Using statistical modelling to link disparate sources of available information to study factors that influence bird distributions

Christos Mammides

July 2012

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Imperial College London in the Department of Biological Sciences, Faculty of Life Sciences
DECLARATION OF ORIGINALITY

The work presented in this thesis is my own. To develop and parameterize the path models in Chapters 2 and 3 I used data from various sources and feedback from experts in the field, all appropriately acknowledged and referenced.
ABSTRACT

The aim of this project is to develop a method to successfully link together various sources of disparate information and data in order to study and understand which factors influence bird communities in different areas. For many areas, especially those of high conservation priority, an enormous amount of information and data may already be available. By using the right tools they could potentially be linked to provide improved understanding of the mechanisms that influence the biodiversity in a region of interest.

To test this hypothesis I used structural equation modelling (SEM) to link data from Kakamega Forest in order to study the effects of various socio-ecological factors on the bird species richness in twenty-two different parts of the forest. Kakamega Forest is Kenya's only rainforest and despite its rich biodiversity, and especially avian diversity, the forest is highly threatened with less than half of its area containing indigenous vegetation.

A similar modelling approach is used to study the factors that influence migratory birds found in thirty-eight Natura 2000 sites in Cyprus, designated under the EU’s Habitats Directive (officially known as the Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora). Using GIS analysis and available data on land use, vegetation cover, habitat diversities, human population densities, and road densities, I developed a path model explaining the observed bird species richness in those areas.

Through this type of analysis I identified and quantified the impact of various habitat variables on bird species richness, which is one of the many measures of species diversity. Compared to other diversity indices, species richness is not influenced by species abundance and therefore it is an appropriate measure for studying the distribution of the species. This type of analysis however, does not allow us to identify which species in particular are impacted by the variables identified so I used generalized linear mixed modelling to study interactions between habitat variables and certain species specific behavioural, morphological, and life history.
characteristics (including food choice, body length and clutch size), to examine how abundances and presence/absence are influenced by those variables in each site.

The results are of significant conservation importance as they give us valuable insight on: a) which factors are most important in determining species richness and b) what species characteristics make birds more vulnerable to change in these factors. Moreover, the results demonstrate that by using an appropriate statistical method there is potential to successfully utilize the enormous amount of available information to derive important conservation conclusions.
ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of my supervisor, Professor Tim Coulson, who I deeply thank for his attentive guidance throughout these years and his remarkably prompt responses to all my enquiries no matter how naïve they sometimes sounded. I am grateful to the BIOTA East Africa consortium for supporting this work and especially to Professor Katrin Böhning-Gaese, Dr. Matthias Schleuning and Professor Gertrud Schaab for their willingness to share their valuable data with me and provide feedback whenever needed for developing the path model for Kakamega Forest. I am also thankful to Jackson Welesi and Joash Ogutu for their assistance in the field; without their help not only I would not have collected the data I needed, but possibly I would still be wondering in the dense forest trying to find my way back to the research station.

Many people have helped with the development of the model for the Natura 2000 sites in Cyprus. I would like to thank Dr. Costas Kadis and Costas Kounnamas of Frederick Research Centre, and Martin Hellicar and Michael Miltiadou of BirdLife Cyprus, for their constant logistic support. I thank the Department of Forests in Cyprus for the exemplary collaboration and especially Nikos Siamarias for preparing and providing all the requested GIS data. Special thanks to all the experts in Cyprus for taking the time to attend the meetings organized and for providing me with valuable feedback on the developed path model. A complete list with their names is found in Appendix II of this thesis.

I would also like to thank all the Silwoodians I have met during these four years, who have contributed to this project possibly more than they actually imagine. Although they are too many to name, I will definitely remember them all in the future years. Particular thanks to Dr. Cristina Banks-Leite for her generous assistance with statistics and to Dr. Navinder Singh for reviewing this thesis and offering constructive suggestions.

This project was funded by the A.G Leventis Foundation and the Research Promotion Foundation in Cyprus (Penek/Support/0308/42).
# TABLE OF CONTENTS

DECLARATION OF ORIGINALITY .................................................................................. 2

ABSTRACT ..................................................................................................................... 3

ACKNOWLEDGEMENTS ............................................................................................... 5

TABLE OF CONTENTS ................................................................................................. 6

LIST OF TABLES ........................................................................................................... 9

LIST OF FIGURES ......................................................................................................... 11

CHAPTER 1 .................................................................................................................. 13

INTRODUCTION ........................................................................................................... 13

1.1 THE ISSUE ........................................................................................................... 14

