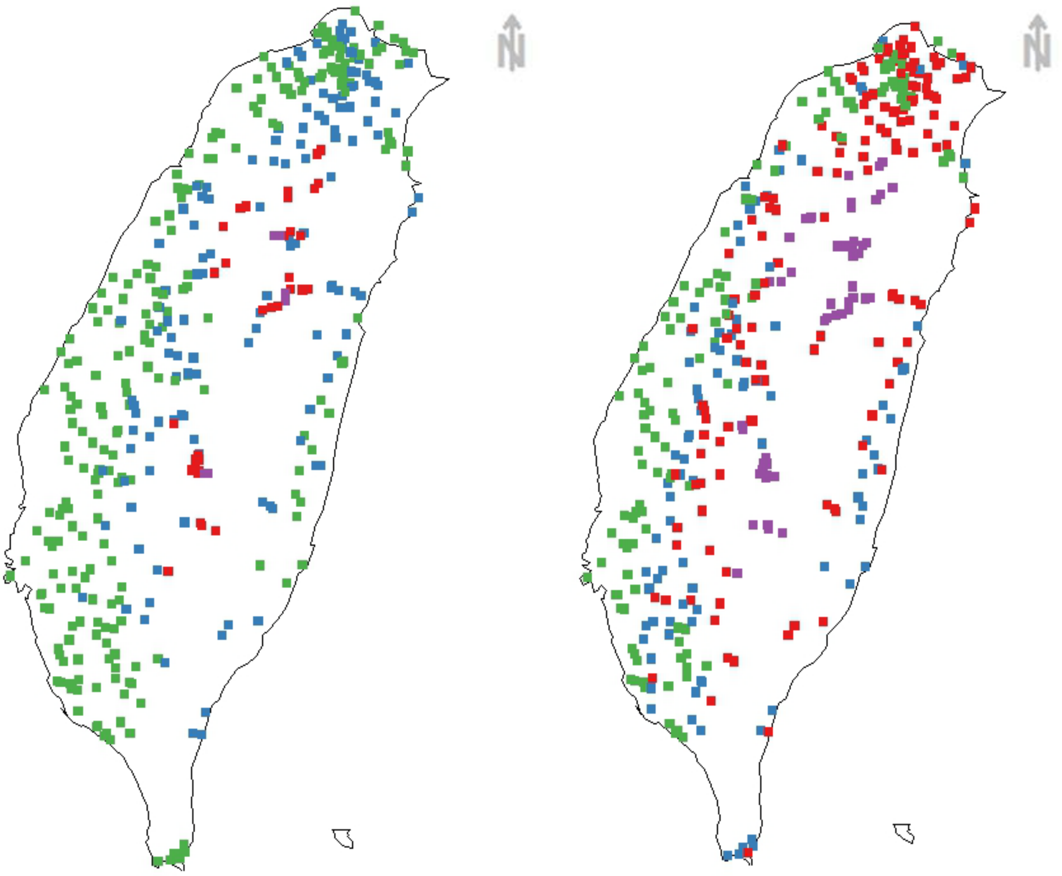


Figure 4. 3 Results of modularity detection (left) and k-means clustering (right) from the Taiwan BBS dataset. Different colours represented different modules or clusters. Colours and order of module (or cluster) are arbitrary. Each square is 1x1 km.

The indicator species selected by modularity and clustering methods are listed in Table 4. 1. These four network modules include: (1) the high elevation zone (purple) with indicator species *Tarsiger indicus*, *Pyrrhula erythaca*, *Troglodytes troglodytes*, *Regulus goodfellowi*, *Carpodacus formosanus*, and *Trochalopteron morrisonianum.* These six species form three combinations of four-species indicator groups (see Table 4. 1, left column of purple group), with the first three species always included; (2) the mid to high elevation zone (red), represented by indicators *Yuhina brunneiceps*, *Aegithalos concinnus* and *Liocichla steerii*, which are similar to indicators species identified based on clustering method (purple in Fig. 4.3b); (3) the low elevation zone (blue) with indicator species of *Psilopogon nuchalis*, *Hypsipetes leucocephalus*, *Alcippe morrisonia* and *Schoeniparus brunneus*; and (4) the plain and urban area (green) with indicator species of *Streptopelia tranquebarica*, *Passer domesticus*, *Streptopelia chinensis* and *Dicrurus macrocercus*. Different indicator species were identified when based on clustering algorithm (Table 4. 1, right column): (1) high elevation zone (purple in Fig. 4.3b): indicators species were the same as those in mid to high elevation zone (red in Fig. 4.3a) for network modularity; (2) mid to high elevation zone (red) included six species, with three species also included in low elevation zone for network modularity (blue); (3) low elevation zone (blue) included five species, with only one species (*H. leucocephalus*) the same as low elevation zone (blue) for network modularity; two of the species (*H. leucocephalus* and *Pomatorhinus musicus*) are also the indicator species for mid to high elevation zone in clustering method; (4) the plain and urban area (green) included four indicator species, with one species (*P. domesticus*) shared with the network modularity in urban area.

Higher A and $sqrtIV^g$ values indicate higher success in predicting site group of unknown sites, and higher B value represents a higher detection rate, thus more efficient indicators. All $sqrtIV^g$ are significant with $p < 0.01$, but A, B, and $sqrtIV^g$ values were generally higher in network modularity than in clustering method (Table 4. 1).

Table 4. 1 The indicator groups selected by modularity (left) and clustering (right) methods of Taiwan Breeding Birds Survey (BBS) data. *SP*= species name codes; *A* = positive predictive value to a site group of each indicator group; *B* = sensitivity of each species group to a site group; *sqrtIV*^{g} = grouped equalised square root value of *A* X *B*.

Modules	Indicator species group			
	SP	*A*	*B*	*sqrtIV*^{g}
Purple	*Regulus goodfellowi* *Tarsiger indicus* *Pyrrhula erythaca* *Troglodytes troglodytes*	0.90	1	0.95
	Carpodacus formosanus *Tarsiger indicus* *Pyrrhula erythaca* *Troglodytes troglodytes*	0.90	1	0.95
	Trochalopteron morrisonianum *Tarsiger indicus* *Pyrrhula erythaca* *Troglodytes troglodytes*	0.90	1	0.95
Red	*Yuhina brunneiceps* *Aegithalos concinnus*	0.86	0.91	0.89

Clusters	Indicator species group			
	SP	*A*	*B*	*sqrtIV*^{g}
Purple	*Yuhina brunneiceps* *Aegithalos concinnus* *Liocichla steerii*	0.97	0.81	0.86
	Aegithalos concinnus *Liocichla steerii*	0.97	0.81	0.86
	Yuhina brunneiceps *Aegithalos concinnus*	0.94	0.83	0.88
Red	*Megapomatorhinus erythrocnemis* *Psilopogon nuchalis* *Alcippe morrisonia*	0.60	0.77	0.68

Blue	*Aegithalos concinnus*	0.86	0.91	0.89
	Yuhina brunneiceps *Aegithalos concinnus* *Liocichla steerii*	0.88	0.86	0.88
	Psilopogon nuchalis *Hypsipetes leucocephalus* *Alcippe morrisonia*	0.77	0.93	0.85
	Psilopogon nuchalis *Hypsipetes leucocephalus* *Schoeniparus brunneus* *Alcippe morrisonia*	0.84	0.85	0.84
	Psilopogon nuchalis *Hypsipetes leucocephalus* *Schoeniparus brunneus*	0.82	0.87	0.84
Green	*Streptopelia tranquebarica* *Passer domesticus*	0.91	0.86	0.88
	treptopelia chinensis *Passer domesticus*	0.84	0.89	0.87
Blue	*Megapomatorhinus erythrocnemis* *Hypsipetes leucocephalus* *Dendrocitta formosae* *Alcippe morrisonia*	0.62	0.72	0.67
	Megapomatorhinus erythrocnemis *Pomatorhinus musicus* *Dendrocitta formosae* *Alcippe morrisonia*	0.61	0.73	0.67
	Dicrurus macrocercus *Pomatorhinus musicus* *Lonchura punctulata* *Hypothymis azurea*	0.87	0.86	0.87
	Dicrurus macrocercus *Pomatorhinus musicus* *Hypsipetes leucocephalus* *Lonchura punctulata*	0.88	0.85	0.86
	Dicrurus macrocercus *Pomatorhinus musicus* *Lonchura punctulata*	0.86	0.86	0.86
Green	*Pycnonotus sinensis* *Acridotheres tristis* *Hirundo rustica*	0.83	0.66	0.74
	Pycnonotus sinensis *Acridotheres tristis* *Hirundo rustica* *Passer domesticus*	0.86	0.65	0.73

	Dicrurus macrocercus *Passer domesticus*	0.87	0.86	0.87

	Pycnonotus sinensis *Acridotheres tristis* *Passer domesticus*	0.81	0.67	0.74

4.4.4 Environmental associates of site groups

The result of RDA for Taiwan BBS (site groups categorised based on network modularity) is illustrated in Figure 4. 4. There are 13 environmental variables, which means that we get 12 constrained ordination axes (one as dummy variable). 30.37% of total variance is explained by environmental factors (constrained axes) (Table 4. 2), with $r^2 = 0.31$ for the global model and the adjusted $r^2 = 0.29$. Along the RDA1, the plain and urban area (green group) was separated from the other three groups, with temperature, elevation, and NDVI as the most influential attribute (Figure 4.4). For the RDA2, the three site groups (purple, red, and blue) were separated mainly along the gradients in elevation, number of land cover types, and NDVI (Figure 4.4).

Table 4. 2 Redundancy analysis f Taiwan Breeding Birds Survey (BBS) data.

	Inertia	Proportion	Rank
Total	0.6237	1	
Constrained	0.1920	0.3079	12
Unconstrained	0.4317	0.6921	265

Inertia is variance

Eigenvalues for constrained axes:

RDA1	RDA2	RDA3	RDA4	RDA5	RDA6	RDA7	RDA8	RDA9	RDA10	RDA11	RDA12
0.10399	0.05217	0.01553	0.00439	0.00409	0.00314	0.00228	0.00200	0.00156	0.00120	0.00091	0.00077

Eigenvalues for unconstrained axes:

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
0.03163	0.01924	0.01827	0.01365	0.01118	0.00945	0.00900	0.00883

(Showed only 8 of all 266 unconstrained eigenvalues)

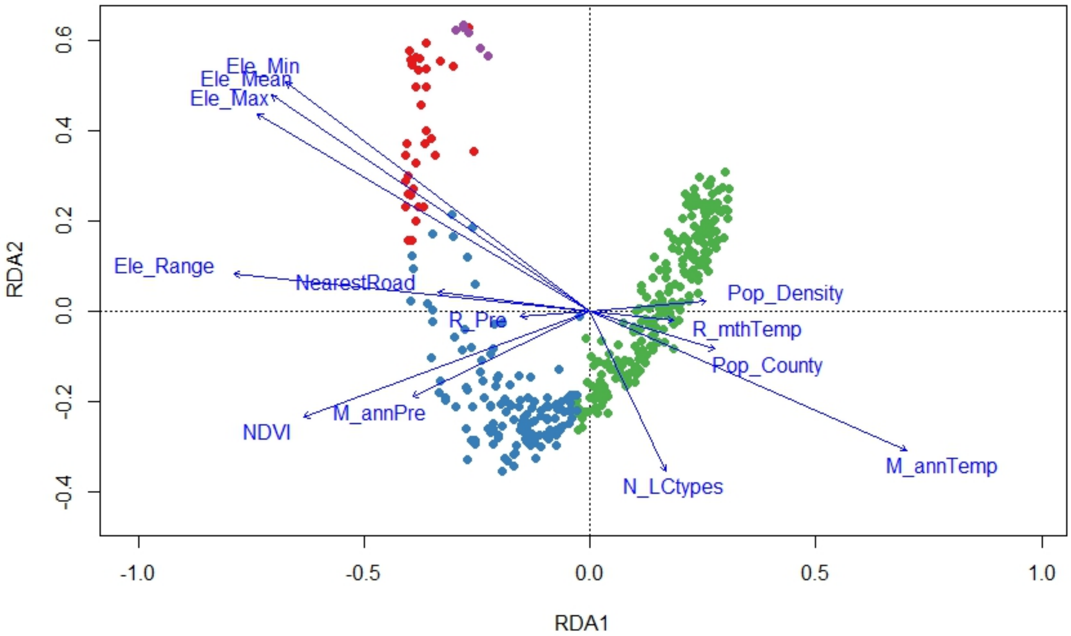


Figure 4. 4 Plot of redundancy analysis (RDA) for environmental associates with site groups of Taiwan BBS data. Points represent sites and different colours represented different modules by the modularity-based indicator species analysis (colours correspond to those in Figure 4.3 left). Environmental attributes included: mean annual temperature (M_annTemp) (° C), range of monthly temperature (R_mthTemp) (° C), mean annual precipitation (M_annPre) (mm), range between maximum and minimum annual precipitation (R_Pre) (mm), Normalised Difference Vegetation Index (NDVI), range of elevation (Ele_Range) (m), mean elevation (Ele_Mean) (m), maximum elevation (Ele_Max) (m), minimum elevation (Ele_Min) (m), number of landcover types (N_LCtypes), population per county (Pop_County) (persons/county), population density (Pop_Density) (person/km^2), nearest distance to road (NearestRoad) (m) for each 1x1 km square.

Five of the 13 environmental variables were removed from further MLR analyses due to high collinearity with retained variables that had lower AIC values (Table 4. 3). Among the eight variables that were analysed in the MLR, the model including maximum elevation height and NDVI had the lowest AIC value (Table 4. 4). The second best model contained maximum elevation height, NDVI, and population density, but the difference in AIC with the best model was > 2 (Table 4. 4).

Table 4. 3 Akaike information criterion (AIC) value for the multinomial logistic regression, with site groups as dependent variable and each single environmental variable as independent variable.

Environmental attributes	**AIC**
Maximum elevation (m)	305.27
Mean elevation (m) *	337.49
Minimum elevation (m) *	388.23
Range of elevation (m) *	428.75
Mean annual temperature (° C) *	464.72
Normalised Difference Vegetation Index (NDVI)	518.37
population density (person/km^2)	642.89
population per county (persons/county) *	713.50
Nearest distance to road (m)	734.84
Mean annual precipitation (mm)	737.64
Number of landcover types	741.01
Range of monthly temperature (° C)	768.50
Range between maximum and minimum annual precipitation (mm)	785.08

* dropped in further analyses due to correlation (rs >0.7) with other variable of lower AIC value

Table 4. 4 The model selection table for Taiwan Breeding Birds Survey (BBS) data. Only the top ten models with low AIC values are shown. The environmental attributes applied in the model selection were: mean annual temperature (M_annTemp) (°C), range of monthly temperature (R_mthTemp) (°C), mean annual precipitation (M_annPre) (mm), range between maximum and minimum annual precipitation (R_Pre) (mm), Normalised Difference Vegetation Index (NDVI), range of elevation (Ele_Range) (m), mean elevation (Ele_Mean) (m), number of landcover types (N_LCtypes), population per county (Pop_County) (persons/county), population density (Pop_Density) (person/km^2), nearest distance to road (NearestRoad) (m) for each 1x1 km square.

Model no.	(Int)	Ele_Max	M_annPre	N_LCtypes	NDVI	NearestRoad	Pop_Density	R_mthTemp	R_Pre	df	logLik	AICc	delta	weight
10	+	+			+					9	-124.983	268.4	0	0.352
42	+	+			+		+			12	-123.021	270.8	2.41	0.105
138	+	+			+				+	12	-123.172	271.1	2.71	0.09
14	+	+		+	+					12	-123.199	271.2	2.77	0.088
26	+	+			+	+				12	-123.517	271.8	3.41	0.064
74	+	+			+			+		12	-124.084	273	4.54	0.036
46	+	+		+	+		+			15	-121.05	273.3	4.9	0.03
140	+	+	+		+				+	15	-121.055	273.3	4.91	0.03
12	+	+	+		+					12	-124.48	273.7	5.33	0.024
58	+	+			+	+	+			15	-121.545	274.3	5.89	0.018

4.5 Discussion

In this study, it was found that the Arrhenius power function had lower error rates than the other three extrapolation functions in the prediction of Taiwan avian species richness. Moreover, while overestimation of species richness occurred when the data were inputted in random order as was conventionally practiced, species richness was underestimated when the data were inputted in nested order; low error rate (-1.55% error rate) was achieved when estimation of these two inputting orders was averaged (see chapter 2 and 4). Comparing with traditional clustering methods, network modularity had higher success in predicting species groups for unknown sites, and the four avian species groups were associated with variation in elevation and NDVI.

The power function (Arrhenius 1921, Preston 1962a, 1962b) and logarithmic function (often erroneously termed the exponential function; Gleason 1922) are commonly applied in the extrapolation of species accumulation curve (Dengler 2009). Selection of an appropriate extrapolation function is critical for satisfactory prediction of species richness. When species are not randomly distributed or when an area is characterised with high environmental heterogeneities, species number increases swiftly with an increase in sampled area (Williams 1943, He and Legendre 1996, Scheiner 2003). In the present study, the Arrhenius power function was better at predicting avian species richness in Taiwan; in comparison, the logarithmic function was the most suitable model for predicting avian species richness in Britain (Chapter 2). This difference is in agreement with previous studies that show the power model is the most appropriate extrapolation function when sampling effort is low and species richness is still far from saturated (Soberón and Llorente 1993, Tjørve 2003). In Taiwan, when many avian species are still not detected in Taiwan BBS, the fast-growing curve of the power function should be more appropriate than log or other functions.

Both the power and the logarithmic functions lead to a convex upward species accumulation curve that lacks upper asymptotes (Dengler and Boch 2008, Dengler 2009), thus risking an overestimation of species richness (Ugland et al. 2003). The sequence of adding in new species affects the curve shape and leads to different predictive results (Ugland et al. 2003). As expected, when data are inputted in random order as conventionally applied, the power function overestimates species richness (Figure 4.2d). However, data input based on nestedness orders provides a novel way in constraining potential overshooting of power functions and helps set a lower bound to the extrapolation curve (Figure 4.2d). A similar moderating effect was observed in the British study (Chapter 2) when the logarithmic function was applied. Interestingly, when species richness predicted based on nestedness order and random order is averaged, the mean is 635.7 species, which is very close to the documented avian species (626 species) in Taiwan, considering that only 410 1x1 km plots out of Taiwan's 36,000 km^2 area (about 1%) are sampled. This accurate estimate of species richness is due to the underestimation and overestimation when data are inputted in nestedness and random orders respectively, so combining them will errors out. The same cancelling-out effect was also observed in the BTO study (Chapter 2); as such, combining nestedness and random orders has potential to become a novel and important approach for predicting species richness, but requires further validation.

The k-means clustering method is descriptive in nature, its robustness is hard to validate and is also an indirect method, relying on distance (difference) between species rather than directly measuring co-occurrence patterns (Milligan 1980, Clarke 1993, Legendre and

Legendre 2012). It is also sensitive to double-zero issue, which is very common in ecological datasets and may cause group misclassification (Dufrêne and Legendre 1997, Legendre and Legendre 2012). The double-zero effect might be particularly severe in a heterogeneous environment (e.g. tropical areas) because the fact that a species is absent in two sites does not indicate that these two sites have similar environments. In this study, indicator species groups in Taiwan had clearer group boundaries when selected by network modularity than by the clustering method. Species composition is different when the site groups are determined based on network module vs. clustering: species overlap is less in module-based than clustering-based site groups (that is, most species occurred in >1 clusters if the site groups are classified based on the cluster algorithm, see species list in Appendix F). My previous study (Chapter 2) also demonstrated the superiority of network modularity over the clustering method, indicating the higher utility of network analysis. Compared with clustering, the British modularity groups also had clearer geographical boundaries, less overlap in species compositions, and higher predictability in unsampled sites.

Taiwan BBS data have been divided into four groups (north, west, east and mid-high elevation) based on geography and altitudes (Ko et al. 2013). By contrast, group classification by indicator value derived from similarity in species composition among sites (De Cáceres and Legendre 2009) can result in a more objective group categorisation. Indeed, the four groups identified by the indicator species algorithms are different to the four groups determined by geography and altitudes (Ko et al. 2013). A similar difference was also found in coral-fungi symbiosis study (Amend et al. 2012), in which indicator species analysis revealed a different classification from the conventional classification methods (e.g. coral phylogeny, environmental attributes such as water temperatures…etc.). Composition of fungal species was highly variable in colonies of a single coral species and capitalising on subjectively selected classification might not reveal factors that determine fungal species composition. Instead, indicator species analysis indicated that several fungal species were only statistically associated with specific corals which thrive in warm water (Amend et al. 2012). Therefore, indicator species analysis is valuable in categorising meaningful species groups, which in turn helps reveal the environmental determinants that define and differentiate communities.

Indicator species in the four sites groups were generally very different between the modularity and clustering method. The exception is the modularity mid to high elevation zone contained the same indicator species as clustering high elevation zone (Table 4.1). These indicator species include *Y. brunneiceps*, *A. concinnus* and *L. steerii* that are commonly observed in mid to high elevations in Taiwan. These three common species are included in the zones of the highest elevation by the clustering method, but are included in the second highest elevation group by the modularity method. An extra site group (high elevation zone) that is attitudinally higher than mid to high elevation zone (Figure 4.4) is identified when based on the network modularity. The six species included in the modularity high elevation zone are generally distributed at higher altitudes but are more spatially limited than *Y. brunneiceps*, *A. concinnus* and *L. steerii*. The two methods might classify indicator species in different ways: the modularity method highlights species of more limited distribution whereas the clustering method favours common species. Therefore, the modularity method may also be valuable in identifying species of conservation concern.

Previous studies on environmental determinants of Taiwan's birds have focused on avian species richness (Lee et al. 2004, Ding et al. 2005, Koh et al. 2006), leaving unexplored how avian species assemblages are environmentally differentiated. While species richness can be easily documented, objective classification of species assemblages is challenging. Capitalising on network analysis, for the first time, Taiwan's avian species assemblage is differentiated, and it is further demonstrated that such differentiation is based primarily on elevation and NDVI. Primary productivity (with NDVI as a surrogate), elevation, and urbanization are also found to be associated with avian species richness in Taiwan (Lee et al. 2004, Ding et al. 2005, Koh et al. 2006), indicating that birds in Taiwan are largely characterised by elevation and NDVI. Similarly, during several glacial periods, high mountain areas in Taiwan were the refugia of Palearctic and Himalayan species which migrated from the continent via landbridges (Tsukada 1966, 1967, Lee et al. 2004). Mountains in Taiwan might thus shelter species that are different from those in the plain areas, leading to a differentiation in bird assemblage with elevation. The difference in NDVI might reflect a difference in primary productivity or human disturbance, which can both affect avian species richness in Taiwan (Lee et al. 2004, Ding et al. 2005, Koh et al. 2006). Places with lower NDVI are typically where a higher degree of human disturbance is experienced, which commonly harbour different species. In Taiwan it is often hard to disentangle the effect of elevation, NDVI, level of urbanization, and temperature, because an increase in elevation is typically accompanied with an increase in NDVI and a decrease in urbanization and temperature. However, altitude should not be taken as a direct driver for biodiversity, because its relationship with biodiversity is not causal. The underlying environmental factors of interest (NDVI, temperature and urbanization here) are the ones to focus on (Hawkins and Felizola Diniz-Filho 2004, Kluge et al. 2006, Field et al. 2009, Fattorini and Ulrich 2012).

Chapter 5: General discussion

A change in community structure reflects underlying environmental variation, and can potentially advance our understanding of biotic communities. In this study, I focused on two community properties: nestedness and co-occurrence. Nestedness is a non-random species distribution pattern in which species in depauperate sites are contained in species-rich sites. On the other hand, co-occurrence networks categorise species assemblages as different units that reflect differential habitat requirements. Capitalising on these community structures can improve conventionally utilised predictive models such as species accumulation curves and indicator value analysis.

Predicted species richness based on species accumulation curves varies markedly with the functions applied. For example, Soberón and Llorente (1993) compared negative exponential, logarithmic, and Clench functions and concluded that functions vary in their performance. Selection of the appropriate extrapolation function is thus critical for satisfactory prediction of species richness, but there is no universal principle for the selection of the optimum function. In general, negative exponential and Clench functions perform better in homogenous environments due to that both functions belong to saturation model; that is, there is an upper asymptote and the accumulation curve eventually approaches a saturation point (Dengler 2009). In comparison, the power function (Arrhenius 1921, Preston 1962a, 1962b) and logarithmic function will create unbounded curves without mathematical asymptotes, and are more suitable for heterogeneous environments where new species continue to be found for longer, despite the fact that this function tends to overestimate species number due to an infinite increase

(Tjørve, 2003; Dengler, 2009). Furthermore, although both logarithmic and power functions are unsaturated curves without asymptotes, the power function outgrows other curves because it follows the power trend, i.e. a variable base is raised to a fixed exponent. In this dissertation, I have shown that accuracy in the prediction of British avian species richness increased with sampling effort and error rates were much lower when fitted with logarithmic function than with exponential or Clench function under both British and 100 km squares spatial scales (Chapter 2). In comparison, the Arrhenius power function had lower error rate than the other three extrapolation functions in the prediction of Taiwan's avian species richness (Chapter 4). Such difference is in agreement with previous studies that the power model is the most appropriate extrapolation function when sampling effort is low and species richness is still far from saturated (Soberón and Llorente 1993, Tjørve 2003). In Taiwan, when many avian species have still not been detected in Taiwan BBS, the fast-growing curve of power function should be more appropriate than logarithmic or other functions. A method to estimate completeness of sample (i.e. the proportion of species observed) derived from Alan Turing's frequency formula has recently been developed (Chao et al. 2013, Chao and Jost 2015). Completeness can be accurately estimated based on the proportion of 'singletons' (species with only one individual in the sample), with more singletons representing low levels of completeness. Such estimation of sampling completeness might help inform how far species richness is away from saturation and is thus helpful in the selection of optimal extrapolation function. Further studies including information of sampling completeness might help to determine the criteria for function selection.

I also demonstrated that ordering data based on nestedness could improve predictability of species richness. Most recent research on nestedness has focused on a depiction of community structure (Picazo et al. 2012, Herczeg and Horvath 2015, Chaves and Ariez 2016) or on an improvement of the algorithms for calculating the nestedness index (Almeida-Neto et al. 2008, Podani et al. 2014). The current study, by capitalising on nestedness, provides a novel approach for improving the predictive power of species accumulation curves. Specifically, while species richness is usually overestimated when the data are inputted in random order as is conventionally practiced, species richness is underestimated when the data are inputted in nested order. An average of projected species richness of these two inputting orders remarkably lowers prediction error rate in Britain and Taiwan even though the two countries have distinctive topographic characteristics, suggesting a general applicability of this novel method and is thus worthy of further validation in other study sites and other taxonomic groupings.

The nestedness-ordered dataset applied in this study can similarly lead to a better prediction when combined with the best fit extrapolation functions (e.g. the logarithmic function for British birds and the power function for Taiwanese birds). For instance, the error rate was <5% based on nestedness order compared with 12% based on random method when logarithmic functions were applied for the estimation of British avian species richness in 100 km squares with 10% sampling effort. This suggests that incorporating data structure (e.g. nestedness) into the analysis could help improve the efficiency of sampling effort, particularly useful in area with limited resource available for field survey (Pearman and Weber 2007, Ashcroft et al. 2010). In the findings for the Taiwan avian data were similar: using approximately 1% sampling data, with the power function, could predict overall species richness with less than 20% error rate. Moreover, conventionally intensive randomisation procedures are needed for the species accumulation curve to achieve higher predictive power (Ugland *et al.* 2003; Chao & Shen 2004; Gray *et al.* 2004) by reducing the probability of including extreme values with a large number of replications. I show here that sampling based on nestedness order is a novel way for capitalising on community structure for the prediction of total species richness. Results of this study have demonstrated that, overall, sampling according to nested order is more favourable than based on random order. The nestedness method applied in this study captures information about community composition and can accumulate as many species with lower sampling effort/areas. Thus taking advantage of the nested structure has the potential to predict community composition by focusing on the most species-rich sub-squares. However, because both British and Taiwan avian assemblages demonstrate a similar level of nestedness, it is not possible here to assess the relationship between the level of nestedness and the predictability of species accumulation curve. More studies that consider different levels of nestedness (e.g. including simulated datasets) are needed to test for the general applicability of nested order.

Relative to spatial nestedness, examination of nestedness over time is less studied, although it can provide additional information on temporal species coexistence (Elmendorf and Harrison 2009, Heino et al. 2009). When studying Britain’s avian species nestnedness, species with more limited occurrence include vagrants which occur less frequently in time rather than in space. In other taxonomic groups, such species may have interannual effects on co-existence (e.g. early-geminating and fast-growing exotic plants outcompete the slow-growing local species). Application of nestedness species accumulation curves could potentially be extended temporally for predicting interannual variation and could have important implications for long-term conservation or monitoring.

Compared with k-means clustering, indicator species identified by network modularity can more successfully assign unsampled sites to the correct species groups. Indicator species analysis is useful for recognising representative species and has been applied to various research topics, such as plant species composition (Flinn et al. 2008, Willerslev et al. 2014), vector ecology (Obsomer et al. 2013), mycological community (Taylor et al. 2014, Varela-Cervero et al. 2015) and intestinal microflora ecology (Seedorf et al. 2014, Planer et al. 2016). The accuracy of site type prediction relies on robust classification of sampled species and sites (Dufrêne and Legendre 1997). The results of clustering, however, can be easily influenced by other factors. Firstly, because cluster analysis is based on dissimilarity among members, the clustering outcome depends on which sites are sampled. Different sampled sites lead to different clustering outcomes, resulting in inconsistent species-site associations. In comparison, network analysis measures species co-occurrence directly and partitions the subgroup objectively and is therefore an efficient substitute for group partitioning in the indicator species analysis. The second concern is that subgroups partitioned by clustering methods are arbitrarily defined by user-defining cut off thresholds and may create artefact groups (Legendre and Legendre 2012). Although k-means clustering is recommended to be utilised in indicator species analysis (Dufrêne and Legendre 1997), this method can be easily affected by the identity of the group to which the original node is assigned (Celebi et al. 2013). Even though Dufrêne and Legendre (1997) suggested repeating the partitioning process until the indicator values of all species are decreasing to find the appropriate number of clusters, there is still no reliable criterion to decide the cluster numbers, and the number of indicator species groups classified by the k-mean clustering needs to be subjectively decided. As such, indicator value does not take into account species absences (DeCáceres et al. 2010); different numbers of site group are repeatedly attempted until the optimal number for acquiring the most confident indicator species value is found (Dufrêne and Legendre 1997, DeCáceres and Legendre 2009). In comparison, the number of site groups is objectively determined by network modularity.

Network analysis with spatial modelling (i.e. the spatial network analysis) has been frequently applied in many fields (Barthélemy 2011), however in ecological research, incorporating ecological processes such as dispersal, competition, or food web into network analysis to uncover geographic pattern remains challenging (e.g. Cumming et al. 2010) and most of the research focus on single species or individual (e.g. Saura and Rubio 2010, Pereira et al. 2011). This study illustrates that network analysis could efficiently classify subgroups of co-occurring avian species in Britain according to both species co-occurrence and the geographical distribution. The complete bird assemblage is sorted into subgroups according to occurrence in coastal or inland area, along the north-south gradient. The sites are divided into several groups, and sites within a module are usually geographically connected compared to the geographical mosaic pattern of clustering results. Also, the Taiwan BBS data have been divided into four groups (plain-urban areas, low elevation zones, mid-high elevations, and high elevations) based on geography and altitudes. This suggests the ability to combine species (co-occurrence) and spatial information by network analysis. This technique has also been used for analysing the large-scale European fauna structure (Araújo et al. 2011) and successfully revealed the robustness of each faunal group by the geographical distribution and the resistance to climate change. Regarding indicator species analysis, it delivers more consist result in defining species-site associations therefore better basis for the predictability. However, such capability has not attracted enough attention and more future research is required.

Scale also affects the predictability of community composition. For example, it has been shown that the most species-rich areas shift northward when the research spatial scale enlarged although the mechanism underlying the shift remains elusive (Lennon et al. 2001, Willis 2002) and biological similarity typically decreases with geographical distance (Soininen et al. 2007).

In the British avian study, the finest spatial scale used for indicator species was 10 km. However, for a mobile taxon such as birds, when data of finer resolution are available, more relations among modules may be uncovered. This increases the robustness of the site group classification and may improve predictability of species composition. How to decide the most appropriate spatial scale requires further research.

In this study, I have demonstrated that it is feasible to utilise orderliness in an assemblage to improve predictability of species richness in unsampled areas. Also I have shown that network analysis can help classify indicator species group. An investigation of both British and Taiwanese bird data supported these conclusions. An assessment of other areas or countries could validate the generality of these pattern-based algorithms. Future research should also explore topics regarding nestedness along the temporal dimension and the significance of scale in predicting species richness and composition.

Appendix A

A.1 Nestedness order of 10 km sub-squares in Britain

* Codes are listed in nested order, from the species-richest site to the depauperate. Read by row, left to right.

* For complete grid reference and map, see http://www.bto.org/volunteer-surveys/birdatlas/methods/correct-grid-references/know-your-place

ND27 TL87 TF74 TF84 TF62 TQ77 TM47 SZ19 TR06 SU70 TM46 SK80 TG50 SU56 NG82 TG04 TL91 SD47 SZ39 NR26
SU86 SH37 TG40 NC96 NT48 TL77 NH80 NR64 TL78 TQ81 SU40 SE27 TL89 NT96 TG30 NS48 TQ58 NC55 SU10 TQ96
NT67 TL19 ND35 NT16 TG31 SD84 SU50 SH57 NY22 TL15 TG42 TL97 TR26 NR78 NT25 NU00 NS47 NO35 TM45 NR86
SD27 SU66 SD41 SU31 SJ24 TG32 NR39 SZ49 TL29 NH62 SU11 SU67 SE74 TM11 NH85 TR35 TM34 TF63 TR16 NC86
NM45 SH28 TQ86 TM22 ND06 SN69 TQ49 SU30 SU94 TF92 NU22 SU97 TQ91 NX66 SU36 SK46 NN80 SU77 NT86 SH48
SD48 NZ71 TL48 SK47 SE30 TL88 NZ52 SU76 TQ06 NC82 NR98 SU60 SK32 SY68 NN79 SD72 SD83 TR36 SD18 NH95
TF55 TQ40 NH79 TL31 NH91 NX67 NS28 NC76 NR97 NR83 SU01 NG77 SU84 TM57 NR73 SD57 NH52 SD17 NS83 TQ87
SK29 SU32 NN81 NO59 NR89 SZ29 TF61 NR84 SJ58 SD26 SK20 NC14 SP29 SU75 TM23 NS37 NO04 TM01 ND05 SH38
SD75 SK12 NS03 SJ61 SO96 NR88 NM48 SK28 TM28 SP92 TL57 SK27 TL08 NH63 TM35 SK22 TQ98 SZ59 TQ55 SY67
TF71 TM59 NH33 TF70 TL21 ND34 TM02 NB43 SJ68 TL56 NR25 NN17 NT68 TQ76 TF10 SN20 SK67 TL30 NT84 NG25

TQ92 TF00 NS07 NY36 SK07 NO42 NM94 TL99 NZ04 SE36 SO78 NS57 NH45 TL92 SX97 SE40 NC13 SJ07 NX74 SJ85
NJ01 NY00 SE20 SE82 NJ05 TG41 ST55 SU73 ND37 SJ69 TL39 TL37 SK51 TQ16 SU42 NS35 SD59 SU20 NH54 NX06
NY55 NM55 SD62 SU46 NM54 TQ00 TA14 SD55 SD65 NR99 SP93 SJ96 NT06 SU96 NH70 SJ04 NY25 SU21 NY60 TQ07
SU88 SH87 NT42 NX46 NK02 HY20 NM86 NY15 TL03 SH46 SC27 NX76 TR01 NJ06 SE64 SZ38 TL81 TL14 SJ18 SJ81
NS94 NH77 TA04 TQ59 NO10 SP91 NC36 NM83 SP77 NY01 NT35 TM58 TL10 NY86 NJ26 NO65 SU80 SO70 NM82 NR57
NL94 NY95 SJ92 NK05 NH53 NH55 NG99 TL09 NR36 NR72 TL11 NT77 TR05 NO66 SN52 ST44 SZ68 SO30 NS38 TF11
SH63 NH90 ND16 SP89 NR96 NX96 NR77 TQ19 SK21 NH59 NX68 SJ74 NS77 NN51 SC49 SZ07 NC46 SN96 ND25 NC26
TQ64 SE24 TM13 SU83 SJ75 SH97 NN96 SO39 SY89 HY44 NS27 NZ43 SK53 SH61 NC02 SV81 TQ78 TQ57 TL04 NM42
NX40 SK60 SR99 NN85 NZ35 SO60 SU44 SE26 SU85 TQ08 SN42 SD96 TQ09 SU62 NR75 NO36 TF81 SJ80 NX15 SK11
TF80 TA16 NM71 NS65 TF64 NN72 NR95 NH20 TR25 NH78 NR85 SY58 TQ01 NG08 NT27 TL98 ND01 SU47 SK78 SD85
NC85 SD38 NM66 SU26 TM24 SU33 NX85 NH42 SJ73 NT58 NT19 TQ67 SO12 NM44 SD39 NR34 NC53 NS89 SE92 NH43
NF72 SU82 NH64 SU58 SH71 NG74 NR46 SK90 SH47 NT24 NO25 NY02 NX37 NY51 SN79 SK36 TM49 NT17 SN62 NH30
NN53 SJ37 TL28 SD29 NO40 SO61 NH65 SY99 NG32 SS42 SD28 SU93 NR79 NF75 SJ30 SZ08 SD31 NJ04 SC28 TQ66
SU09 NC64 NH32 SO15 TF30 SJ50 TR04 SU57 NR63 SD74 SJ51 SD95 TF52 SJ17 TQ33 NM73 SD82 NU20 NO16 SU34
NO06 SH39 NC91 SJ08 TM39 SJ95 NO53 SD64 NY54 SK26 HY22 SH76 SU41 SE09 NC80 SK87 TL22 SJ31 SO57 SU12
NO14 ND03 NO01 SH44 SE91 TL20 SC39 TL26 SD56 NN18 SU87 SD32 NS68 NN95 ND23 NR65 NN69 SE35 NO00 NH69
NO03 SK69 NG88 SN77 NC21 NM81 SJ72 NZ10 NY03 TQ63 NC60 NX16 SH36 SE05 SN63 ND49 SU43 NM93 NM49 SN40
SU91 SD09 NN55 SE63 SU72 SU55 ND26 NX91 NX89 NS29 SD67 SJ63 NY08 SK00 NT07 SU00 NC56 NR15 NN75 TG14
SK68 SX64 TL02 NC25 NS41 NH87 NZ42 SD86 NX45 NC71 SK02 HY50 SE29 SU22 NM15 NT15 NR35 NX25 NZ00 NT83
NZ29 TQ79 NT92 NS93 NS40 NY30 TQ68 NT47 SU53 SD61 HY45 SY78 NZ27 NN84 HY32 NH67 TF94 SE73 NJ41 TQ73

HY42 SN14 NS75 SE19 SH56 NS58 SE50 NJ13 TA17 NM62 SN30 SX98 NY12 SD60 HY75 NC45 NH93 NH89 SK79 SO38
NJ46 SK18 NH94 NR87 NH22 SH75 NS98 ST59 SJ43 NH26 NR74 SJ87 SP97 TQ39 HU31 NS49 SZ09 SK48 NN48 NY45
NM76 SU18 SK09 NT31 SE80 SH43 SJ41 TL86 SE17 NH12 SJ71 SK57 SD69 SU37 NJ14 NZ36 SO79 SN31 TL58 TL44
NG92 NO49 TF82 NZ37 NY35 NN34 NY37 NH46 NR45 SD54 NX44 SD58 HY62 NT74 TG10 SK39 NF99 ND04 NC65 HY30
TM14 NX56 NF76 NB24 SH12 SE34 NY94 NZ16 NM70 SJ57 TQ65 SJ40 SK66 NH68 NM64 NX98 NY06 SH77 NN83 SJ91
SY08 SJ60 HY51 TQ72 TA27 NT26 SY98 TF44 HY43 NR44 SD46 NF74 NS09 SH95 NX35 NH10 TG01 NS36 SE88 NJ02
SU61 NT43 NK14 NC03 NR47 NH51 NJ24 NN86 NS05 NT05 SJ53 SO28 SJ22 NM56 NZ05 NN71 NO41 TL70 NR62 SK43
ST56 NM04 NY75 NX57 NT00 TM03 SD73 SO51 SJ46 TL05 SJ14 NS18 NM67 SE18 SU98 NH75 NS61 TQ23 ND13 TL69
SE10 SX87 ST84 TQ93 SJ62 NY65 SO31 SP44 NR71 NN82 TQ46 HY21 SP86 SO84 SN95 NY52 SH27 SD90 NJ03 SK13
NM52 NY87 NC66 NH82 NC31 NT65 TL13 HU35 NT56 NX69 NU21 NR68 NC75 SN41 TQ53 NT37 SD78 SU27 NN30 TF93
TL59 NT82 NG30 SS59 NB13 SU03 NS19 SJ66 NN06 NZ03 SU65 NT95 TQ13 NZ15 TQ75 SP28 TF72 TQ04 NS46 NF73
NC35 NX78 SS43 SJ82 NY41 SK91 NM34 NH66 NB03 SJ45 SP58 SD91 SK86 SH52 SN67 TF16 SH70 TM12 NO37 TQ43
TQ27 SO77 NX27 SE83 NC90 NG20 TQ70 NN28 NH56 NO02 NT13 NZ61 NY13 TG22 SE87 NH49 NT66 SN54 NN91 NW96
NY63 NT73 SO11 NG09 TL41 SP45 SN84 SO47 NN59 HY31 NT44 NH76 SP46 NH81 TQ29 SH72 NM35 SO67 TQ25 NS86
TQ89 SN74 NX95 TQ44 NG36 NX64 SZ48 SO02 HY53 NC44 NR27 TR34 HY64 HY33 SK19 TL71 NN74 NC30 NX34 NS08
SH53 SU13 TF60 SJ84 NH19 TM44 SP37 ST90 NR70 ST60 NG89 NS55 NO21 SJ78 NO45 SE25 NN45 SE01 NF87 SK54
NG42 TL96 NG33 NG91 NG41 TL60 NY97 SK94 TR15 TV59 TQ56 NH41 SP48 SH50 NX18 NN00 NR94 TQ54 NT04 SU52
TL84 ND29 SJ20 SX84 SK08 SO83 HY74 NN65 TG21 TL12 NY76 ST49 NJ15 NZ02 NO30 ST24 NY91 NT14 SN50 HY40
SJ02 SU59 NH72 ND15 NT33 SU17 NY92 NR61 NY16 NX77 SJ94 NX43 NO13 NY70 NB54 NM97 SE98 NS76 SH74 NT18
NC01 NS67 TQ05 NS32 TQ45 SN78 NO50 NY42 NM53 NT45 SN51 SP50 NC74 TQ17 SD97 ND48 NS02 NH96 SJ93 NG83