1.2 THE METHOD ..................................................................................................... 16

1.2.1 Background information and description .................................................... 16

1.2.2 Steps involved in SEM analysis ..................................................................... 17

1.3. ADVANTAGES AND LIMITATIONS OF SEM .................................................. 19

1.4 SELECTED STUDIES ILLUSTRATING THE ADVANTAGES OF SEM .............. 22

1.5 USING SEM TO LINK DISPARATE SOURCES OF DATA .................................. 26

1.6 A HYPOTHETICAL EXAMPLE ........................................................................... 27

1.7 THESIS OUTLINE ............................................................................................... 31

CHAPTER 2 .................................................................................................................. 33

PARAMETERIZING THE MODEL FOR KAKAMEGA FOREST ................................. 33

2.1 ABSTRACT ........................................................................................................... 34

2.2 INTRODUCTION .................................................................................................. 34

2.3 METHODS ......................................................................................................... 35

2.3.1 Study Area ..................................................................................................... 35
LIST OF TABLES

Table 1.1 Pearson correlation coefficients between the variables examined in the three hypothetical studies ........................................................................................................................................................................... 29
Table 1.2 The covariance matrix of the variables included in hypothetical example ..... 30
Table 2.1 List of variables included in the path model developed based on the available literature on Kakamega Forest .......................................................................................................................................................................................... 40
Table 2.2 Number of species in each bird guild examined ........................................ 46
Table 2.3 Basic descriptive statistics for the continuous variables included in the SEM analysis ................................................................................................................................................................................................................. 48
Table 2.4 The resulting path coefficients, based on the final model, for the three different measures of the bird communities in Kakamega Forest ................................................................. 52
Table 2.5 The R-square values, for the different bird categories examined, showing the percentage of explained variance by the final path model .......................................................................................................................... 52
Table 2.6 The resulting path coefficients, based on the final model for the forest specialists, forest generalists and forest visitors ........................................................................................................................................... 53
Table 2.7 The resulting path coefficients, based on the final model, for frugivores and insectivores .................................................................................................................................................................................. 54
Table 2.8 Path coefficients, based on the final model, showing the cumulative indirect effects of habitat disturbance on all the bird guilds examined .............................................................................. 54
Table 2.9 Results of multiple regression analyses (standardized beta coefficients) for the overall bird species diversity, richness and abundance .............................................................................. 56
Table 3.1 Basic descriptive statistics for each species richness category in the Natura 2000 sites ................................................................................................................................................................................................................. 70
Table 3.2 Results of the goodness of fit tests for the experts’ model ......................... 72
Table 3.3 P-values of each path coefficient for each species richness category tested using the refined model ........................................................................................................................................................................... 72
Table 3.4 Results of the goodness of fit tests for the final model ................................ 73
Table 3.5 The resulting beta coefficients, estimated through multiple regression analyses, for each bird category ........................................................................................................................................................................ 75
Table 4.1 Results of the GLMM analysis showing the variance explained by the random effects .................................................................................................................................................................................................................... 85
Table 4.2 Results of the GLMM analysis showing the importance of the independent variables and their interactions on the presence/absence of species .............................................................................. 85
Table 5.1 Number of species in each forest guild used in the analysis ....................... 96
Table 5.2 Variance explained by the random effects in the presence/absence model ... 98
Table 5.3 Results showing the importance of the depended variables and their interactions on presence/absence of the species .................................................................................................................... 99
Table 5.4 Variance explained by the random effects in the abundance model .......... 101
Table 5.5 Results showing the importance of the dependent variables examined and their interactions on the abundances of the species .................................................. 103

Table A.1 Results of the multiple regression analyses, beta coefficients, for FS, FG and FV using the five equations described in Chapter 3.................................................. 131

Table A.2 Results of the multiple regression analyses, beta coefficients, for frugivores and insectivores using the five equations described in Chapter 3........................................... 131

Table A.3 Variance explained by the random effects ................................................................. 134

Table A.4 Parameter estimates for Presence/Absence Model, using Dataset 1............ 134
LIST OF FIGURES