NG73 NC11 SH73 NT29 NN94 NJ36 NC84 NT90 NN93 NY11 TF14 NJ50 SJ83 SO69 SO50 SM70 NY40 NR37 NO55 NT34

SP25 TQ14 NZ01 TL52 TQ41 TA02 ST87 SO37 SU51 SP55 NH48 NM46 SY09 NS15 ST93 NH06 NC51 NC24 NX47 SO34

NF96 NH98 SJ33 NC37 SE16 NN92 TF21 TF22 SO27 TM29 SJ39 SD88 NG61 SJ23 TR03 NN76 SU63 NC23 SJ76 SD45

SK56 SJ70 SP95 NH73 NS63 NT64 SU99 ST36 SU14 SO98 SH49 NJ33 NC70 TQ35 SP74 SZ79 SE07 SO58 NO24 TA05

SM93 SX99 SY97 NR92 SK55 SK15 SK50 SO48 SN57 SE21 SK63 NN70 TQ26 HU32 NZ51 NJ35 TL06 NM74 ST50 NN38

ST57 SE90 SK75 SO71 SH33 NT75 SC16 HU58 NJ25 NS45 NN10 SO88 NS66 SK71 NT71 NS39 TL16 NJ32 SC26 SJ97

SN68 SO62 NO64 NO31 SD66 SH60 NT30 SU23 SN35 SJ52 TQ84 SO94 HU44 SD49 NX29 SZ58 SN66 SE37 SP34 NC50

TL46 NN19 NS84 NT12 NH15 SJ99 NH92 NN31 HY63 TF02 NG85 NY50 SP40 SH68 SO49 ST47 NZ07 SJ06 SN58 SN65

NT02 SU45 NZ21 SK37 NO76 SU81 NS04 NT03 HP61 SD81 NR49 HP50 SU74 SK35 NX55 NJ66 NG52 SK58 SS21 NT28

TL76 TL24 ST61 NC83 SO93 TL79 SD89 ND02 SN92 SE61 SN82 NN90 SE14 ST66 SU54 ST39 SP85 NX87 TR37 SP10

SK24 SU15 SD93 NM43 NM75 TF01 NC63 NS85 TA22 SO59 SC48 NM96 TF25 NF98 SE28 NM69 SO09 NU04 SO87 SN61

NN24 SU90 SD76 TQ42 SP94 NM63 NG94 SD37 NJ31 SE42 ST51 NZ13 NG72 NC10 SP90 SO73 NC54 SO95 NG87 SP76

SN99 TG23 SO05 TL00 SP75 SS53 SH85 NM32 NZ28 SD87 TF26 NY88 SU28 TL18 NR69 SP35 NG35 NH25 NZ06 SX88

NS72 TF09 SO04 SD79 NY56 NH18 NR56 SK65 SH94 NG44 SJ01 ST23 NR58 SO40 TM21 NT60 TL32 TQ32 ND07 NT23

SE43 TQ15 TL68 NN63 TQ97 NY62 HU69 SK59 SH96 NS91 SP19 NH23 SK96 SJ86 TQ85 SH84 SU08 TG02 SS44 NT57

SE15 NS17 TM15 SP01 NJ83 NS96 ST04 SH64 NJ40 SN89 SK64 NT22 NT53 ST69 TM38 SE86 TL72 TL50 NN68 SK41

NZ41 NM80 SO68 SN94 SE60 SK44 SD50 TL82 TQ74 NN89 NS64 SK33 TF33 NS56 NH31 NN07 TL94 TM33 SZ69 SX05

SP13 SX73 NM39 SU71 NF97 NB53 NT54 NR60 NY90 TQ48 SK49 SO20 SS78 NN50 NT55 SP99 HT93 ND14 TL53 SX17

NT32 SH81 SN86 NC40 TF73 TG11 NO12 SK61 TQ62 SD30 ST46 TR13 SJ55 SS22 NH44 NH74 TG33 SJ42 TF06 SU19

SJ32 SY59 NG96 NS73 SH58 NU02 NT62 SE00 NG53 SJ90 SO21 TF83 NT52 TQ71 TR09 SK06 NX99 SS88 NS95 NY14

SO92	SS54	SK45	SE95	ST95	NS99	SK38	NS71	NY99	NH35	NN29	NX38	TQ50	SH92	SN73	NF86	SU92	SP09	NO05	SU79
ST32	SE31	SP65	NB56	SD19	NC95	NY38	SK34	SH65	SJ11	SS79	NR38	TL74	NS06	NT63	NY33	TL80	SE49	NY17	TM05
SE72	NM84	NN33	TF49	NF71	SC37	NN41	ST35	NX65	NZ08	NC22	SO86	NU23	NT36	ST20	SO14	ST38	NT61	NY34	SK14
NS92	SC38	SH67	NZ24	SX77	SP05	SN81	NY57	NO20	TQ02	TQ47	SO36	TM07	TQ11	TQ03	SP64	TL27	SH83	ST29	NR76
TL01	NT08	SO41	SP54	TA15	NY23	SE85	NY24	NX75	SO13	SP12	TQ38	SN76	ND39	NC81	TL83	SE33	TQ24	TM04	NG93
TM25	NY48	NG81	NS97	SO82	SD68	SS49	ND12	SU07	NO54	SJ79	NN54	TG13	NO68	SH22	NR93	SE46	NY53	NT39	NY96
NX79	NN46	SE65	SE44	SU35	HU43	TQ61	SP06	SK04	NN04	HU68	SU02	TF20	SH93	TF91	SO46	SO89	NZ09	NX92	SO91
NJ91	SJ67	NY47	SO75	NT70	ST13	SK25	SJ59	SU95	SE71	SK03	TQ22	SJ49	NY32	NO26	NX04	SE69	NN40	NG62	NT94
TA26	SE06	SD99	SO43	SN72	NG70	SO35	SD42	NM26	SP53	SW87	NH84	ST94	SS64	SD92	SN75	SO01	SZ89	NX84	NT10
SP32	SH55	NJ65	NM65	NM85	SH62	NH39	NH40	NH34	ST58	SO08	SO65	NU03	SP68	SD34	SP36	HP51	NY80	SK88	NT76
NX97	SO24	SK83	ST96	SD77	SE97	SV91	SS84	SK05	NH05	SE62	SP51	TF15	NH09	NX58	NX86	HY52	SE54	SE47	NN64
ST98	NZ31	SO72	NN20	TA11	SP42	NF77	ST74	TM48	SE08	HU34	SP47	ST48	NS81	NZ25	HU49	TL67	NS90	NY26	NS33
SP30	ST45	TQ99	NO15	NC00	NO46	SO90	SE58	SO63	TL75	SO26	TL23	SN12	SK74	SP22	NZ40	HY41	NX59	SO55	SO74
SP80	SE75	ST11	ND24	HU15	ST76	NY85	SE51	NG97	SU24	NJ81	SE89	TA21	NO19	NM25	NY21	NZ19	SN53	SE45	TL66
NB45	NU13	NO47	SP88	NT20	SO53	NJ92	ST03	TF50	HU46	TL65	SE02	NB32	SJ28	NU01	NZ23	TL90	NS62	SP17	NZ20
TG03	SN83	ST08	NO69	NY31	NS69	NY93	NO27	NR24	TQ34	SS48	SO97	SN98	TL45	SE53	ST14	TF39	SO19	NY66	NN09
SY88	SN64	ST54	TQ94	SK84	HU36	NO99	NG78	SY79	SM83	SJ05	TL93	TA13	NK03	NG24	TG00	HU25	ST79	SS80	SE39
SN03	SD98	NY81	SX59	SE57	SO44	NY43	SU78	TL25	NG26	NS82	NM98	NT85	TA30	ST28	SP84	NN01	HU42	SO52	SU29
HU45	SS94	NB91	SO29	NR82	SP70	TQ10	SO42	NN43	SD71	NX30	SH90	NY82	NX05	TR24	SN25	NM77	SU16	SD63	TL34
SP15	NS79	TQ36	TA08	SE79	ST80	TQ12	TL36	NG60	NB35	SK85	SN87	SD51	HU54	HU27	SN43	SJ21	SN85	SE55	NS59

SO22	SP83	NT46	SN33	SD70	SS19	SO45	HU33	TQ82	NG50	NS51	TL42	SJ03	TF34	NY59	ST92	NO34	SO76	SY49	NS78
HU24	SP52	SH86	SO18	TL47	NG45	SK17	NF67	NF84	NX07	SN45	NC12	SW32	NZ11	TL61	NB55	TQ60	SW62	ST33	NX26
SN34	SP18	HU59	SJ65	NY98	SP56	SP81	SJ54	SU64	SS73	SN46	NG55	SJ98	NC52	NS30	NT93	SO99	SU05	NY74	TL07
NF66	NJ16	SK40	NM72	NO11	HU56	NC27	NT51	ST81	SY69	NC73	SD08	NM79	NX36	SP38	SP07	NM78	ST43	SH45	NC34
SK30	TM26	SJ27	NF70	SK01	HU37	SW53	SO33	NM91	NB44	SU48	TF41	NH08	SE03	SM72	NX88	HZ27	NN11	SP04	SY19
NX13	NJ34	SO03	TR14	ND36	NY89	TQ83	SK82	SH91	NO08	NJ22	NM90	SP43	SP26	SU38	NY10	NG51	NT72	SN71	SK23
SU89	SP69	TA20	SS74	SX69	TM00	HU39	NC20	NT41	SO32	TA31	SW72	NO57	SE93	NJ70	TG12	TM18	SJ64	NY73	NJ56
SJ15	SV80	SK98	TF47	SX35	NG71	NU11	TM09	SN55	SE11	SP11	SX96	TQ95	NY61	NO75	NH86	NZ22	SN93	HY55	NO51
SO23	NS74	TL55	SE67	NO23	NJ27	SX56	ST77	SP96	SN32	SE99	SP73	NM31	ST83	NB46	NN56	TM17	NH83	ST89	SX78
SK42	SK70	NM22	NH88	TA18	SX46	NM95	NY71	SE78	SD94	TR02	NL69	SN44	SO10	NT80	TM36	SE32	TF48	NJ45	TF17
SE81	HU28	SH66	TQ52	SP87	TL40	NN66	NN52	SS97	NF95	NN57	SZ57	SM80	SN56	SU39	ST78	SS69	NO33	NH21	NJ90
NF82	NH58	HP60	NM60	SE66	NN23	ST70	TQ30	SW86	SJ13	SX83	ST16	NO67	TQ51	TA00	SP16	NB21	NH47	ST06	ST65
TA06	NT21	SD44	NY67	NZ30	NM89	SM82	TL49	SN04	NN67	NG14	TF27	SN21	SJ77	HU38	SJ34	SH80	SN24	HU41	NN87
NU10	SK81	NG76	ND38	SY39	TL35	ST99	SN91	NC92	NM68	NC42	NN62	NZ17	SP49	NJ86	SP57	TF43	SE04	HU16	SJ56
SU06	TF04	HT94	SO64	SS93	TL51	NS14	TQ28	NU12	SU49	SP82	NJ23	SO81	TM08	SX89	SJ12	TG51	HU48	SP02	ST91
NY78	NC94	SX74	NW95	NJ12	NL58	NG19	NY44	NN49	SO56	NM88	NB23	SC47	TL95	NY84	NG95	SX68	SW83	NT87	SS91
SX86	SS09	SH34	TF36	TF05	SE96	NB00	SU04	SJ48	ST75	NY18	NO39	SK16	SK73	NN35	ST00	SP98	NN15	SH23	SU69
SK97	TL33	SE48	HU53	SK89	SN97	NH04	NN22	TF57	HY60	ND28	NS80	NM40	SX15	SP24	SE56	ST71	NN73	ND17	ST97
NX48	TF51	SO06	SE70	SO85	ST18	TM37	SS92	ST85	NN02	NY49	NA92	NB12	ST34	NU05	NG98	NB22	SD20	SS89	SW73
NO88	NN16	NY64	NS26	HU47	NO60	ST09	NJ30	SR89	NO63	NN44	NG43	NO17	NC15	SE23	NH01	SJ36	NO29	NO44	SP61

HY61 NS24 NX08 NJ11 SP00 SN60 NX19 SE68 NT01 NO87 SO54 SX54 TQ31 NJ61 NZ80 TF19 ST01 TF90 SX06 TF07
TF42 SJ26 HP40 TA10 NG34 SP39 NN77 SS71 ND19 SS41 NS31 NM24 TQ37 SK93 NF85 NY69 SM81 NB01 SK72 SP62
SE77 SS90 NK06 SE38 ND47 SX85 ST30 NN27 HY54 NX17 SS87 SK77 SD33 SH13 TG20 NJ42 SP72 TQ21 NN32 NC61
NS52 SJ25 NN58 TA07 NZ62 SJ88 SK52 SW75 SP21 SW93 NB34 NO07 SN80 SP71 NG80 SP63 SS98 SN22 SX65 TQ20
SY29 NG15 SW84 SE59 SX37 SP67 SJ00 NB41 SH82 TA23 NO22 TM06 SW64 NT09 SE84 SS72 SE12 NX33 SS51 SP78
NY39 SN90 NS60 SP23 ST68 NX90 NS70 SW43 SJ44 NT11 SD53 HU57 SS40 NG54 SX95 TQ69 NO32 NN12 NS54 SS31
SO66 NO56 SS82 TL62 SP66 SO16 ST22 TM27 SP33 NG46 TF35 TL54 SX63 NB31 SU25 NT81 TA03 SX04 SX36 NY72
NY77 SX18 NT50 NH17 SS70 SO25 NH57 SJ35 NB90 ST82 NJ96 NT91 SX27 NG47 SX08 TL43 TF45 TF12 NS34 SO80
NF81 NH27 NH29 NN21 SS61 TL17 SJ16 SN15 ST42 NW97 SX67 NF60 SO00 NY20 SP27 NB14 SK10 NZ45 TF37 NG84
ST88 ST12 NH36 NS42 TF28 NN37 SS81 SS83 NA91 SS32 NN39 SD52 NY09 NJ20 HU14 NY83 SX79 ST86 SP08 SX45
HY10 SE76 NN47 HU26 SJ38 NJ10 NJ80 SP79 NZ26 SX44 NA74 SX47 SP60 NO18 SS96 NZ50 NZ33 SK76 NX09 NN36
TL64 NL93 NC41 NG75 SW42 NJ95 SN88 TF56 NS53 NR67 ST67 SD43 NZ34 ST07 SK99 NM87 SH54 SX16 SX48 ST19
SZ47 ST10 NH28 TA01 SN01 NF88 NJ51 SH32 NC16 SK31 NX14 TF40 TF38 NB42 TL63 SE13 SS58 NF09 NT40 SK62
SS50 NY58 NX39 NO89 SX66 SP03 SH98 SH88 NN61 TR23 NY04 NN60 NS25 SO07 TQ88 NF83 SJ47 SX76 SS60 NY46
NJ44 ST37 NJ00 TF18 ST63 SY66 ST73 NS87 NZ14 ST25 SX25 NK04 SP41 TL85 SP31 NA00 TA12 ST21 NG23 ST72
NZ70 TQ18 NY28 NO48 NH02 NH11 SX55 NL57 SC17 NC93 SD80 SD16 NJ55 SN11 NS16 SU68 NS20 SW71 TF32 NY19
NY68 SE94 NH71 NJ93 NJ43 NJ75 NN97 NM99 SW61 NO28 SP59 SE52 NN08 NO61 SX28 SX19 NN13 SD40 NB02 SX39
ST64 TL38 SS52 NM57 SW95 NJ17 NM41 NN42 ST27 NB10 SN00 SS62 ST40 TA32 NB30 NO43 NH00 NO38 NO79 NZ18
NZ90 NM61 SS14 NG90 NF19 NJ84 NS44 NX49 NO52 SM92 NJ71 ND59 NG79 SW74 SY77 NR50 NH37 NG40 NN03 NC43
ST31 NZ32 NA10 SX57 NH24 TM19 NB33 NC33 NG56 SX58 SH29 NJ73 NY29 NZ81 NT69 NX03 TF29 NX28 SK92 NM92

TF24 NM47 NY07 SN59 SW85 TG24 ST41 NC32 TL73 NJ54 TM16 NN78 TA40 HZ17 SH78 SZ99 NG65 SP20 NJ76 SM94
SJ89 NJ60 SS30 SS20 NZ53 NO78 SY87 HU67 SN36 SY48 NO09 NR59 TA24 SW98 NY79 NH13 ST52 NZ46 TF58 NS43
HU40 TR07 NJ72 SN13 NZ44 NZ60 SY18 SX49 NC62 NH60 SS63 HU30 SO17 SW63 NY27 ST02 NN26 SW54 NJ64 SJ10
NG38 SN70 NH16 NA93 HU55 NB20 ST53 SX75 NR51 NG49 SM91 SN23 NJ62 NH07 SX07 SH51 NR16 HY73 NB11 NN99
NS88 NG29 TF46 TF13 SD21 HW83 SP14 SW97 HU66 NJ74 SY28 TF03 SW33 SN10 TA09 TA41 NG37 NK15 SC36 SX38
NH14 NG64 NC72 NG31 NK13 NX54 NJ94 TF08 TF31 SW96 NJ21 HX62 NF56 ST62 NH50 NG86 NO58 ST17 TQ80 SM90
NM33 SW76 NH38 TF23 NN14 ND33 NJ53 SM84 SX26 SS75 SH24 NJ52 HZ26 NN98 NB52 NS21 SX29 NO74 SW52 NX24
NM21 SS68 SW94 SK95 NZ72 NX93 SM50 NH03 NH61 NN88 NO86 NJ85 NF68 NG66 SJ29 NU14 NS50 HZ16 SE41 NJ82
SN02 NG18 NJ63 TV49 NN05 NA90 NG63 HW63 NS10 SS99 TV69 NX23 TA33 TR33 SS38 NA64 NM29 NO77 SY07 HY34
ST26 NG39 NF89 NS01 HP62 SE22 NF80 SX94 NZ12 HU77 NM59 NL68 NH97 NG07 NM23 SZ28 NM51 NM37 NM38 NG13
NM19 NZ91 SM62 NN25 TR12 HX51 SV90 SX09 HY23 SW82 SM73 NM16 NZ38 SD36 SW44 SM71 NB40 SS39 NT49 SS10
NZ39 NL79 NM05 HY35 NF61 NA81 SY38 SS11 SW65 TR46

A.2 Nestedness order of species in Britain

* Codes are listed in nested order, from the most abundant species to the rarest. Read by row, left to right.

* For reference of Latin and common names to the two-letter code, see http://www.bto.org/about-birds/birdfacts/british-list

WR S. PW B. R. SG SL ST CH WW MA MP HS WP D. BT K. GT C. HM
CK M. JD SF GR LI CT GC L. H. RO PH SI Y. GO WH CD BF FP SH
CC RB TC MG LT BC MH GL TO GS CU SD BH GW SW W. J. OC SN HG
BZ LR P. G. CO TP MS LB DI SM CS WC NH TS RN TU LW RK CM RT
LO WO MT KF RL WT CG PE GB GH SK SC LG YW BO RG WK CB TD RP
T. CA GP HC F. SU RC Q. PF GG LS RW TW SA ML GD CN RM RZ CR
DN SE HY TY GJ E. EA KI HH AE N. BK RH GU RA PO LE LP GK SV
PU WA PM HB HF WN BV RY NJ GI AC GA GX NX TE MN CE DO OP RE
AF TM KT WM CF GN FF BX SB FC MX MR CP CW CY EG GY WL CX TT
CL RI TN CI ND PT OL TL WS WE DW XB GF BY RS BW DC MO KR BN
BR HZ BL AK MU SZ NK MW RU PG ED SP AV NS VI HP OD RQ SQ SO
SJ BA DR RX LM BG LN WY LU PS RF SS NG GV WX BJ SR IG GE JS
TK GZ IC UG AN AA IN TF LV BI MY KN PD HD WG NB UD LX RV EO
VS BU DT EM HN KE FW FV ET AS OQ SY WJ

A.3 Nestedness order of 10 km sub-squares in Taiwan

* Codes are listed in nested order, from the species-richest site to the depauperate. Read by row, left to right.

* For complete site reference and map, see https://sites.google.com/a/birds-tesri.twbbs.org/bbs-taiwan/method-1/yang-qu-de-tu

A03-01	A18-02	B16-01	B30-01	B32-11	B13-02	C37-04	A19-05	B16-02	B32-01	A09-33	B14-04	B13-03	C14-04	B32-10
A27-04	B30-04	B21-01	B11-01	B13-01	A27-19	B37-02	A19-06	C30-03	B14-01	C37-05	C14-01	C37-01	A19-02	B10-01
C16-01	A27-28	B10-02	A18-03	B32-04	A32-04	C37-02	A16-05	A33-14	B14-02	A34-07	A09-08	B38-04	B32-03	B06-07
C30-02	B29-02	B14-03	A33-26	C30-04	A35-06	A16-04	A27-20	C14-02	A33-15	A20-01	C14-03	B28-01	C30-01	A04-16
A21-01	B37-01	A09-31	A40-11	A28-06	A17-04	A19-04	B10-03	A01-12	A34-24	A40-12	A29-03	A09-29	A28-05	A36-05
C28-01	B30-02	A32-02	A33-04	B38-03	A26-02	A09-04	A18-04	B38-07	A12-02	A29-06	B30-07	B35-01	A28-08	A34-02
A19-03	A07-02	A35-03	C37-03	A35-04	A40-03	B06-01	A27-40	A29-01	A40-16	B22-01	B32-02	A34-03	A39-02	A34-05
B10-14	A34-23	A20-03	A35-05	B30-03	A34-06	A09-13	A34-08	A20-05	A04-48	A26-03	A36-01	A40-15	A32-06	A33-09
A39-01	A18-01	A34-49	A17-02	A24-02	B28-06	A04-31	A01-07	A37-05	A34-38	A09-50	A33-07	B38-01	A33-16	A20-02
A09-46	A40-02	B28-04	A09-24	A36-15	A22-01	A41-01	B06-06	A37-01	A26-01	A33-08	A37-02	A29-16	A17-10	A20-04
A33-02	A37-11	B10-13	A19-01	A41-02	A40-14	A17-01	A27-02	A04-45	A05-02	A35-15	A09-57	A03-20	B13-04	A28-02
A04-50	A09-15	A12-01	A04-02	A29-04	A05-01	A09-56	A36-02	A40-01	A27-41	A09-20	A27-15	A28-01	A33-01	A21-02
A34-48	A37-10	A09-25	A04-28	A34-21	A10-01	A04-44	A27-07	A32-07	A09-52	A36-03	A39-08	A33-11	A36-14	A29-02

A02-01 A27-03 A29-12 A24-01 A28-15 B38-02 A29-05 A29-20 A27-12 A37-03 A32-03 A37-04 A05-15 A09-05 A33-06
A09-01 A35-02 A35-19 A16-01 A16-03 B33-01 A27-27 A33-27 A04-26 A16-02 A38-01 A04-30 A28-10 A33-10 A32-01
A10-03 A04-34 A27-31 A36-16 A33-13 A35-16 A04-33 A09-14 A05-17 A27-32 A34-04 A36-04 A29-17 A09-30 A27-25
A32-08 A04-06 A10-02 A02-03 A17-09 B39-01 B33-02 A09-19 A09-55 A37-09 A34-32 A34-01 A28-04 A01-02 A03-16
A09-45 A04-11 A09-03 A33-22 A28-16 A34-35 A33-18 A27-29 A27-33 A04-41 A34-43 A33-28 A04-49 A02-02 A09-36
A35-08 A35-01 A29-08 A33-03 B32-05 A01-05 A27-22 A10-10 A27-01 A04-03 A28-09 A17-03 A28-03 A29-18 A09-48
A40-10 A27-17 A04-01 A05-04 A34-14 A02-05 A09-26 A35-13 A09-40 A04-12 A34-45 A17-11 A27-39 A33-05 A39-03
A27-05 A19-13 A17-20 A37-08 A27-10 A34-20 A09-06 A01-01 A40-09 A04-43 A27-30 A09-02 A03-07 A03-19 A34-17
A04-27 A34-40 A04-23 A27-06 A09-17 A03-18 A04-32 A29-23 A04-18 A33-21 A29-07 A28-07 A35-12 A40-04 A35-09
A03-03 B14-09 A09-51 A12-06 A04-51 A04-09 A34-39 A27-43 A09-27 A09-54 A09-18 A26-04 A27-24 A04-20 A10-09
A25-01 A34-44 A34-42 A04-19 A09-22 A05-05 A04-24 A22-02 A04-04 A34-18 A35-22 A10-07 A29-19 A09-35 A09-38
A27-21 A09-39 A09-58 A37-07 A35-07 A34-26 A35-18 A04-35 A09-41 A34-47 A01-08 A04-10 A04-05 A34-41 A39-07
A05-06 A27-38 A33-12 A34-22 A09-32 A04-46 A09-53 A04-25 A04-22 A34-27 A05-14 A04-36 A01-03 A09-37 A29-21
A34-31 A35-17 A36-10 A09-43 A28-14 A28-13 A04-21 A09-16 A28-12 A04-17 A07-01 A27-18 A35-10 A09-44 A09-49
A33-19 A12-10 A34-33 A33-29 A34-34

A.4 Nestedness order of species in Taiwan

* Codes are listed in nested order, from the most abundant species to the rarest. Read by row, left to right.

* For reference of Latin and common names to the code, see

https://docs.google.com/viewer?a=v&pid=sites&srcid=YmlyZHMtdGVzcmkudHdiYnMub3JnfGJicy10YWl3YW58Z3g6NWQyZDAxN2M2Njg1NTc0NA

C406	C468	C372	C408	C331	C277	C621	C471	C366	C361	C470	C045	C387	C276	C388
C474	C460	C097	C549	C472	C459	C473	C355	C100	C625	C333	C119	C358	C319	C274
C571	C551	C364	C371	C390	C271	C344	C570	C404	C505	C154	C104	C107	C320	C624
HODI	C293	C541	C299	C476	C128	C146	C283	C456	C373	MOTS	C280	C481	C379	C415
C044	C428	C360	C464	C131	C542	C384	C173	C509	C385	C178	C356	C094	C457	C466
C430	C521	C093	C526	C342	C405	C524	C602	C091	C548	C458	C482	C053	C496	C418
C508	C397	C187	C225	C491	C561	C552	C562	C175	C048	C400	C490	C626	C479	C089
DOAN	C394	C226	C022	C183	C535	C318	C314	C272	C116	C369	C392	C398	C480	C149
C414	C359	C577	C306	C134	C170	C419	C161	C499	C401	C115	C179	C109	C288	C165
C547	C607	C455	C555	C046	C556	C448	C413	C395	C103	C478	C483	C489	DOCA	C487
C124	DOAP	C101	C512	C114	C335	C347	C334	C611	C142	C402	C608	C623	C412	C463
C475	C087	C302	HYBU	C573	C507	C486	C301	C579	C511	C217	C098	C374	C544	C518
C620	C028	C403	C477	C180	B004	C340	C336	C258	C622	C297	C465	C383	C082	C139

ACGA	C550	C121	C130	C208	C613	C434	C132	C129	C219	C565	C263	C581	C096	C416
C186	C595	C294	DOCY	C349	C539	C284	B013	C558	C427	C367	C386	C308	C606	C312
C291	C162	C329	C200	B008	DOGO	B011	C569	C171	BYBU	C598	C143	C560	C155	B012
C120	STNI	C594	C204	C409	C423	C510	HYGA	C019	C316	C203	C266	B009	C519	C578
C176	B020	C617	C133	C024	C322	C262	C168	C016	C591	C559	C047	ZODA	C615	C188
C250	B018	C527	C092	C202	C034	C193	C026	C177	C113	C246	C025	C206		

*HYGA: *Garrulax canorus* X *Garrulax taewanus*; ZODA: *Zoothera dauma dauma*; STNI: *Strix nivicola*; DOGO: *Anser domestica*; ACGA: *Acridotheres grandis*; DOCA: *Cairina moschata* (domestic); HODI: *Horornis diphone/canturians*; DOAP: *Anas platyrhynchos* var. domestica; MOTS: *Motacilla tschutschensis*; DOCY: *Cygnus atratus* (Domestic); DOAN: A*nas platyrhynchos* (Domestic); BYBU: *Bycanistes bucinator*; HYBU: *Pycnonotus sinensis* X *Pycnonotus taivanus*.

Appendix B

B.1 Minimum set for estimating species number by assemblage structure

To identify how many sub-squares were required to predict 90% and 100% of the total number of species in Britain; species accumulation was calculated with increasing number of sub-squares, using two orderliness: random and nested order.

B.1.1 Randomly-select method

From each 100 km square in Britain, we randomly selected one of 10 km sub-square and recorded the number of species in it. Continued to randomly select the second sub-square other than the previous one and added number of new species to form a new species subset. This process was repeated until all species were included. Lastly, repeated the whole procedure (i.e. species accumulate from 1 sub-square to maximum number of sub-squares in the 100 km square) 1000 times and calculated the average number of minimum sub-squares needed to include 90% or 100% species. The minimum number for including 90% or 100% species was calculated in all fifty-five 100 km square.

An example using only 10 sub-squares was demonstrated in Table A- 1. Each time we randomly selected one of the 10 sub-squares and their species composition recorded. The minimum number of sub-squares required to include the whole community (species A to G) can then be calculated. For the real data set, similar process was implemented except that all sub-squares in a 100 km square were included.

Table A- 1 Example of randomly-select method. (a) A 10-cell square containing species A to G; (b) the randomly selected sub-squares are combined to create a new species subset. This process is repeated until the whole community (species A to G) has been recorded. In this sample, four sub-squares are the minimum required for recording all seven species in this square.

(a)

Sub-square	I	II	III	IV	V	VI	VII	VIII	IX	X
Species composition	ABD	BDF	ABC	AB	ABCDE	BEF	CEFG	ACF	ABCFG	BCDEFG
Total species number	3	3	3	2	5	3	4	3	5	6

(b)

Round of select	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
New sub-square selected	VI	VIII	IX	V	VII	X	I	III	IV	II
New subset of species	BEF	BEFAC	BEFACG	BEFACGD	BEFACGD	BEFACGD	BEFACGD	BEFACGD	BEFACGD	BEFACGD
Total species number	3	5	6	7	7	7	7	7	7	7

B.1.2 Nestedness-select method

For this method, we selected sub-square based on the nested structure of the 100 km square, which was rearranged by local unexpectedness in descending order according to its nestedness. The procedure was similar to the randomly-select method except that the sub-squares were chosen according to the nested order, starting from the most species-rich site until at least 90% or 100% species were included.

B.2 Minimum sub-squares required for estimating species richness: result

A mean of 29.2±5.7% (±SD) of 10 km sub-squares was required for the random-selection method to include 90% of all species within the 100 km square (ranging from 16.3±3.3% to 40.2±7.6% for each square). Including all species requires 74.3±7.9 to 93.7±6.0% of sub-squares (with a mean of 86.2±6.9%). Comparatively, only 4.0%-28.8% (with mean = 14.8±7.1%) of all sub-squares were required to incorporate 90% of species using nestedness-selection method, but including all species needs much more sub-squares (41.6%-100.0%, with a mean of 76.9±15.7%). The nestedness-selection method required fewer sub-squares to estimate 90% of species than random method, but had larger variation in minimum number for these sub-squares.

Appendix C

C.1 Determing the number of clusterings in indicator species value analysis

Dufrêne and Legendre (1997) is followed for the determination of the optimal number of clusterings for the indicator species value analysis. This method firstly calculates the indicator value (IndVal) of each avian species for each clustering level (e.g. IndVal for species 1 to species 273 at clustering level 2, 3, 4 etc.). The difference in IndVal between consecutive cluster levels for each species is then calculated (e.g. the difference in IndVal between cluster 2 and cluster 3 for species a) and plotted. Because larger IndVal represents a better predictor, the optimal clustering level is when the difference in IndVal (cluster n+1 minus cluster n) is the largest and is most positive. In this case (Figure A-1), because the value (difference in IndVal, y axis) is the highest between cluster 2 and cluster 3, and is of similar high value between cluster 4 and cluster 5 (both also have smaller negative value unlike between cluster 7 and cluster 8), it is most optimal to divide the avian assembly into three or five clusters. However, either three or five clusters has lower predictability than the method based on network modularity, as demonstrated briefly in Figure A-2 (below, for three clusters) and comprehensively in Chapter 3 (for five clusters).

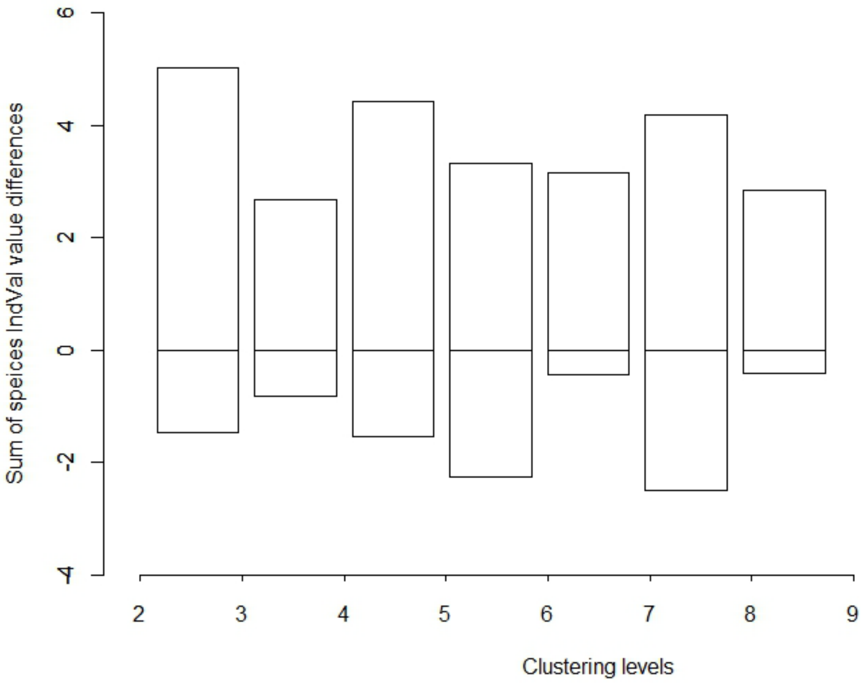
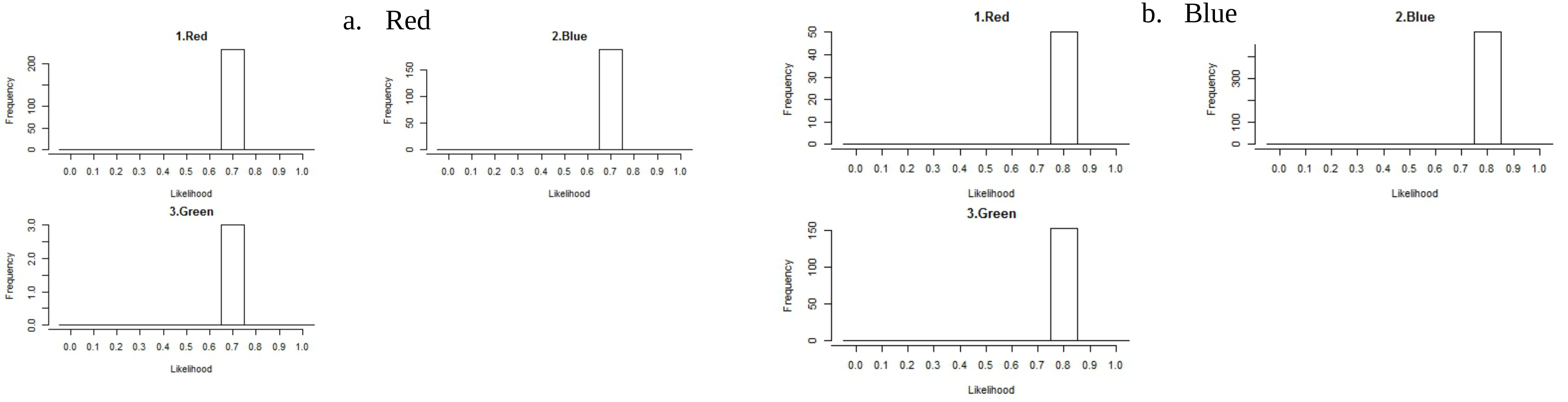


Figure A- 1 Sum of species indicator value (IndVal) differences consecutive cluster levels for each species in the BTO dataset.

a. Red

b. Blue



c. Green

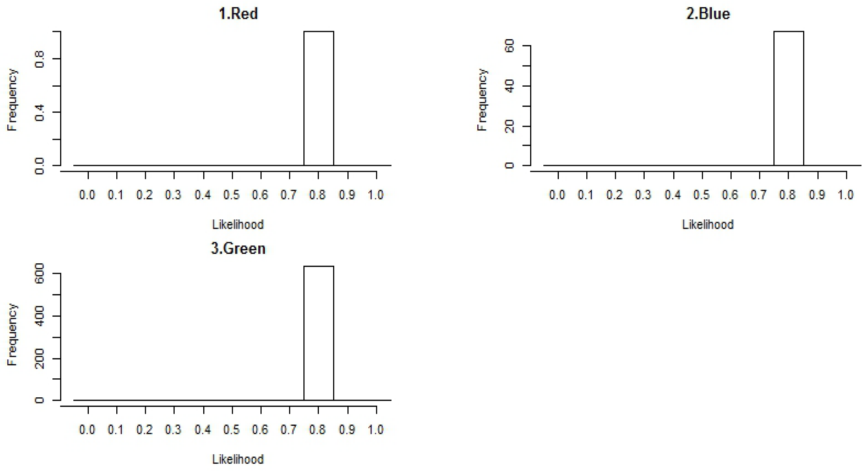


Figure A- 2 The predictive ability of the site group to the unsampled sites when predicted based on 40% sampling and under each of three sampling clustering indicator species group. The title (a-c) indicates each of the three assigned population clustering and the x-axis is the likelihood of assigning an unsampled site to the target site group, comparing with the original site memberships (i.e. the site group assigned by the population clustering). For example, figure (a) is the indicators derived from the red group and figure (a.1) is the likelihood to assign an unsampled site, which is originally assigned to the red group in population clustering, to the red site group by the red indicators.

Appendix D

D.1 Sensitivity analysis for sampling size of site group detecting

A preliminary test has done for deciding the sampling size to effectively represent popular modularity, before the site group detecting process. 10%, 20%, and 40% random samplings from the whole BTO dataset were partitioned based on network modularity randomtrap method. The three sampling modularity (10%, 20%, and 40%) was compared with the corresponding sampling sites in population modularity using the *Fowlkes-Mallows index* (*FM index*) (Fowlkes and Mallows 1983).

D.2 Sensitivity analyses for sampling size of site group detecting: result

The result of sensitivity analysis showed the 10% and 20% sampling both had four modules whereas the 40% sampling had five. The green module that represents a small coastal site group was hardly detected in the 10% and 20% sampling for its sparse and scatter distribution. Table A- 2 showed the FM index of 10% and 20% sampling were both lower than 40% sampling. The module distribution was illustrated in Figure A- 1. There was no green module in either 10% or 20% sampling modularity result.

Table A- 2 The Fowlkes-Mallows index of 10%, 20%, and 40% sampling modularity when comparing the similarity to population modularity. FM= the Fowlkes-Mallows index; E= expected value of the Fowlkes-Mallows index; V= the variance under the assumption of no relation between the sampling and population modularity.

Sampling size	***FM***	***E***	***V***
10%	0.71	0.43	<0.01
20%	0.73	0.33	<0.01
40%	0.83	0.34	<0.01

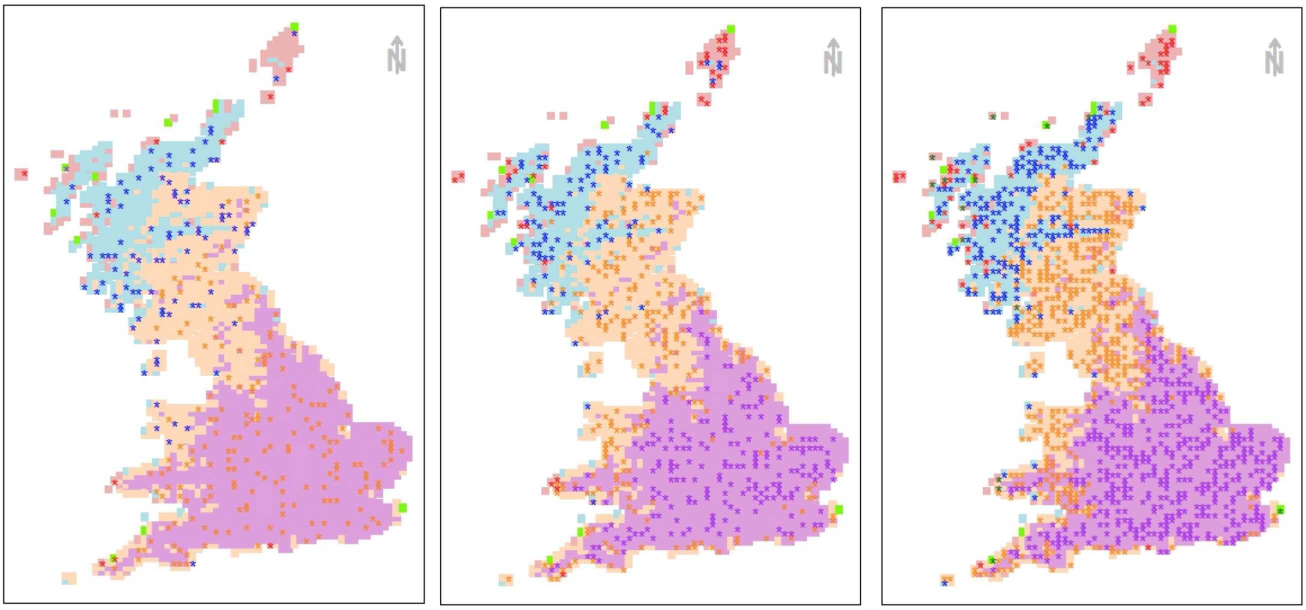


Figure A- 3 Assigned module of sites based on 10%, 20%, and 40% (left to right) sampling (colour asterisk) and whole sampling (background colour) of BTO dataset with randomtrap algorithm. Different colours represented different modules. Each square is a 10 km sub-square.

Appendix E

E.1 Species composition in determined network modules and k-means clusters in Britain.

Table A- 3 Species list for the four indicator groups selected by modularity method (left) and k-means clusters (right).

*Species names represented by two-letter code (see http://www.bto.org/about-birds/birdfacts/british-list)

** Module and cluster colours are assigned arbitrary.

Modules	Species
Red	AA AC AE AF B. BC BH BV BW BZ C. CA CB CD CE CF CH CK CL CM CN CR CS CU CX D. DI DN DO E. EA F. FP GB GJ GK GL GP GU GX H. HB HC HG HH HM HS JD K. KI L. LB LG LI LR LW MA MH ML MP MS MX ND NK NX OC OD PE PH PM PU PW Q. R. RA RB RC RE RG RH RK RM RN RO RP RS RZ S. SA SB SC SD SE SF SG SI SL SN SO SQ ST SU SV SW T. TE TL TM TT TU TW TY W. WC WE WH WM WN WP WR WS WW Y.

Clusters	Species
Red	AE AF AK AN AS AV B. BA BC BF BG BH BI BJ BK BL BN BO BR BT BU BW BX BY BZ C. CA CB CC CD CE CF CG CH CI CK CL CM CN CO CP CR CS CT CU CW CX CY D. DC DI DN DR DT DW E. ED EG EM ET F. FC FF FP FV FW G. GA GB GC GD GE GF GG GH GI GJ GL GN GO GP GR GS GT GU GV GW GX GY H. HB HC HD HF HG HH HM HP HS HY HZ J. JD K. KF KI KN KR L. LB LE LG LI LM LN LO LP LR LS LT LU LW LX M. MA MG MH ML MN MO MP MR MS MT MU MW MY N. NB NG NH NJ NS OC OL OP OQ P. PE PF PG PH PO PT PU PW Q. R. RA RB RC RE RG RI RK RL RM RN RO RP RQ RS RT RU RV RW RX RY RZ S. SA SC SD SE SF SG SH SI SJ SK SL SM SN SP SS ST SU SV SW SZ T. TC TD TE TN TO TP TS TT TU TW TY UD UG VI VS W. WA WC WG WH WK WL WN WO WP WR WS WT WW XB Y. YW

Blue	AC AE AF AK B. BA BC BF BG BH BJ BK BL BO BT BV BW BY BZ C. CA CB CC CD CE CF CG CH CI CK CM CN CO CP CR CS CT CU CX CY D. DI DN DO DR E. EA ED EO ET F. FF FP G. GA GB GC GD GF GG GH GI GJ GK GL GN GO GP GR GS GT GU GV GW GX GY GZ H. HB HC HD HF HG HH HM HS IC IG IN J. JD JS K. KF KI KT L. LB LE LG LI LN LR LT LW M. MA MG MH ML MN MP MR MS MX ND NJ NK NX OC OD OL OP P. PE PF PH PM PO PS PT PU PW Q. R. RA RB RC RE RF RG RH RK RL RM RN RO RP RS RT RU RZ S. SA SB SC SD SE SF SG SH SI SJ SK SL SM SN SO SP SQ SS ST SU SV SW SY SZ T. TC TD TE TK TL TM TO TP TS TT TU TW TY W. WA WC WE WH WJ WK WM WN WO WP WR WS WT WW Y. YW
Green	F. GB GU GX KI PU RA RC RH SA TY
Purple	AE AF AK AN AS AV B. BA BC BF BG BH BI BJ BK BL BN BO BR BT BU BW BX BY BZ C. CA CB CC CD CE CF CG CH CK CL CM CN CO CP CR CS CT CU CW CX CY D. DC DI DN DO DR DT DW E. ED EG EM ET F. FC FF FP FV FW G. GA GB GC GD GE GF GG GH GI GJ GL GN GO GP GR GS GT GU GV GW GX GY H. HB HC HD HF HG HH HM HN HP HS HY HZ J. JD JS K. KF KI KN KR KT L. LB LE LG LI LM LO LP LR LS LT LU LV LW LX M. MA MG MH ML MN MO MP MR MS MT MU MW MY N. NB NG NH NJ NS OC OL OP OQ P. PE PF PG PH PO PT PW Q. R. RB RC RE RG RI RK RL RM RN RO RP RQ RS RT RU RV RW RX RY RZ S. SA SC SD SE SF SG SH SI SJ SK SL SM SN SR SS ST SU SV SW T. TC TD TE TN TO TP TS TT TU TW UD UG VI VS W. WA WC WG WH WK WL WN WO WP WR WS WT WW WX Y. YW

Blue	AA AC AE AF AK B. BA BC BF BG BH BK BL BO BT BV BW BY BZ C. CA CB CC CD CE CF CG CH CI CK CL CM CN CO CP CR CS CT CU CX CY D. DI DN DO DR E. EA ED F. FF FP GA GB GC GD GG GH GI GJ GK GL GN GO GP GR GS GT GU GV GW GX GY H. HB HC HG HH HM HS IC IG J. JD K. KI L. LB LE LG LI LN LR LT LW M. MA MG MH ML MP MR MS MX ND NK NX OC OD OL OP P. PE PF PH PM PO PS PT PU PW Q. R. RA RB RC RE RF RG RH RK RL RM RN RO RP RS RT RU RW RZ S. SA SB SC SD SE SF SG SH SI SJ SK SL SM SN SO SP SQ SS ST SU SV SW SZ T. TC TD TE TL TM TO TP TS TT TU TW TY W. WA WC WE WH WK WM WN WO WP WR WS WW Y.
Green	AF AK AN B. BC BF BH BK BL BN BO BR BT BW BX BY BZ C. CA CB CC CD CE CF CG CH CK CL CM CN CO CR CS CT CU CW CX D. DC DI DN DO DR DW E. ED EG F. FC FF FP G. GA GB GC GD GE GF GG GH GI GJ GL GN GO GP GR GS GT GU GW GX GY H. HB HC HF HG HH HM HN HP HS HY HZ J. JD JS K. KF KI KR KT L. LB LE LG LI LM LO LP LR LS LT LV LW M. MA MG MH ML MN MO MP MR MS MT MU MW MX N. NG NH NJ NS OC OL OP P. PE PF PG PH PO PT PU PW Q. R. RA RB RC RE RG RI RK RL RM RN RO RP RT RW RY RZ S. SA SC SD SE SF SG SH SI SK SL SM SN SR ST SU SV SW T. TC TD TN TO TP TS TT TU TW TY W. WA WC WH WK WL WN WO WP WR WT WW WX Y. YW
Purple	AC AE AF AK B. BC BF BH BJ BK BL BN BO BT BV BW BX BY BZ C. CA CB CC CD CE CF CG CH CI CK CM CN CO CP CR CS CT CU CX CY D. DI DN DO E. EA ED EG EO ET F. FC FF FP G. GA GB GC GD GF GG GH GI GJ GK GL GN GO GP GR GS GT GU GW GX GY GZ H. HB HC HD HF HG HH HM HP HS HY HZ IN J. JD JS K. KF KI KR KT L. LB LE LG LI LN LO LP LR LS LT LU LW M. MA MG MH ML MN MP MR MS MT MX ND NG NH NJ NK NX OC OD OL OP P. PD PE PF PG PH PM PO PS PT PU PW Q. R. RA RB RC RE RG RH RI RK RL RM RN RO RP RS RT RU RW RX RY RZ S. SA SB SC SD SE SF SG SH SI SJ SK SL SM SN SO SP SQ SR ST SU SV SW SY SZ T. TC TD TE TK TM TO TP TS TT TU TW TY W. WA WC WE WG WH WJ WK WL WM WN WO WP WR WS WT WW WY XB Y. YW

Orange	AC AE AF AK AV B. BA BC BF BG BH BK BL BN BO BR BT BV BW BX BY BZ C. CA CB CC CD CE CF CG CH CI CK CL CM CN CO CP CR CS CT CU CW CX CY D. DI DN DO DR DW E. EA ED EG F. FC FF FP G. GA GB GC GD GF GG GH GI GJ GK GL GN GO GP GR GS GT GU GV GW GX GY GZ H. HB HC HD HF HG HH HM HP HS HY HZ IN J. JD K. KE KF KI KN KR KT L. LB LE LG LI LM LN LO LP LR LS LT LU LW LX M. MA MG MH ML MN MO MP MR MS MT MU MW MX N. NB ND NG NH NJ NS OC OD OL OP P. PD PE PF PG PH PM PO PS PT PU PW Q. R. RA RB RC RE RG RH RI RK RL RM RN RO RP RS RT RU RW RX RY RZ S. SA SB SC SD SE SF SG SH SI SJ SK SL SM SN SO SP SQ SR SS ST SU SV SW SZ T. TC TD TE TF TK TL TM TO TP TS TT TU TW TY UD VI VS W. WA WC WE WG WH WK WL WM WN WO WP WR WS WT WW WY XB Y. YW

Orange	AC AE AF AK AV B. BA BC BF BG BH BK BL BN BO BR BT BW BX BY BZ C. CA CB CC CD CE CF CG CH CI CK CL CM CN CO CP CR CS CT CU CW CX CY D. DI DN DO DR DW E. EA ED EG F. FC FF FP G. GA GB GC GD GF GG GH GI GJ GL GN GO GP GR GS GT GU GV GW GX GY GZ H. HB HC HF HG HH HM HP HS HY IN J. JD K. KE KF KI KN KR KT L. LB LE LG LI LN LO LP LR LS LT LU LW LX M. MA MG MH ML MN MO MP MR MS MT MU MX N. NB ND NH NJ NS NX OC OP P. PE PF PG PH PO PT PU PW Q. R. RA RB RC RE RG RH RI RK RL RM RN RO RP RQ RS RT RU RW RY RZ S. SA SC SD SE SF SG SH SI SK SL SM SN SP SQ SS ST SU SV SW T. TC TD TE TF TL TM TO TP TS TT TU TW TY VI W. WA WC WH WK WN WO WP WR WS WT WW XB Y. YW

Appendix F

F.1 Species composition in determined network modules and k-means clusters in Taiwan.

Table A- 4 Species list for the four indicator groups selected by modularity method in Taiwan

Site group	Species
Purple	*Dendrocopos leucotos, Brachypteryx montana, Regulus goodfellowi, Cuculus optatus, Carpodacus formosanus, Trochalopteron morrisonianum, Locustella alishanensis, Corvus macrorhynchos, Heterophasia auricularis, Cinclidium leucurum, Tarsiger indicus, Pyrrhula erythaca, Prunella collaris, Delichon dasypus, Parus monticolus, Yuhina brunneiceps, Nucifraga caryocatactes, Muscicapa ferruginea, Tarsiger johnstoniae, Actinodura morrisoniana, Sitta europaea, Horornis acanthizoides, Suthora verreauxi, Phylloscopus inornatus, Liocichla steerii, Periparus ater, Fulvetta formosana, Troglodytes troglodytes*
Red	*Apus pacificus, Dendrocopos leucotos, Dicrurus macrocercus, Spilornis cheela, Megapomatorhinus erythrocnemis, Cuculus poliocephalus, Dicrurus aeneus, Zoothera dauma dauma, Apus nipalensis, Enicurus scouleri, Dendrocopos canicapillus, Brachypteryx montana, Horornis fortipes, Pomatorhinus musicus, Cyanoderma ruficeps, Megalaima nuchalis, Accipiter gularis, Regulus goodfellowi, Cuculus optatus, Arborophila crudigularis, Ianthocincla ruficeps, Carpodacus formosanus, Myophonus insularis, Trochalopteron morrisonianum, Locustella alishanensis, Pnoepyga formosana, Corvus macrorhynchos, Heterophasia auricularis, Cinclidium leucurum, Tarsiger indicus,*

	Turdus pallidus, Turdus poliocephalus, Motacilla alba, Columba pulchricollis, Pericrocotus solaris, Pyrrhula erythaca, Bambusicola thoracicus, Poecile varius, Turdus chrysolaus, Delichon dasypus, Accipiter virgatus, Garrulus glandarius, Ictinaetus malayensis, Cinclus pallasii, Fringilla montifringilla, Zoothera dauma, Streptopelia orientalis, Parus monticolus, Yuhina brunneiceps, Nucifraga caryocatactes, Muscicapa ferruginea, Dicaeum ignipectus, Ficedula parva, Hypsipetes leucocephalus, Aegithalos concinnus, Tarsiger johnstoniae, Streptopelia chinensis, Actinodura morrisoniana, Sitta europaea, Horornis acanthizoides, Abroscopus albogularis, Ianthocincla poecilorhyncha, Horornis diphone/canturians, Parus holsti, Suthora verreauxi, Ficedula hyperythra, Liocichla steerii, Spinus spinus, Niltava vivida, Otus spilocephalus, Hypothymis azurea, Syrmaticus mikado, Emberiza spodocephala, Periparus ater, Falco peregrinus, Phoenicurus fuliginosus, Picus canus, Erpornis zantholeuca, Treron sieboldii, Zosterops japonicus, Otus lettia, Accipiter trivirgatus, Fulvetta formosana, Prinia inornata, Pyrrhula nipalensis, Dendrocitta formosae, Schoeniparus brunneus, Glaucidium brodiei, Alcippe morrisonia, Lophura swinhoii, Troglodytes troglodytes, Hierococcyx sparverioides
Blue	*Pitta nympha, Acridotheres cristatellus, Apus pacificus, Ardea alba, Dendrocopos leucotos, Dicrurus macrocercus, Spilornis cheela, Megapomatorhinus erythrocnemis, Egretta garzetta, Dicrurus aeneus, Zoothera dauma dauma, Apus nipalensis, Eophona migratoria, Enicurus scouleri, Dendrocopos canicapillus, Charadrius dubius, Brachypteryx montana, Emberiza pusilla, Horornis fortipes, Pomatorhinus musicus, Tachybaptus ruficollis, Cyanoderma ruficeps, Passer rutilans, Megalaima nuchalis, Regulus goodfellowi, Cuculus optatus, Arborophila crudigularis, Ianthocincla ruficeps, Caprimulgus affinis, Garrulax taewanus, Myophonus insularis, Trochalopteron morrisonianum, Locustella alishanensis, Urocissa caerulea, Pnoepyga*

formosana, Corvus macrorhynchos, Heterophasia auricularis, Acridotheres javanicus, Cinclidium leucurum, Lonchura striata, Tringa ochropus, Copsychus malabaricus, Amaurornis phoenicurus, Turdus pallidus, Pycnonotus sinensis, Turdus poliocephalus, Spizixos semitorques, Motacilla alba, Oriolus traillii, Pericrocotus divaricatus, Riparia riparia, Strix nivicola, Columba pulchricollis, Butastur indicus, Pericrocotus solaris, Hirundapus cochinchinensis, Rallina eurizonoides, Prinia flaviventris, Motacilla cinerea, Pyrrhula erythaca, Bambusicola thoracicus, Cecropis striolata, Poecile varius, Turdus chrysolaus, Accipiter soloensis, Nycticorax nycticorax, Delichon dasypus, Pernis ptilorhynchus, Accipiter virgatus, Garrulus glandarius, Acridotheres fuscus, Ictinaetus malayensis, Cinclus pallasii, Coracina macei, Fringilla montifringilla, Zoothera dauma, Streptopelia orientalis, Parus monticolus, Phylloscopus coronatus, Yuhina brunneiceps, Nucifraga caryocatactes, Hirundo tahitica, Lanius cristatus, Muscicapa ferruginea, Dicaeum ignipectus, Streptopelia tranquebarica, Hypsipetes leucocephalus, Urocissa erythrorhyncha, Aegithalos concinnus, Acridotheres tristis, Hirundo rustica, Ixobrychus cinnamomeus, Pycnonotus taivanus, Streptopelia chinensis, Sinosuthora webbiana, Actinodura morrisoniana, Sitta europaea, Horornis acanthizoides, Calliope calliope, Columba livia, Pandion haliaetus, Passer montanus, Pica pica, Lonchura punctulata, Prinia crinigera, Riparia chinensis, Lanius schach, Abroscopus albogularis, Cisticola juncidis, Ianthocincla poecilorhyncha, Centropus bengalensis, Urosphena squameiceps, Horornis diphone/canturians, Terpsiphone atrocaudata, Parus holsti, Phoenicurus auroreus, Phylloscopus inornatus, Ficedula hyperythra, Liocichla steerii, Phylloscopus proregulus, Niltava vivida, Anthus rubescens, Otus spilocephalus, Cisticola exilis, Bubulcus ibis, Motacilla tschutschensis, Hypothymis azurea, Gorsachius melanolophus, Ianthocincla chinensis, Milvus migrans, Emberiza spodocephala, Phylloscopus borealis, Periparus ater, Cacatua galerita, Phoenicurus fuliginosus, Nisaetus nipalensis, Picus canus, Dicaeum minullum, Erpornis zantholeuca, Treron sieboldii, Butorides striata, Anas

	platyrhynchos (Domestic), Zosterops japonicus, Alcedo atthis, Chalcophaps indica, Accipiter gentilis, Ardea cinerea, Otus lettia, Accipiter trivirgatus, Muscicapa latirostris, Strix leptogrammica, Prinia inornata, Pyrrhula nipalensis, Dendrocitta formosae, Anthus hodgsoni, Glareola maldivarum, Schoeniparus brunneus, Aix galericulata, Phasianus colchicus, Actitis hypoleucos, Glaucidium brodiei, Alcippe morrisonia, Lophura swinhoii, Monticola solitarius, Pycnonotus sinensis X Pycnonotus taivanus, Buteo buteo, Copsychus saularis, Hierococcyx sparverioides
Green	*Gracula religiosa, Pitta nympha, Acridotheres cristatellus, Apus pacificus, Ardea alba, Dicrurus macrocercus, Anthus richardi, Spilornis cheela, Garrulax canorus, Garrulax canorus X Garrulax taewanus, Megapomatorhinus erythrocnemis, Anas crecca, Egretta garzetta, Cuculus poliocephalus, Dicrurus aeneus, Apus nipalensis, Tringa stagnatilis, Enicurus scouleri, Dendrocopos canicapillus, Alauda gulgula, Larus fuscus, Sternula albifrons, Charadrius dubius, Horornis fortipes, Pomatorhinus musicus, Tachybaptus ruficollis, Cyanoderma ruficeps, Passer rutilans, Mesophoyx intermedia, Gallinago megala, Calidris temminckii, Megalaima nuchalis, Recurvirostra avosetta, Xenus cinereus, Pluvialis fulva, Accipiter gularis, Horornis diphone, Hydrophasianus chirurgus, Lonchura oryzivora, Cuculus optatus, Accipiter nisus, Arborophila crudigularis, Caprimulgus affinis, Garrulax taewanus, Myophonus insularis, Trochalopteron morrisonianum, Urocissa caerulea, Cuculus micropterus, Corvus macrorhynchos, Emberiza rustica, Gallinago gallinago, Heterophasia auricularis, Acridotheres javanicus, Cinclidium leucurum, Fulica atra, Anas querquedula, Emberiza tristrami, Anthus gustavi, Chlidonias leucopterus, Euodice malabarica, Lonchura striata, Tringa ochropus, Copsychus malabaricus, Amaurornis phoenicurus, Turdus pallidus, Pycnonotus sinensis, Turdus poliocephalus, Ianthocincla sannio, Spizixos semitorques, Motacilla alba, Anas acuta, Calidris acuminata,*

Oriolus traillii, Ardeola bacchus, Pericrocotus divaricatus, Riparia riparia, Columba pulchricollis, Sturnia sinensis, Butastur indicus, Gallirallus striatus, Pericrocotus solaris, Muscicapa griseisticta, Sturnus cineraceus, Rallina eurizonoides, Sturnia malabarica, Prinia flaviventris, Motacilla cinerea, Bambusicola thoracicus, Eurystomus orientalis, Tringa totanus, Anthus cervinus, Cecropis striolata, Turdus chrysolaus, Accipiter soloensis, Halcyon coromanda, Anas penelope, Nycticorax nycticorax, Egretta sacra, Acrocephalus orientalis, Delichon dasypus, Pernis ptilorhynchus, Charadrius alexandrinus, Accipiter virgatus, Acridotheres fuscus, Anas zonorhyncha, Zoothera dauma, Streptopelia orientalis, Chloris sinica, Calidris subminuta, Tringa nebularia, Yuhina brunneiceps, Hirundo tahitica, Lanius cristatus, Gallinula chloropus, Dicaeum ignipectus, Calidris ruficollis, Falco tinnunculus, Streptopelia tranquebarica, Psittacula krameri, Hypsipetes leucocephalus, Treron formosae, Egretta eulophotes, Threskiornis aethiopicus, Acridotheres tristis, Hirundo rustica, Anser domestica, Ixobrychus cinnamomeus, Acridotheres grandis, Pycnonotus taivanus, Streptopelia chinensis, Sinosuthora webbiana, Himantopus himantopus, Rostratula benghalensis, Calliope calliope, Columba livia, Pandion haliaetus, Passer montanus, Pica pica, Lonchura punctulata, Prinia crinigera, Turdus eunomus, Turnix suscitator, Hypsipetes amaurotis, Riparia chinensis, Lanius schach, Abroscopus albogularis, Cisticola juncidis, Anas clypeata, Cairina moschata (domestic), Centropus bengalensis, Urosphena squameiceps, Horornis diphone/canturians, Terpsiphone atrocaudata, Ardea purpurea, Sturnus sericeus, Anas platyrhynchos var. domestica, Ixobrychus sinensis, Phoenicurus auroreus, Tringa brevipes, Phylloscopus inornatus, Phylloscopus proregulus, Cisticola exilis, Motacilla citreola, Bubulcus ibis, Motacilla tschutschensis, Oriolus chinensis, Cygnus atratus (Domestic), Larus crassirostris, Limosa limosa, Hypothymis azurea, Gorsachius melanolophus, Platalea minor, Elanus caeruleus, Saxicola maurus, Chlidonias hybrida, Calidris alpina, Gracupica nigricollis, Milvus migrans, Lonchura atricapilla, Emberiza spodocephala, Phylloscopus

	borealis, Acridotheres burmannicus, Cacatua galerita, Falco peregrinus, Phoenicurus fuliginosus, Dicaeum minullum, Erpornis zantholeuca, Treron sieboldii, Butorides striata, Anas platyrhynchos (Domestic), Zosterops japonicus, Porzana fusca, Alcedo atthis, Chalcophaps indica, Charadrius mongolus, Ardea cinerea, Otus lettia, Accipiter trivirgatus, Aythya fuligula, Muscicapa latirostris, Alauda arvensis, Sturnus vulgaris, Phylloscopus fuscatus, Prinia inornata, Ninox japonica, Aplonis panayensis, Emberiza rutila, Bycanistes bucinator, Eudynamys scolopaceus, Dendrocitta formosae, Anthus hodgsoni, Estrilda melpoda, Estrilda astrild, Glareola maldivarum, Sterna hirundo, Schoeniparus brunneus, Phasianus colchicus, Charadrius hiaticula, Actitis hypoleucos, Glaucidium brodiei, Alcippe morrisonia, Tarsiger cyanurus, Monticola solitarius, Pycnonotus sinensis X Pycnonotus taivanus, Copsychus saularis, Calidris ferruginea, Troglodytes troglodytes, Tringa glareola, Phalacrocorax carbo

Table A- 5 Species list for the four indicator groups selected by clustering method in Taiwan

Site group	**Species**
Purple	*Apus pacificus, Dendrocopos leucotos, Dicrurus macrocercus, Spilornis cheela, Megapomatorhinus erythrocnemis, Egretta garzetta, Cuculus poliocephalus, Dicrurus aeneus, Zoothera dauma dauma, Apus nipalensis, Enicurus scouleri, Dendrocopos canicapillus, Brachypteryx montana, Horornis fortipes, Pomatorhinus musicus, Cyanoderma ruficeps, Passer rutilans, Megalaima nuchalis, Accipiter gularis, Regulus goodfellowi, Cuculus optatus, Arborophila crudigularis, Ianthocincla ruficeps, Carpodacus formosanus, Myophonus insularis,*

	Trochalopteron morrisonianum, Locustella alishanensis, Urocissa caerulea, Pnoepyga formosana, Corvus macrorhynchos, Heterophasia auricularis, Cinclidium leucurum, Tarsiger indicus, Turdus pallidus, Pycnonotus sinensis, Turdus poliocephalus, Spizixos semitorques, Motacilla alba, Columba pulchricollis, Pericrocotus solaris, Hirundapus cochinchinensis, Prinia flaviventris, Motacilla cinerea, Pyrrhula erythaca, Bambusicola thoracicus, Cecropis striolata, Poecile varius, Turdus chrysolaus, Prunella collaris, Delichon dasypus, Accipiter virgatus, Garrulus glandarius, Ictinaetus malayensis, Cinclus pallasii, Fringilla montifringilla, Zoothera dauma, Streptopelia orientalis, Parus monticolus, Phylloscopus coronatus, Yuhina brunneiceps, Nucifraga caryocatactes, Hirundo tahitica, Muscicapa ferruginea, Dicaeum ignipectus, Ficedula parva, Hypsipetes leucocephalus, Urocissa erythrorhyncha, Aegithalos concinnus, Tarsiger johnstoniae, Streptopelia chinensis, Sinosuthora webbiana, Actinodura morrisoniana, Sitta europaea, Horornis acanthizoides, Calliope calliope, Passer montanus, Prinia crinigera, Abroscopus albogularis, Ianthocincla poecilorhyncha, Horornis diphone/canturians, Parus holsti, Suthora verreauxi, Phoenicurus auroreus, Phylloscopus inornatus, Ficedula hyperythra, Liocichla steerii, Spinus spinus, Niltava vivida, Otus spilocephalus, Hypothymis azurea, Syrmaticus mikado, Gorsachius melanolophus, Emberiza spodocephala, Phylloscopus borealis, Periparus ater, Falco peregrinus, Phoenicurus fuliginosus, Picus canus, Erpornis zantholeuca, Treron sieboldii, Zosterops japonicus, Chalcophaps indica, Otus lettia, Accipiter trivirgatus, Fulvetta formosana, Prinia inornata, Pyrrhula nipalensis, Dendrocitta formosae, Schoeniparus brunneus, Aix galericulata, Glaucidium brodiei, Alcippe morrisonia, Lophura swinhoii, Troglodytes troglodytes, Hierococcyx sparverioides
Red	*Pitta nympha, Acridotheres cristatellus, Apus pacificus, Ardea alba, Dicrurus macrocercus, Spilornis cheela, Megapomatorhinus*

erythrocnemis, Egretta garzetta, Dicrurus aeneus, Apus nipalensis, Enicurus scouleri, Dendrocopos canicapillus, Alauda gulgula, Charadrius dubius, Brachypteryx montana, Emberiza pusilla, Horornis fortipes, Pomatorhinus musicus, Tachybaptus ruficollis, Cyanoderma ruficeps, Passer rutilans, Mesophoyx intermedia, Megalaima nuchalis, Accipiter gularis, Cuculus optatus, Accipiter nisus, Arborophila crudigularis, Ianthocincla ruficeps, Caprimulgus affinis, Garrulax taewanus, Myophonus insularis, Trochalopteron morrisonianum, Locustella alishanensis, Urocissa caerulea, Corvus macrorhynchos, Emberiza rustica, Heterophasia auricularis, Acridotheres javanicus, Cinclidium leucurum, Lonchura striata, Copsychus malabaricus, Amaurornis phoenicurus, Turdus pallidus, Pycnonotus sinensis, Spizixos semitorques, Motacilla alba, Oriolus traillii, Pericrocotus divaricatus, Riparia riparia, Strix nivicola, Columba pulchricollis, Butastur indicus, Pericrocotus solaris, Hirundapus cochinchinensis, Muscicapa griseisticta, Rallina eurizonoides, Prinia flaviventris, Motacilla cinerea, Bambusicola thoracicus, Cecropis striolata, Poecile varius, Turdus chrysolaus, Accipiter soloensis, Nycticorax nycticorax, Egretta sacra, Delichon dasypus, Pernis ptilorhynchus, Accipiter virgatus, Garrulus glandarius, Acridotheres fuscus, Ictinaetus malayensis, Cinclus pallasii, Coracina macei, Fringilla montifringilla, Anas zonorhyncha, Zoothera dauma, Streptopelia orientalis, Parus monticolus, Phylloscopus coronatus, Yuhina brunneiceps, Hirundo tahitica, Lanius cristatus, Muscicapa ferruginea, Gallinula chloropus, Dicaeum ignipectus, Streptopelia tranquebarica, Hypsipetes leucocephalus, Aegithalos concinnus, Acridotheres tristis, Hirundo rustica, Pycnonotus taivanus, Streptopelia chinensis, Sinosuthora webbiana, Actinodura morrisoniana, Horornis acanthizoides, Calliope calliope, Columba livia, Pandion haliaetus, Passer montanus, Pica pica, Lonchura punctulata, Prinia crinigera, Riparia chinensis, Lanius schach, Abroscopus albogularis, Cisticola juncidis, Ianthocincla poecilorhyncha, Centropus bengalensis, Urosphena squameiceps, Horornis diphone/canturians, Terpsiphone atrocaudata, Parus holsti, Phoenicurus auroreus,

	Phylloscopus inornatus, Ficedula hyperythra, Liocichla steerii, Phylloscopus proregulus, Niltava vivida, Anthus rubescens, Otus spilocephalus, Bubulcus ibis, Motacilla tschutschensis, Oriolus chinensis, Hypothymis azurea, Gorsachius melanolophus, Ianthocincla chinensis, Milvus migrans, Emberiza spodocephala, Phylloscopus borealis, Cacatua galerita, Phoenicurus fuliginosus, Nisaetus nipalensis, Picus canus, Dicaeum minullum, Erpornis zantholeuca, Treron sieboldii, Butorides striata, Anas platyrhynchos (Domestic), Zosterops japonicus, Porzana fusca, Alcedo atthis, Chalcophaps indica, Ardea cinerea, Otus lettia, Accipiter trivirgatus, Muscicapa latirostris, Strix leptogrammica, Prinia inornata, Pyrrhula nipalensis, Eudynamys scolopaceus, Dendrocitta formosae, Anthus hodgsoni, Schoeniparus brunneus, Phasianus colchicus, Actitis hypoleucos, Glaucidium brodiei, Alcippe morrisonia, Lophura swinhoii, Monticola solitarius, Pycnonotus sinensis X Pycnonotus taivanus, Buteo buteo, Copsychus saularis, Hierococcyx sparverioides
Blue	*Pitta nympha, Acridotheres cristatellus, Apus pacificus, Ardea alba, Dicrurus macrocercus, Anthus richardi, Spilornis cheela, Garrulax canorus, Megapomatorhinus erythrocnemis, Anas crecca, Egretta garzetta, Cuculus poliocephalus, Dicrurus aeneus, Apus nipalensis, Tringa stagnatilis, Eophona migratoria, Dendrocopos canicapillus, Alauda gulgula, Sternula albifrons, Charadrius dubius, Horornis fortipes, Pomatorhinus musicus, Tachybaptus ruficollis, Cyanoderma ruficeps, Mesophoyx intermedia, Calidris temminckii, Megalaima nuchalis, Xenus cinereus, Pluvialis fulva, Horornis diphone, Cuculus optatus, Arborophila crudigularis, Caprimulgus affinis, Garrulax taewanus, Myophonus insularis, Urocissa caerulea, Cuculus micropterus, Corvus macrorhynchos, Gallinago gallinago, Heterophasia auricularis, Acridotheres javanicus, Cinclidium leucurum, Fulica atra, Anas querquedula, Emberiza tristrami, Euodice malabarica, Lonchura striata, Tringa ochropus, Copsychus malabaricus,*

Amaurornis phoenicurus, Turdus pallidus, Pycnonotus sinensis, Turdus poliocephalus, Ianthocincla sannio, Spizixos semitorques, Motacilla alba, Anas acuta, Calidris acuminata, Oriolus traillii, Ardeola bacchus, Pericrocotus divaricatus, Riparia riparia, Columba pulchricollis, Sturnia sinensis, Butastur indicus, Gallirallus striatus, Pericrocotus solaris, Muscicapa griseisticta, Rallina eurizonoides, Sturnia malabarica, Prinia flaviventris, Motacilla cinerea, Bambusicola thoracicus, Eurystomus orientalis, Tringa totanus, Anthus cervinus, Cecropis striolata, Turdus chrysolaus, Accipiter soloensis, Halcyon coromanda, Anas penelope, Nycticorax nycticorax, Acrocephalus orientalis, Delichon dasypus, Pernis ptilorhynchus, Charadrius alexandrinus, Accipiter virgatus, Ictinaetus malayensis, Anas zonorhyncha, Zoothera dauma, Streptopelia orientalis, Calidris subminuta, Tringa nebularia, Parus monticolus, Yuhina brunneiceps, Hirundo tahitica, Lanius cristatus, Gallinula chloropus, Dicaeum ignipectus, Calidris ruficollis, Falco tinnunculus, Streptopelia tranquebarica, Psittacula krameri, Hypsipetes leucocephalus, Aegithalos concinnus, Treron formosae, Acridotheres tristis, Hirundo rustica, Ixobrychus cinnamomeus, Pycnonotus taivanus, Streptopelia chinensis, Sinosuthora webbiana, Himantopus himantopus, Rostratula benghalensis, Calliope calliope, Columba livia, Pandion haliaetus, Passer montanus, Pica pica, Lonchura punctulata, Prinia crinigera, Turdus eunomus, Turnix suscitator, Hypsipetes amaurotis, Riparia chinensis, Lanius schach, Abroscopus albogularis, Cisticola juncidis, Anas clypeata, Cairina moschata (domestic), Centropus bengalensis, Urosphena squameiceps, Horornis diphone/canturians, Terpsiphone atrocaudata, Ardea purpurea, Sturnus sericeus, Ixobrychus sinensis, Phoenicurus auroreus, Tringa brevipes, Phylloscopus inornatus, Liocichla steerii, Phylloscopus proregulus, Niltava vivida, Cisticola exilis, Bubulcus ibis, Motacilla tschutschensis, Oriolus chinensis, Larus crassirostris, Limosa limosa, Hypothymis azurea, Gorsachius melanolophus, Platalea minor, Elanus caeruleus, Chlidonias hybrida, Gracupica nigricollis, Milvus migrans, Lonchura atricapilla, Emberiza

	spodocephala, Phylloscopus borealis, Acridotheres burmannicus, Falco peregrinus, Phoenicurus fuliginosus, Picus canus, Erpornis zantholeuca, Treron sieboldii, Butorides striata, Anas platyrhynchos (Domestic), Zosterops japonicus, Porzana fusca, Alcedo atthis, Chalcophaps indica, Charadrius mongolus, Accipiter gentilis, Ardea cinerea, Otus lettia, Accipiter trivirgatus, Aythya fuligula, Muscicapa latirostris, Phylloscopus fuscatus, Prinia inornata, Ninox japonica, Aplonis panayensis, Emberiza rutila, Bycanistes bucinator, Dendrocitta formosae, Anthus hodgsoni, Estrilda melpoda, Glareola maldivarum, Schoeniparus brunneus, Phasianus colchicus, Actitis hypoleucos, Glaucidium brodiei, Alcippe morrisonia, Monticola solitarius, Pycnonotus sinensis X Pycnonotus taivanus, Calidris ferruginea, Tringa glareola, Phalacrocorax carbo
Green	*Gracula religiosa, Acridotheres cristatellus, Ardea alba, Dicrurus macrocercus, Anthus richardi, Spilornis cheela, Garrulax canorus, Garrulax canorus X Garrulax taewanus, Megapomatorhinus erythrocnemis, Anas crecca, Egretta garzetta, Dicrurus aeneus, Apus nipalensis, Tringa stagnatilis, Enicurus scouleri, Dendrocopos canicapillus, Alauda gulgula, Larus fuscus, Sternula albifrons, Charadrius dubius, Pomatorhinus musicus, Tachybaptus ruficollis, Cyanoderma ruficeps, Mesophoyx intermedia, Gallinago megala, Megalaima nuchalis, Recurvirostra avosetta, Pluvialis fulva, Hydrophasianus chirurgus, Lonchura oryzivora, Cuculus optatus, Caprimulgus affinis, Garrulax taewanus, Myophonus insularis, Trochalopteron morrisonianum, Urocissa caerulea, Corvus macrorhynchos, Gallinago gallinago, Acridotheres javanicus, Fulica atra, Anthus gustavi, Chlidonias leucopterus, Euodice malabarica, Lonchura striata, Tringa ochropus, Copsychus malabaricus, Amaurornis phoenicurus, Turdus pallidus, Pycnonotus sinensis, Spizixos semitorques, Motacilla alba, Calidris acuminata, Ardeola bacchus, Sturnia sinensis, Butastur indicus, Gallirallus striatus, Pericrocotus*

solaris, Muscicapa griseisticta, Sturnus cineraceus, Sturnia malabarica, Prinia flaviventris, Motacilla cinerea, Bambusicola thoracicus, Tringa totanus, Anthus cervinus, Cecropis striolata, Turdus chrysolaus, Anas penelope, Nycticorax nycticorax, Egretta sacra, Acrocephalus orientalis, Delichon dasypus, Charadrius alexandrinus, Accipiter virgatus, Acridotheres fuscus, Anas zonorhyncha, Streptopelia orientalis, Chloris sinica, Calidris subminuta, Tringa nebularia, Yuhina brunneiceps, Hirundo tahitica, Lanius cristatus, Gallinula chloropus, Falco tinnunculus, Streptopelia tranquebarica, Hypsipetes leucocephalus, Egretta eulophotes, Threskiornis aethiopicus, Acridotheres tristis, Hirundo rustica, Anser domestica, Ixobrychus cinnamomeus, Acridotheres grandis, Pycnonotus taivanus, Streptopelia chinensis, Sinosuthora webbiana, Himantopus himantopus, Rostratula benghalensis, Calliope calliope, Columba livia, Pandion haliaetus, Passer montanus, Pica pica, Lonchura punctulata, Prinia crinigera, Turdus eunomus, Turnix suscitator, Riparia chinensis, Lanius schach, Abroscopus albogularis, Cisticola juncidis, Anas clypeata, Cairina moschata (domestic), Centropus bengalensis, Horornis diphone/canturians, Sturnus sericeus, Anas platyrhynchos var. domestica, Ixobrychus sinensis, Phoenicurus auroreus, Tringa brevipes, Phylloscopus inornatus, Cisticola exilis, Motacilla citreola, Bubulcus ibis, Motacilla tschutschensis, Cygnus atratus (Domestic), Hypothymis azurea, Gorsachius melanolophus, Elanus caeruleus, Saxicola maurus, Chlidonias hybrida, Calidris alpina, Gracupica nigricollis, Milvus migrans, Lonchura atricapilla, Emberiza spodocephala, Phylloscopus borealis, Acridotheres burmannicus, Cacatua galerita, Falco peregrinus, Phoenicurus fuliginosus, Treron sieboldii, Butorides striata, Anas platyrhynchos (Domestic), Zosterops japonicus, Porzana fusca, Alcedo atthis, Charadrius mongolus, Ardea cinerea, Otus lettia, Accipiter trivirgatus, Muscicapa latirostris, Alauda arvensis, Sturnus vulgaris, Prinia inornata, Aplonis panayensis, Eudynamys scolopaceus, Dendrocitta formosae, Estrilda melpoda, Estrilda astrild, Glareola maldivarum, Sterna hirundo, Phasianus colchicus, Charadrius hiaticula, *Actitis hypoleucos, Alcippe morrisonia, Tarsiger cyanurus, Monticola solitarius, Copsychus saularis, Troglodytes troglodytes, Tringa glareola*

Glossary of Terms

Affiliation network are two mode networks that describe not only the direct linkage between members but also the events these members attend to (biologically, the habitat that species occupy) and the relationship among events.

Asymptote (of species accumulation curve) represents the saturation point in species richness, it can help determine the least sampling effort for estimating total richness in ecologically similar regions.

Clench function is derived from the **Michaelis-Menten kinetics** which is an enzymatic reaction rate curve based on the concentration of a substrate. Biologically, this means that the probability of adding new species increases (up to an asymptote) as more time is spent in the field.

Edge (in a network) represents the connection by a relationship between a pair of node (the basic unit of network).

Fowlkes-Mallows index is an index compares two clusters and evaluates the extent of similarity or dissimilarity between the clusters.

Indicator species define characteristics of habitat and can be applied to predict the habitat type of unknown or unsampled sites.

Michaelis-Menten kinetics is an equation in biochemistry describing the rate of enzymatic reactions by relating reaction rate to the concentration of a substrate. The reaction plot is a saturate curve with asymptote representing maximum reaction rate achieved by the system.

Modularity defines subgroups (termed modules) in a network. The connections between nodes (the basic unit of network) are measured and nodes possess a higher degree of connections are assigned in the same module, with lower degree of connections to nodes in other modules.

Mode (in a network) means the type of nodes (e.g. two-mode network), or the type of network (e.g. default mode network).

Modules (in a network) are subgroups within a network. Nodes in one module have stronger or more frequent connections among them comparing to nodes outside this module.

Module hub is the node with the highest degree in a module and represents the most dominant member (i.e. the member that links to most other members). This type of node can efficiently indicate the presence of other members in this module.

Network (ecological) representing biological interactions in an ecological communities. In this type of network nodes usually are species and the edges stands for interactions between species. The property of ecological networks can be used to illustrate the stability or fragility of ecological communities.

Nestedness is a measure of structure in an ecological system. In a species-sites system, species richness in each site are arranged from speciose site to depauperate site and the species in depauperate site are a subset of those in speciose ones. Nestness is measured by ‘temperature’ which indicates the system entropy. Therefore low nestedness reflects high turnover of species in space or time wheareas high nestedness means ordered accumulation or loss.

Node (in a network) is the basic element for constructing a network.

Rarefaction curve is a technique to assess species richness from a given number of individual samples, based on rarefaction which reduce the curve length for comparing species richness data among sets with different sample sizes. This curve is a plot of the number of species as a function of the number of samples by multiple random re-sampling process.

Species accumulation curve is a plot descries the relationship between species richness and sampling effort or area. Species number accumulates along with the increase of sampling size or sampling area.

Species distribution model applies environmental variables to modelling species spatial distribution based on species’ environment requirements (realized ecological niche).

TWINSPAN is a short form of Two Way Indicator Species Analysis which is a numerical method on habitat association of species assemblage.

List of References

Ackerly, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences* 164:164–184.

Agresti, A., and M. Kateri. 2011. Categorical data analysis. Springer, Berlin Heidelberg.

Allan, E., P. Manning, F. Alt, J.Binkenstein, S. Blaser, N. Blüthgen, S. Böhm, F. Grassein, N. Hölzel, V. H. Klaus, T. Kleinebecker, E. K. Morris, Y. Oelmann, D. Prati, S. C. Renner, M. C. Rillig, M. Schaefer, M. Schloter, B. Schmitt, I. Schöning, M. Schrumpf, E. Solly, E. Sorkau, J. Steckel, I. Steffen-Dewenter, B. Stempfhuber, M. Tschapka, C. N. Weiner, W. W. Weisser, M. Werner, C. Westphal, W. Wilcke, and M. Fischer. 2015. Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecology Letters* 18:834–843.

Almeida-Neto, M., P. R. J. Guimarães, R. D. Loyota, and W. Ulrich. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* 117:1227–1239.

Amend, A. S., D. J. Barshis, and T. A. Oliver. 2012. Coral-associated marine fungi form novel lineages and heterogeneous assemblages. *The ISME Journal* 6:1291–1301.

de Araújo, C. B., L. O. Marcondes-Machado, and G. C. Costa. 2014. The importance of biotic interactions in species distribution models: A test of the Eltonian noise hypothesis using parrots. *Journal of Biogeography* 41:513–523.

Araújo, M. B., A. Rozenfeld, C. Rahbek, and P. A. Marquet. 2011. Using species co-occurrence networks to assess the impacts of climate change. *Ecography* 34:897–908.

Araújo, M. B., W.Thuiller, and R. G. Pearson. 2006. Climate warming and the decline of amphibians and reptiles in Europe. Journal of Biogeography 33:1712–1728.

Arita, H. T., J. Andrés Christen, P. Rodríguez, and J. Soberón. 2008. Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications. *The American Naturalist* 172:519–32.

Arrhenius, O. 1921.Species and Area. *The Journal of Ecology* 9:95.

Ashcroft, M. B., J. R. Gollan, D. P. Faith, G. A. Carter, S. A. Lassau, S. G. Ginn, M. W. Bulbert, and G. Cassis. 2010. Using generalised dissimilarity models and many small samples to improve the efficiency of regional and landscape scale invertebrate sampling. *Ecological Informatics* 5:124–132.

Atmar, W., and B. D. Patterson. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96:373–382.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157:101–118.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200:1–19.

Baber, J. M., E. Fleishman, K. J. Babbitt, and T. L. Tarr. 2004. The relationship between wetland hydroperiod and nestedness patterns in assemblages of larval amphibians and predatory macroinvertebrates. *Oikos*, 107: 16-27.

Balvanera, P., I. Siddique, L. Dee, A. Paquette, F. Isbell, A. Gonzalez, J. Byrnes, M. I. O’Connor, B. A. Hungate, and J. N. Griffin. 2014. Linking biodiversity and ecosystem services: Current uncertainties and the necessary next steps. *BioScience* 64:49–57.

Banerjee, S., A. E. Gelfand, A. O. Finley, H. Sang, S. Banerjee, and A. O. Finley. 2008. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 70:825–848.

Barthélemy, M. 2011. Spatial networks. *Physics Reports* 499:1–101.

Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* 19:134–143.

Baselga, A., and M. B.Araújo. 2010. Do community-level models describe community variation effectively? *Journal of Biogeography* 37:1842–1850.

Bebber, D. P., F. H. C. Marriott, K. J. Gaston, S. A. Harris, and R. W. Scotland. 2007. Predicting unknown species numbers using discovery curves. *Proceedings of the Royal Society B* 274:1651–8.

Béguinot, J. 2015. Extrapolation of the species accumulation curve for incomplete species samplings: a new nonparametric approach to estimate the degree of sample completeness and decide when to stop sampling. *Annual Research and Review in Biology* 8:1–9.

Biella, P., J.Ollerton, M.Barcella, andS.Assini. 2017.Network analysis of phenological units to detect important species in plant-pollinator assemblages: can it inform conservation strategies? Community Ecology 18:1–10.

Bloch, C. P., C. L. Higgins, and M. R. Willig. 2007. Effects of large-scale disturbance on metacommunity structure of terrestrial gastropods : temporal trends in nestedness. *Oikos* 116:395–406.

Blondel, V. D., J. Guillaume, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008:10008.

Borcard, D. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73:1045–1055.

Braun-Blanquet. 1932. Plant Sociology. McGraw-Hill Book Company, New York.

Brose, U., N. D. Martinez, R. J. Williams, U. Brose, N. E. O. D. Martinez, and R. J. Williams. 2003. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology* 84:2364–2377.

Brualdi, R. A., and J. G. Sanderson. 1999. Nested species subsets, gaps, and discrepancy. *Oecologia* 119:256–264.

Bruun, H. H., and J. Moen. 2003. Nested communities of alpine plants on isolated mountains: relative importance of colonization and extinction. *Journal of Biogeography* 30:297–303.

Bunge, J., and M. Fitzpatrick. 1993. Estimating the number of species: a review. *Journal of the American Statistical Association* 88:364–373.

De Cáceres, M., and P. Legendre. 2009. Associations between species and groups of sites : indices and statistical inference. *Ecology* 90:3566–3574.

De Cáceres, M., P. Legendre, and M. Moretti. 2010. Improving indicator species analysis by combining groups of sites. *Oikos* 119:1674–1684.

De Cáceres, M., P. Legendre, S. K.Wiser, and L. Brotons. 2012. Using species combinations in indicator value analyses. *Methods in Ecology and Evolution* 3:973–982.

Cardoso, P., F. Rigal, P. A.V. Borges, and J. C. Carvalho. 2014. A new frontier in biodiversity inventory: a proposal for estimators of phylogenetic and functional diversity. *Methods in Ecology and Evolution* 5:452–461.

Cardoso, P., N. Scharff, C. Gaspar, S. S. Henriques, R. Carvalho, P. H. Castro, J. B. Schmidt, I. Silva, T. Szüts, A. DeCastro, and L. C. Crespo. 2008. Rapid biodiversity assessment of spiders (Araneae) using semi-quantitative sampling: a case study in a Mediterranean forest. *Insect Conservation and Diversity* 1:71–84.

Cassini, M. H. 2011. Ecological principles of species distribution models: the habitat matching rule. *Journal of Biogeography* 38:2057–2065.

Cayuela, L., N. J. Gotelli, and R. K. Colwell. 2015. Ecological and biogeographic null hypotheses for comparing rarefaction curves. *Ecological Monographs* 85:437–455.

Celebi, M. E., H. A. Kingravi, and P. A. Vela. 2013. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications* 40:200–210.

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs* 84:45–67.

Chao, A., and L. Jost. 2015a. Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution* 6:873–882.

Chao, A., and L. Jost. 2015b. Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution* 6:873–882.

Chao, A., and T. -J. Shen. 2004. Nonparametric prediction in species sampling. *Journal of Agricultural, Biological, and Environmental Statistics* 9:253–269.

Chao, A., Y. T. Wang, and L. Jost. 2013. Entropy and the species accumulation curve: A novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution* 4:1091–1100.

Chave, J. 2004. Neutral theory and community ecology. *Ecology Letters* 7:241–253.

Chaves, L. F., and N. Ariez. 2016. Nestedness patterns of sand fly (Diptera: Psychodidae) species in a neotropical semi-arid environment. *Acta Tropica* 153:7–13.

Cheruvelil, K. S., S. Yuan, K. E. Webster, P. N. Tan, J. F. Lapierre, S. M. Collins, C. E. Fergus, C. E. Scott, E. N. Henry, P. A. Soranno, C. T. Filstrup, and T. Wagner. 2017. Creating multithemed ecological regions for macroscale ecology: Testing a flexible, repeatable, and accessible clustering method. *Ecology and Evolution* 7:3046–3058.

Chytrý, M., L. Tichý, J. Holt, and Z. Botta-Dukát. 2002. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* 13:79–90.

Clark, J. S., A. E. Gelfand, C. W. Woodall, and K. Zhu. 2014. More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications* 24:990–999.

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18:117–143.

Clench, H. K. 1979. How to make regional lists of butterflies: some thoughts. *Journal of the Lepidopterist's Society* 33:216–231.

Colwell, R. K., A. Chao, N. J. Gotelli, S. Y. Lin, C. X. Mao, R. L. Chazdon, and J. T. Longino. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology* 5:3–21.

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 345:101–18.

Connolly, S. R., M. Dornelas, D. R. Bellwood, and T. P. Hughes. 2009. Testing species abundance models: a new bootstrap approach applied to Indo-Pacific coral reefs. *Ecology* 90:3138–3149.

Cook, R. R., and J. F. Quinn. 1995. The influence of colonization in nested species subsets. *Oecologia* 102:413–424.

Cousins, S. H. 1991. Species diversity measurement: choosing the right index. *Trends in Ecology and Evolution* 6:190–192.

Csardi G, N. T. 2006. The igraph software package for complex network research. *InterJournal, Complex Systems* 1695.

Cumming, G. S., Ö. Bodin, H. Ernstson, and T. Elmqvist. 2010. Network analysis in conservation biogeography: challenges and opportunities. *Diversity and Distributions* 16:414–425.

Cutler, A. 1991. Nested faunas and extinction in fragmented habitats. *Conservation Biology* 5:496–505.

D'Amen, M., C. Rahbek, N. E. Zimmermann, and A. Guisan. 2017. Spatial predictions at the community level: from current approaches to future frameworks. *Biological Reviews* 92:169–187.

Dengler, J. 2009. Which function describes the species-area relationship best? A review and empirical evaluation. *Journal of Biogeography* 36:728–744.

Dengler, J., and S. Boch. 2008. Sampling-design effects on properties of species-area relationships - A case study from Estonian dry grassland communities. *Folia Geobotanica* 43:289–304.

Devoto, M., S. Bailey, P. Craze, and J. Memmott. 2012. Understanding and planning ecological restoration of plant-pollinator networks. *Ecology Letters* 15:319–328.

Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 in M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Harvard University Press, Cambridge, Massachusetts, USA.

Díaz-Francés, E., and J. Soberón. 2005. Statistical estimation and model selection of species-accumulation functions. *Conservation Biology* 19:569–573.

Ding, T., C. Juan, R. Lin, Y. Tsai, J. Wu, and Y. Yang. 2014. The 2014 CWBF Checklist of the Birds of Taiwan. Taipei, Taiwan.

Ding, T. S., H. W. Yuan, S. Geng, Y. S. Lin, and P. F. Lee. 2005. Energy flux, body size and density in relation to bird species richness along an elevational gradient in Taiwan. *Global Ecology and Biogeography* 14:299–306.

Dirnböck, T. 2003. A regional impact assessment of climate and land-use change on alpine vegetation. *Journal of Biogeography* 30:401–417.

Doncaster, C. P., V. A. Chávez, C. Viguier, R. Wang, E. Zhang, X. Dong, J. A. Dearing, P. G. Langdon, and J. G. Dyke. 2016. Early warning of critical transitions in biodiversity from compositional disorder. *Ecology* 97:3079–3090.

Dufrêne, M., and P. Legendre. 1997. Species assemblages and indicator species : the need for a flexible asymmetrical approach. *Ecological Monographs* 67:345–366.

Dunne, J. A., R. J. Williams, and N. D. Martinez. 2002. Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences* 99:12917–22.

Ehouman, N. M., S. Tiho, and M. Dagnogo. 2012. Co-occurrence of earthworms in Lamto savanna: A null model analysis of community structure. *European Journal of Soil Biology* 53:40–47.

Elith, J., and J. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40:677–697.

Elmendorf, S. C., and S. P. Harrison. 2009. Temporal variability and nestedness in California grassland species composition. *Ecology* 90:1492–1497.

Engen, S., V. Grøtan, and B. E. Sæther. 2011. Estimating similarity of communities: A parametric approach to spatio-temporal analysis of species diversity. *Ecography* 34:220–231.

Estrada-peña, A., and J. DeFuente. 2016. Species interactions in occurrence data for a community of tick-transmitted pathogens. *Scientific Data* 3:1–13.

Estrada, E. 2007. Food webs robustness to biodiversity loss: the roles of connectance, expansibility and degree distribution. *Journal of Theoretical Biology* 244:296–307.

Farine, D. R., C. J. Garroway, and B. C. Sheldon. 2012. Social network analysis of mixed-species flocks: exploring the structure and evolution of interspecific social behaviour. *Animal Behaviour* 84:1271–1277.

Fattorini, S., and W. Ulrich. 2012. Drivers of species richness in European Tenebrionidae (Coleoptera). *Acta Oecologica* 43:22–28.

Ferrier, S., and A. Guisan. 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology* 43:393–404.

Ficetola, G. F., W. Thuiller, and C. Miaud. 2007. Prediction and validation of the potential global distribution of a problematic alien invasive species - the American bullfrog. *Diversity and Distributions* 13:476–485.

Field, R., B. A. Hawkins, H.V. Cornell, D. J. Currie, J. A. F. Diniz-Filho, J. F. Guégan, D. M. Kaufman, J. T. Kerr, G. G. Mittelbach, T. Oberdorff, E. M. O'Brien, and J. R. G. Turner.

2009. Spatial species-richness gradients across scales: A meta-analysis. *Journal of Biogeography* 36:132–147.

Fischer, J., and D. B. Lindenmayer. 2002. Treating the nestedness temperature calculator as a ¨black box¨ can lead to false conclusions. *Oikos* 99:193–199.

Fischer, J., and D. B. Lindenmayer. 2005. Nestedness in fragmented landscapes: a case study on birds, arboreal marsupials and lizards. *Journal of Biogeography* 32:1737–1750.

Fleishman, E., R. B. Blair, and D. D. Murphy. 2001. Empirical validation of a method for umbrella species selection. *Ecological Applications* 11:1489–1501.

Fleishman, E., and D. D. Murphy. 1999. Patterns and processes of nestedness in a Great Basin butterfly assemblages: applications to conservation planning. *Oecologia* 119:133–139.

Fleishman, E., and R. Mac Nally. 2002. Topographic determinants of faunal nestedness in Great Basin butterfly assemblages: applications to conservation planning. *Conservation Biology* 16:422–429.

Fleishman, E., C. J. Betrus, R. B. Blair, R. Mac Nally, and D. D. Murphy. 2002. Nestedness analysis and conservation planning: the importance of place, environment, and life history across taxonomic groups. *Oecologia* 133: 78-89.

Flinn, K. M., M. J. Lechowicz, and M. J. Waterway. 2008. Plant species diversity and composition of wetlands within an upland forest. *American Journal of Botany* 95:1216–1224.

Fortunato, S. 2010. Community detection in graphs. *Physics Reports* 486:75–174.

Fowlkes, E. B., and C. L. Mallows. 1983. A method for comparing two hierarchical clusterings. *Jounal of the American Statistical Association* 78:553–569.

García-Callejas, D., and M. B. Araújo. 2016. The effect of model and data complexity on predictions from species distributions models. *Ecological Modelling* 326:4–12.

Gavish, Y., C. J. Marsh, M. Kuemmerlen, S. Stoll, P. Haase, and W. E. Kunin. 2017. Accounting for biotic interactions through alpha-diversity contraints in stacked species distribution models. *Methods in Ecology and Evolution* Online fir.

Genini, J., M. C. Côrtes, P. R. J. Guimarães, and M. Galetti. 2012. Mistletoes play different roles in a modular host–Parasite network. *Biotropica* 44:171–178.

Gibbons, D. W., J. B. Reid, and R. A. Chapman. 1993. The new atlas of breeding birds in Britain and Ireland: 1988–1991. Poyser, London.

Gilbert, A. J. 2009. Connectance indicates the robustness of food webs when subjected to species loss. *Ecological Indicators* 9:72–80.

Gilman, S. E., M. C. Urban, J. Tewksbury, G. W. Gilchrist, and R. D. Holt. 2010. A framework for community interactions under climate change. *Trends in Ecology and Evolution* 25:325–331.

Gleason, H. A. 1922. On the relation between species and area. *Ecology* 3:158–162.

Godsoe, W., and L. J. Harmon. 2012. How do species interactions affect species distribution models? *Ecography* 35:811–820.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81:2606–2621.

Gotelli, N. J., M. J. Anderson, H. T. Arita, A. Chao, R. K. Colwell, S. R. Connolly, D. J. Currie, R. R. Dunn, G. R. Graves, J. L. Green, J.-A. Grytnes, Y.-H. Jiang, W. Jetz, S. Kathleen Lyons, C. M. McCain, A. E. Magurran, C. Rahbek, T. F. L. V. B. Rangel, J. Soberón, C. O. Webb, and M. R. Willig. 2009. Patterns and causes of species richness: a general simulation model for macroecology. *Ecology letters* 12:873–86.

Gotelli, N. J., andR. K.Colwell. 2001.Quantifying biodiversity : procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4:379–391.

Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond’s assembly rules model. *Ecology* 83:2091–2096.

Gray, J. S., K. I. Ugland, and J. Lambshead. 2004. On species accumulation and species-area curves. *Global Ecology and Biogeography* 13:567–568.

Grimm, V., and S. F. Railsback. 2012. Pattern-oriented modelling: a “multi-scope” for predictive systems ecology. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367:298–310.

Grundel, R., K. J. Frohnapple, D. N. Zaya, G. A. Glowacki, C. J. Weiskerger, T. A. Patterson, and N. B. Pavlovic. 2014. Geographic coincidence of richness, mass, conservation value, and response to climate of U.S. land birds. *Ecological Applications* 24:791–811.

Guillera-Arroita, G., J. J. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P. E. Lentini, M. A. Mccarthy, R. Tingley, and B. A. Wintle. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* 24:276–292.

Guimerà, R., and L. A. N. Amaral. 2005. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics* 2005:1–13.

Guisan, A., C. H. Graham, J. Elith, F. Huettmann, M. Dudik, S. Ferrier, R. Hijmans, A. Lehmann, J. Li, L. G. Lohmann, B. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, S. E. Williams, M. S. Wisz, and N. E. Zimmermann. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13:332–340.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993–1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186.

Gwinn, D. C., M. S. Allen, K. I. Bonvechio, M.V. Hoyer, and L. S. Beesley. 2015. Evaluating estimators of species richness: The importance of considering statistical error rates. *Methods in Ecology and Evolution* 7:294–302.

Harms, K. E., R. Condit, S. P. Hubbell, and R. B. Foster. 2001. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology* 89:947–959.

Hawkins, B. A., and J. A. Felizola Diniz-Filho. 2004. “Latitude” and geographic patterns in species richness. *Ecography* 27:268–272.

He, F., and P. Legendre. 1996. On species-area relations. *American Naturalist* 148:719–737.

He, F., P. Legendre, and J.V LaFrankie. 1996. Spatial pattern of diversity in a tropical rain forest in Malaysia. Journal of Biogeography 23:57–74.

Heino, J., H. Mykrä, and T. Muotka. 2009. Temporal variability of nestedness and idiosyncratic species in stream insect assemblages. *Diversity and Distributions* 15:198–206.

Herczeg, R., and G. F.Horvath. 2015. Species composition and nestedness of small mammal assemblages in two disturbed marshlands. *North-Western Journal of Zoology* 11:183–193.

Higgins, C. L., M. R. Willig, and R. E. Strauss. 2006. The role of stochastic processes in producing nested patterns of species distributions. *Oikos* 114:159–167.

Hill, J. K., M. A. Gray, C. V. Khen, S. Benedick, N. Tawatao, and K. C. Hamer. 2011. Ecological impacts of tropical forest fragmentation: how consistent are patterns in species richness and nestedness? *Philosophical Transactions of the Royal Society B: Biological Sciences*: 366: 3265-3276.

Hill, M. O. 1979. TWINSPAN: a fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Cornell University, New York, USA.

Holt, A. R., K. J. Gaston, and F. He. 2002. Basic and applied ecology occupancy-abundance relationships and spatial distribution : a review. *Basic and Applied Ecology* 13:1–13.

Holt, R. D. 2009. Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences* 106:19659–19665.

Holt, R. D., and M. Barfield. 2009. Trophic interactions and range limits: the diverse roles of predation. *Proceedings of the Royal Society B* 276:1435–1442.

Hortal, J., P. Garcia-Pereira, and E. García-Barros. 2004. Butterfly species richness in mainland Portugal : predictive models of geographic distribution patterns. *Ecography* 27:68–82.

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

Hubbell, S. P. 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology* 19:166–172.

Hubbell, S. P. 2006. Neutral theory and the evolution of ecological equivalence. Ecology 87:1387–1398.

Hylander K., C. Nilsson, B. Gunnar Jonsson, and T. Göthner. 2005. Differences in habitat quality explain nestedness in a land snail meta-community. *Oikos*: 108: 351-361.

Jeong, H., S. P. Mason, A.L. Barabási, and Z. N. Oltvai. 2001. Lethality and centrality in protein networks. *Nature* 411:41–2.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution* 19:101–108.

Joseph, K., C. H. Tan, and K. M. Carley. 2012. "Beyond local, categories and friends: clustering foursquare users with latent topics. Pages 22–29Proceedings of the 2012 ACM conference on ubiquitous computing. ACM.

Keeling, M. J. 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* 294:813–817.

Kéfi, S., E. L. Berlow, E. A. Wieters, L. N. Joppa, S. A. Wood, U. Brose, and S. A. Navarrete. 2015. Network structure beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology* 96:291–303.

Kempen, B., D. J. Brus, G. B. M. Heuvelink, and J. J. Stoorvogel. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma* 151:311–326.

Kissling, W. D., C. F. Dormann, J. Groeneveld, T. Hickler, I. Kühn, G. J. Mcinerny, J. M. Montoya, C. Römermann, K. Schiffers, F. M. Schurr, A. Singer, J. C. Svenning, N. E. Zimmermann, and R. B. O'Hara. 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography* 39:2163–2178.

Kluge, J., M. Kessler, and R. R. Dunn. 2006. What drives elevational patterns of diversity? A test of geometric constraints, climate and species pool effects for pteridophytes on an elevational gradient in Costa Rica. *Global Ecology and Biogeography* 15:358–371.

Ko, C. J., M. W. Fan, Y. X. Jiang, W. J. Yu, Y. Y. Lo, K. W. Huang, R. S. Lin, C. C. Cheng, and P. F. Lee. 2013. 2012 Taiwan Breeding Bird Survey Annual Report. Nantou, Taiwan.

Koh, C. N., P. F. Lee, and R. S. Lin. 2006. Bird species richness patterns of northern Taiwan: primary productivity, human population density, and habitat heterogeneity. *Diversity and Distributions* 12:546–554.

Koleff, P., K. J. Gaston, and J. J. Lennon. 2003. Measuring beta diversity for presence-absence. *Journal of Animal Ecology* 72:367–382.

Krause, A. E., K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor. 2003. Compartments revealed in food-web structure. *Nature* 426:282–5.

Larsen, P. E., D. Field, and J. A. Gilbert. 2012. Predicting bacterial community assemblages using an artificial neural network approach. *Nature Methods* 9:621–5.

Latimer, A. M., S. Banerjee, H. Sang, E. S. Mosher, and J. A. Silander. 2009. Hierarchical models facilitate spatial analysis of large data sets: A case study on invasive plant species in the northeastern United States. *Ecology Letters* 12:144–154.

LeLay, G., R. Engler, E. Franc, and A. Guisan. 2010. Prospective sampling based on model ensembles improves the detection of rare species. *Ecography* 33:1015–1027.

Leathwick, J. R., and M. P. Austin. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82:2560–2573.

Lee, P., T.Ding, F. Hsu, and S. Geng. 2004. Breeding bird species richness in Taiwan : distribution on gradients of elevation , primary productivity and urbanization. *Journal of Biogeography* 31:307–314.

Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271–280.

Legendre, P., and L. Legendre. 2012. Numerical Ecology. 3rd Editio. Elsevier, Oxford.

Lehmann, A., J. M. Overton, and J. R. Leathwick. 2002. GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling* 157:189–207.

Lennon, J. J., P. Koleff, J. J. D. Greenwood, and K. J. Gaston. 2001. The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology* 70:966–979.

Levy, R., and E. Borenstein. 2013. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences* 110:12804–12809.

Lischke, H., N. E. Zimmermann, J. Bolliger, S. Rickebusch, and T. J. Löffler. 2006. TreeMig: A forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale. *Ecological Modelling* 199:409–420.

Lusseau, D. 2003. The emergent properties of a dolphin social network. *Proceedings of the Royal Society B: Biological Sciences* 270 Suppl: S186-8.

MacArthur, R. H. 1967. The theory of island biogeography. Princeton University Press, Princeton NJ.

Magnussen, S. 2014. Species richness estimation from probabilities of observing new species. *Journal of Forest Research* 19:194–204.

Magura, T. 2002. Carabids and forest edge: Spatial pattern and edge effect. *Forest Ecology and Management* 157:23–37.

McAbendroth, L., A. Foggo, S. D. Rundle, and D. T. Bilton. 2005. Unravelling nestedness and spatial pattern in pond assemblages. *Journal of Animal Ecology* 74:41–49.

McGill, B., and C. Collins. 2003. A unified theory for macroecology based on spatial patterns of abundance. *Evolutionary Ecology* Research 5:469–492.

McInerny, G. J., and D. W. Purves. 2011. Fine-scale environmental variation in species distribution modelling: regression dilution, latent variables and neighbourly advice. *Methods in Ecology and Evolution* 2:248–257.

Meier, E. S., F. Kienast, P. B. Pearman, J. C. Svenning, W. Thuiller, M. B. Araújo, A. Guisan, and N. E. Zimmermann. 2010. Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography* 33:1038–1048.

Melo, A. S., R. A. S. Pereira, A. J. Santos, G. J. Shepherd, G. Machado, H. F. Medeiros, and R. J. Sawaya. 2003. Comparing species richness among assemblages using sample units : why not use extrapolation methods to standardize different sample sizes ? *Oikos* 101:398–410.

Merow, C., M. J. Smith, and J. A. Silander. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36:1058–1069.

Miller, R. I. ., and R. G., Wiegert. 1989. Documenting completeness, species-area relations, and the species-abundance distribution of a regional flora. *Ecology* 70:16–22.

Milligan, G. W. 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* 45:325–342.

Mokany, K., and S. Ferrier. 2011. Predicting impacts of climate change on biodiversity: a role for semi-mechanistic community-level modelling. *Diversity and Distributions* 17:374–380.

Mokany, K., T. D. Harwood, J. M. Overton, G. M. Barker, and S. Ferrier. 2011. Combining α- and β-diversity models to fill gaps in our knowledge of biodiversity. *Ecology letters* 14:1043–51.

Montoya, J. M., and R.V. Sol. 2002. Small world patterns in food webs. *Journal of Theoretical Biology* 214:405–12.

Moody, J., D. Mcfarland, and S. Bender. 2005. Dynamic network visualization. *American Journal of Sociology* 110:1206–1241.

Moody, J., and D. R. White. 2003. Structural cohesion and embeddedness: a hierarchical concept of social groups. *American Sociological Review* 68:103–127.

Moreno, C. E., and G. Halffter. 2000. Assessing the completeness of bat biodiversity inventories using species accumulation curves. *Journal of Applied Ecology* 37:149–158.

Morin, P. J. 1999. Community Ecology. Blackwell Science, Inc.

Mouton, A. M., B. DeBaets, and P. L. M. Goethals. 2010. Ecological relevance of performance criteria for species distribution models. *Ecological Modelling* 221:1995–2002.

Mutshinda, C. M., R. B. O’Hara, and I. P. Woiwod. 2009. What drives community dynamics? *Proceedings of the Royal Society B* 276:2923–2929.

Mutshinda, C. M., R. B. O'Hara, and I. P. Woiwod. 2011. A multispecies perspective on ecological impacts of climatic forcing. *Journal of Animal Ecology* 80:101–107.

Mac Nally, R., G. Horrocks, and A. F. Bennett. 2002. Nestedness in fragmented landscapes : birds of the box-ironbark forests of south-eastern Australia. *Ecography* 6:651–660.

Navia, A. F., E. Cortés, and P. A. Mejía-Falla. 2010. Topological analysis of the ecological importance of elasmobranch fishes: A food web study on the Gulf of Tortugas, Colombia. *Ecological Modelling* 221:2918–2926.

Newman, M. 2004. Analysis of weighted networks. *Physical Review E* 70:56131.

Newman, M. E. J. 2010. Networks: An Introduction. Oxford University Press.

Newman, M., and M. Girvan. 2004. Finding and evaluating community structure in networks. *Physical Review E* 69:26113.

Ney-Nifle, M., and M. Mangel. 1999. Species–Area Curves Based on Geographic Range and Occupancy. *Journal of Theoretical Biology* 196:327–342.

Norton, J., J. W. Lewis, and D. Rollinson. 2004. Temporal and spatial patterns of nestedness in eel macroparasite communities. *Parasitology* 129:203–211.

Obsomer, V., M. Dufrêne, P. Defourny, and M. Coosemans. 2013. *Anopheles* species associations in Southeast Asia: indicator species and environmental influences. *Parasites and Vectors* 6:136.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, and H. Wagner. 2013. vegan: Community Ecology Package. R package version 2.0-7.

Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* 83:171–193.

Olesen, J. M., J. Bascompte, Y. L. Dupont, and P. Jordano. 2007. The modularity of pollination networks. *Proceedings of the National Academy of Sciences* 104:19891–6.

Olito, C., and J. W. Fox. 2015. Species traits and abundances predict metrics of plant-pollinator network structure, but not pairwise interactions. *Oikos* 124:428–436.

Oppel, S., A. Meirinho, I. Ramírez, B. Gardner, A. F. O'Connell, P. I. Miller, and M. Louzao. 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation* 156:94–104.

Ovaskainen, O., J. Hottola, and J. Shtonen. 2010. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology* 91:2514–2521.

Ovaskainen, O., and J. Soininen. 2011. Making more out of sparse data: hierarchical modeling of species communities. *Ecology* 92:289–295.

Paine, R. T. 1966. Food web complexity and species diversity. *The American Naturalist* 100:65–75.

Paine, R. T. 1974. Intertidal community structure. *Oecologia* 15:93–120.

Patterson, B. D. 1990. On the temporal development of nested subset patterns of species composition. *Oikos* 59:330–342.

Patterson, B. D., and W. Atmar. 1986. Nested subsets and the structure of insular mammlian faunas and archipelagos. *Biological Journal of Linnen Society* 28:65–82.

Pearman, P. B., and D. Weber. 2007. Common species determine richness patterns in biodiversity indicator taxa. *Biological Conservation* 138:109–119.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12:361–371.

Pearson, R. G., W. Thuiller, M. B. Araújo, E. Martinez-Meyer, L. Brotons, C. McClean, L. Miles, P. Segurado, T. P. Dawson, and D. C. Lees. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* 33:1704–1711.

Pereira, M., P. Segurado, and N. Neves. 2011. Using spatial network structure in landscape management and planning: a case study with pond turtles. *Landscape and Urban Planning* 100:67–76.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231–259.

Picazo, F., D. T. Bilton, J. L. Moreno, D. Sánchez-Fernández, and A. Millán. 2012. Water beetle biodiversity in Mediterranean standing waters: assemblage composition, environmental drivers and nestedness patterns. *Insect Conservation and Diversity* 5:146–158.

Planer, J. D., Y. Peng, A. L. Kau, L.V. Blanton, I. M. Ndao, P. I. Tarr, B. B. Warner, and J. I. Gordon. 2016. Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature* 534:263–266.

Podani, J. 2006. Braun-Blanquet's legacy and data analysis in vegetation science. *Journal of Vegetation Science* 17:113–117.

Podani, J., F. Jordan, and D. Schmera. 2014. A new approach to exploring architecture of bipartite (interaction) ecological networks. *Journal of Complex Networks* 2:168–186.

Podani, J., and D. Schmera. 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence - absence data. *Oikos* 120:1625–1638.

Pons, P., and M. Latapy. 2005. Computing communities in large networks using random walks. *Journal of Graph Algorithms Application* 10:284–293.

Poore, M. E. D. 1955. The use of phytosociological methods in ecological investigations: I. The Braun-Blanquet system. *Journal of Ecology* 43:226–244.

Prell, C. 2011. Social network analysis: history, theory and methodology. SAGE Publications Ltd.

Prendergast, J. R. 1997. Species richness covariance in higher taxa: empirical tests of the biodiversity indicator concept. *Ecography* 20:210–217.

Prendergast, J. R., R. M. Quinn, J. H. Lawton, B. Eversham, and D. W. Gibbons. 1993. Rare species, the coincidence of idversity hotsopts and conservation strategies. *Nature* 365:335–337.

Prentice, I. C., A. Bondeau, W. Cramer, S.Harrison, T. Hickler, W. Lucht, S. Sitch, B. Smith, and M. T. Sykes. 2007. Dynamic global vegetation modeling: quantifying terrestrial ecosystem responses to large-scale environmental change. In *Terrestrial Ecosystems in a Changing World* (pp. 175-192). Springer Berlin Heidelberg.

Preston, F. W. 1962a. The canonical distribution of commonness and rarity: Part I. *Ecology* 43:185–215.

Preston, F. W. 1962b. The canonical distribution of commonness and rarity: Part II. *Ecology* 43:410–432.

Prober, S. M., J. W. Leff, S. T. Bates, E. T. Borer, J. Firn, W. S. Harpole, E. M. Lind, E. W. Seabloom, P. B. Adler, J. D. Bakker, E. E. Cleland, N. M. Decrappeo, E. Delorenze, N. Hagenah, Y. Hautier, K. S. Hofmockel, K. P. Kirkman, J. M. H. Knops, K. J. LaPierre, A. S. Macdougall, R. L. Mcculley, C. E. Mitchell, A. C. Risch, M. Schuetz, C. J. Stevens, R. J. Williams, and N. Fierer. 2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecology Letters* 18:85–95.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Randin, C. F., T. Dirnböck, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33:1689–1703.

Ricotta, C., M. Carboni, and A. T. R. Acosta. 2015. Let the concept of indicator species be functional! *Journal of Vegetation Science* 26:839–847.

Ricotta, C., and S. Pavoine. 2015. A multiple-site dissimilarity measure for species presence/absence data and its relationship with nestedness and turnover. *Ecological Indicators* 54:203–206.

Rodríguez-Gironés, M. A., and L. Santamaría. 2006.A new algorithm to calculate the nestedness temperature of presence-absence matrices. *Journal of Biogeography* 33:924–935.

Rollinson, C. R., M. W.Kaye, andC. D.Canham. 2016.Interspecific variation in growth responses to climate and competition of five eastern tree species. Ecology 97:1003–1011.

Romdal, T. S., R. K. C. Olwell, and C. Rahbek. 2005. The influence of band sum area, domain extent, and range sizes on the latitudinal mid-domain effect. *Ecology* 86:235–244.

Rosenzweig, M. L., andZiv. 1999. The echo pattern of species diversity: pattern and processes. *Ecography* 22:614–628.

Rosindell, J., S. P. Hubbell, and R. S. Etienne. 2011. The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. *Trends in Ecology and Evolution* 26:340–348.

Rudolf, V. H. W., and N. L. Rasmussen. 2013. Ontogenetic functional diversity: Size structure of a keystone predator drives functioning of a complex ecosystem. *Ecology* 94:1046–1056.

Ruhí, A., E. Chappuis, D. Escoriza, M. Jover, J. Sala, D. Boix, S. Gascón, and E. Gacia. 2014. Environmental filtering determines community patterns in temporary wetlands: A multi-taxon approach. *Hydrobiologia* 723:25–39.

Sattler, T., G. B. Pezzatti, M. P. Nobis, M. K. Obrist, T. Roth, and M. Moretti. 2014. Selection of multiple umbrella species for functional and taxonomic diversity to represent urban biodiversity. *Conservation Biology* 28:414–426.

Saura, S., and L. Rubio. 2010. A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. *Ecography* 33:523–537.

Scheiner, S. M. 2003. Six types of species-area curves. *Global Ecology and Biogeography* 12:441–447.

Schouten, M. A., P. A. Verweij, A. Barendregt, R. J. Kleukers, and P. C. De Ruiter. 2007. Nested assemblages of Orthoptera species in the Netherlands: the importance of habitat features and life-history traits. *Journal of Biogeography* 34: 1938-1946.

Schröder, B., and R. Seppelt. 2006. Analysis of pattern-process interactions based on landscape models-Overview, general concepts, and methodological issues. *Ecological Modelling* 199:505–516.

Seedorf, H., N. W. Griffin, V. K. Ridaura, A. Reyes, J. Cheng, F. E. Rey, M. I. Smith, G. M. Simon, R. H. Scheffrahn, D. Woebken, A. M. Spormann, W. VanTreuren, L. K. Ursell, M. Pirrung, A. Robbins-Pianka, B. L. Cantarel, V. Lombard, B. Henrissat, R. Knight, and J. I. Gordon. 2014. Bacteria from diverse habitats colonize and compete in the mouse gut. *Cell* 159:253–266.

Segurado, P., and M. B. Araújo. 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31:1555–1568.

Sinclair, S. J., M. D. White, and G. R. Newell. 2010. How rseful are species distribution models for managing biodiversity under future climates ? *Ecology and Society* 15:8–20.

Skelly, D. K., E. E. Werner, and S. A. Cortwright. 1999. Long-term distributional dynamics of a Michigan amphibian assemblage. *Ecology* 80:2326–2337.

Smith, E. P., and G. van Belle. 1984. Nonparametric estimation of species richness. *Biometrics* 40:119–129.

Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115-1123.

Soberón, J. M., and J. B. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology* 7:480–488.

Socolar, J. B., J. J. Gilroy, W. E. Kunin, and D. P. Edwards. 2016. How Should Beta-Diversity Inform Biodiversity Conservation? *Trends in Ecology and Evolution* 31:67–80.

Soininen, J., R. McDonald, and H. Hillebrand. 2007. The distance decay of similarity in ecological communities. *Ecography* 30:3–12.

Solé, R.V., and J. Bascompte. 2006. Spatial self-organization: from pattern to process. Self-Organization in Complex Ecosystems (MPB-42) (Pages 65–126). Princeton University Press, Princeton.

Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. *Oecologia* 85:74–79.

Sutherland, W. J. 2006. Predicting the ecological consequences of environmental change: a review of the methods. *Journal of Applied Ecology* 43:599–616.

Taylor, D. L., T. N. Hollingsworth, J. W. McFarland, N. J. Lennon, C. Nusbaum, and R. W. Ruess. 2014. A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological Monographs* 84:3–20.

Tews, J., U. Brose, V. Grimm, K. Tielbörger, M. C. Wichmann, M. Schwager, and F. Jeltsch. 2004. Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography* 31:79–92.

Thompson, G. G., and P. C. Withers. 2003. Effect of species richness and relative abundance on the shape of the species accumulation curve. *Austral Ecology* 28:355–360.

Thompson, G. G., P. C. Withers, E. R. Pianka, and S. A. Thompson. 2003a. Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. Austral Ecology 28:361–383.

Thompson, G. G., P. C. Withers, E. R. Pianka, and S. A.Thompson. 2003b. Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. *Austral Ecology* 28:361–383.

Thuiller, W. 2003. BIOMOD: Optimising predictions of species distributions and projecting potential future shift under global change. *Global Change Biology* 9:1353–1362.

Thuiller, W., M. B. Araújo, and S. Lavorel. 2004. Do we need land-cover data to model species distributions in Europe ? *Journal of Biogeography* 31:353–361.

Tichý, L., and M. Chytrý. 2006. Statistical determination of diagnostic species for site groups of unequal size. *Journal of Vegetation Science* 17:809–818.

Tjørve, E. 2003.Shapes and functions of species – area curves : a review of possible models. *Journal of Biogeography* 30:827–835.

Trainor, A. M., O. J. Schmitz, J. S. Ivan, and T. M. Shenk. 2014. Enhancing species distribution modeling by characterizing predator-prey interactions. *Ecological Applications* 24:204–216.

Ugland, K. I., J. S. Gray, and K. E. Ellingsen. 2003. The species-accumulation curve and estimation of species richness. *Journal of Animal Ecology* 72:888–897.

Ulrich, W., M. Almeida-Neto, and N. J. Gotelli. 2009. A consumer's guide to nestedness analysis. *Oikos* 118:3–17.

Ulrich, W., and J. Buszko. 2007. Sampling design and the shape of species–area curves on the regional scale. *Acta Oecologica* 31:54–59.

Ulrich, W., andN. J. Gotelli. 2007. Null model analysis of species nestedness patterns. *Ecology* 88:1824–31.

Ulrich, W., and M. Zalewski. 2007. Are ground beetles neutral? *Basic and Applied Ecology* 8:411–420.

Václavík, T., J. A. Kupfer, and R. K. Meentemeyer. 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). *Journal of Biogeography* 39:42–55.

Varela-Cervero, S., M. Vasar, J. Davison, J. M. Barea, M. Öpik, and C. Azcón-Aguilar. 2015. The composition of arbuscular mycorrhizal fungal communities differs among the roots, spores and extraradical mycelia associated with five Mediterranean plant species. *Environmental Microbiology* 17:2882–2895.

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. Fourth Edi. Springer, New York.

Voris, H. K. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* 27:1153–1167.

Voytek, J. B., and B. Voytek. 2012. Automated cognome construction and semi-automated hypothesis generation. *Journal of Neuroscience Methods* 208:92–100.

Waddle, J. H., R. M. Dorazio, S. C. Walls, K. G. Rice, J. Beauchamp, M. J. Schuman, and F. J. Mazzotti. 2010. A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications* 20:1467–75.

Walther, B. A., and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28:815–829.

Ward, D. F. 2006. Modelling the potential geographic distribution of invasive ant species in New Zealand. *Biological Invasions* 9:723–735.

Watt, A. S. 1947. Pattern and process in plant community. *Journal of Ecology* 35:1–22.

Watt, M. S., J. K. Stone, I. A. Hood, and L. K. Manning. 2011. Using a climatic niche model to predict the direct and indirect impacts of climate change on the distribution of Douglas-fir in New Zealand. *Global Change Biology* 17:3608–3619.

Webb, N. R. 1989. Studies on the invertebrate fauna of fragmented heathland in Dorset, UK, and the implications for conservation. *Biological Conservation* 47:153–165.

Wethered, R., and M. Lawes. 2005. Nestedness of bird assemblages in fragmented Afromontane forest: the effect of plantation forestry in the matrix. *Biological Conservation* 123:125–137.

White, D. R., and F. Harary. 2001. The cohesiveness of blocks in social networks: Node connectivity and conditional density. *Sociological Methodology* 31:305–359.

Whitehead, H. 1995. Investigating structure and temporal scale in social organizations using identified individuals. *Behavioral Ecology* 6:199–208.

Whittaker, R. 1975. Communities and Ecosystems. 2nd edition. Macmillan, New York.

Wiegand, T., F. Jeltsch, I. Hanski, and V. Grimm. 2003. Using pattern-oriented modeling for revealing hidden information : a key for reconciling ecological theory and application *Oikos* 65:209–222.

Willerslev, E., J. Davison, M. Moora, M. Zobel, E. Coissac, M. E. Edwards, E. D. Lorenzen, M. Vestergard, G. Gussarova, J. Haile, J. Craine, L. Gielly, S. Boessenkool, L. S. Epp, P. B. Pearman, R. Cheddadi, D. Murray, K. A. Brathen, N. Yoccoz, H. Binney, C. Cruaud, P. Wincker, T. Goslar, I. G. Alsos, E. Bellemain, A. K. Brysting, R. Elven, J. H. Sonstebo, J. Murton, A. Sher, M. Rasmussen, R. Ronn, T. Mourier, A. Cooper, J. Austin, P. Moller,

D. Froese, G. Zazula, F. Pompanon, D. Rioux, V. Niderkorn, A. Tikhonov, G. Savvinov, R. G. Roberts, R. D. E. MacPhee, M. T. P. Gilbert, K. H. Kjaer, L. Orlando, C. Brochmann, and P. Taberlet. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506:47–51.

Williams, C. B. 1943. Area and number of species. *Nature* 152:264–267.

Willis, K. J. 2002. Species Diversity--Scale Matters. *Science* 295:1245–1248.

Wilson, R. J. 1996. Introduction to Graph Theory. Longman, Harlow.

Woodcock, B. A., J. Redhead, A. J. Vanbergen, L. Hulmes, S. Hulmes, J. Peyton, M. Nowakowski, R. F. Pywell, and M. S. Heard. 2010. Impact of habitat type and landscape structure on biomass, species richness and functional diversity of ground beetles. *Agriculture, Ecosystems and Environment* 139:181–186.

Wright, D. H. 1998.A comparative analysis of nested subset patterns of species composition. *Oecologia* 113:1–20.

Wright, D. H., and J. H. Reeves. 1992. On the meaning and measurement of nestedness of species assemblages. *Oecologia* 92:416–428.

Yackulic, C. B., J. Reid, J. D. Nichols, J. E. Hines, R. Davis, and E. Forsman. 2014. The roles of competition and habitat in the dynamics of populations and species distributions. *Ecology* 95:265–279.

Yiu, M. L., and N. Mamoulis. 2004. Clustering Objects on a Spatial Network. Pages 443–454. Proceedings of the 2004 ACM SIGMOD international conference on Management of data. ACM.

Yuan, S., P. N. Tan, K. S. Cheruvelil, S. M. Collins, and P. A. Soranno. 2015. Constrained spectral clustering for regionalization: Exploring the trade-off between spatial contiguity and landscape homogeneity. Page 1-10. Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Paris, France.

Zhang, H., M. Korayem, E. You, and D. J. Crandall. 2012. Beyond co-occurrence: discovering and visualizing tag relationships from geo-spatial and temporal similarities. Pages 33–42. Proceedings of the fifth ACM international conference on Web search and data mining. ACM.

Zimmermann, N. E., T. C. Edwards, C. H. Graham, P. B. Pearman, and J. C. Svenning. 2010. New trends in species distribution modelling. *Ecography* 33:985–989.