Figure 1.1 The path diagram and coefficients of the analysis based on the hypothetical example (values in parentheses represent the R^2 for the endogenous variables)........30
Figure 2.1 Map of Kakamega Forest (Courtesy of Professor Gertrud Schaab, BIOTA E02).................................................................36
Figure 2.2 The path diagram developed based on the available literature on Kakamega Forest.................................................................40
Figure 2.3 Map of Kakamega Forest showing the 22 study sites from which data were collected to parameterize the path model (Dataset 1: Farwig et al. 2006b; Dataset 2: Farwig et al. 2008c)........................................................................................................41
Figure 2.4 The path model for Kakamega Forest developed based on experts’ opinion ........................................................................................................43
Figure 2.5 The resulting path diagram, after the “Human Population Densities” and the “Management Type” variables were excluded, showing the corresponding path coefficients and the R^2 values of each endogenous variable (top right corner)..........49
Figure 2.6 The results of the SEM analysis when the final and more parsimonious model was tested; the path coefficients and the R^2 values (top right corner) of each endogenous variable are also displayed........................................................................49
Figure 3.1 Map of eastern Mediterranean Sea showing the location of Cyprus (Data source: ESRI)........................................................................................................64
Figure 3.2 Path model for the Natura 2000 sites developed based on available literature ................................................................................................................................................64
Figure 3.3 Path model for the Natura 2000 sites developed based on experts’ opinion ........................................................................................................66
Figure 3.4 Map of the SACs (Natura 2000 sites) in Cyprus (shown in green)..............67
Figure 3.5 SEM analysis results based on the refined model showing the corresponding path coefficients and the R^2 square values of the endogenous variables (top right corner) ........................................................................................................71
Figure 3.6 Results of the SEM analyses for each species richness category based on the final model, showing also the corresponding path coefficients and the R^2 for the endogenous variables........................................................................................................74
Figure 4.1 Graphs depicting the interaction effects between the three habitat variables and between the number of EU(25) pairs and habitat breadth.........................................................86
Figure 4.2 Graphs depicting the two-way significant interactions between the habitat variables and the species traits included in the analysis.................................................................87
Figure 5.1 Graphs depicting the interaction effects between the forest guilds and habitat variables examined. .............................................................................................................................100
Figure 5.2 Graph depicting the interaction effect of clutch size with foliage height diversity and distance to the forest edge on the probability of presence of the species .............................................................................................................................101
Figure 5.3 Graphs depicting the interaction effects between the species traits and the habitat variables on bird abundances
CHAPTER 1

INTRODUCTION
1.1 THE ISSUE

To be able to address today’s rates of biodiversity and habitat loss, we need, among other things, a good understanding of the underlying processes that cause loss. Although in some cases the reasons might be obvious, in many others the loss is a result of complex interactions which are not easy to disentangle. This may require long-term studies and large amounts of data which are not always available for a number of reasons, including limited resources (Hobbie et al. 2003, Ferraro and Pattanayak 2006). As a result we often end up with short-term, narrowly focused studies (Michener et al. 1997, Cook et al. 2001), which can only provide a glimpse of the bigger picture. The glimpse is certainly of importance on its own but probably of greater importance when appropriately linked with other related studies, and could be used to gain understanding for complex systems and relationships for which large amount of information might be needed (Furlow and Beretvas 2005).

Linking studies together either qualitatively or quantitatively, through a structured literature review or meta-analytic methods, often leads to new important insights derived from the original results and conclusions (Arnaqvist and Wooster 1995). For instance, Fernandez-Duque and Valeggia (1994), in their paper in which they explain the importance of meta-analysis, provide an example in which by using five published studies they show that the strong negative effects of logging on bird diversity became evident through meta-analysis even though the results of the individual studies differed. The benefit from utilizing and synthesizing the large amounts of existing information is that we can learn more about a particular system and reach conclusions which in some cases might not be obvious from the disparate information already available.

However, the success in quantitatively linking the studies is a function of the information reported, which is often limited (Whitlock 2011). For example, Bender et al. (1998) reviewed more than 200 studies that examined the impact of patch size on various animals but could only use 25 of them in their meta-analysis because, among other reasons, many studies did not include enough information to calculate the required statistics. In some cases the issues associated with the lack of information can be overcome with access to the raw data used in the original studies.
Unfortunately though, once the data for those studies are collected, analysed and the results published, they are usually not made available for other scientists to use (Parr and Cummings 2005, Costello 2009, Reichman et al. 2011, Whitlock 2011) and end up being eventually lost (Michener et al. 1997, Michener 2006).

Yet, conservation biology and ecology are very data demanding fields and scientists increasingly rely on larger datasets and the integration of small scale studies to conduct complex analyses to answer today’s questions (Madin et al. 2007). The cumulative amount of data and information collected across all studies is enormous. Moreover, large amounts of useful data are also available in grey literature, unpublished reports and databases of various governmental and non-governmental organizations. All this information could be of great utility with important conservation implications if appropriately linked together using the right tool.

The goal of this project is to test whether path analysis and Structural Equation Modelling (SEM) can be used to utilize the dissimilar sources of disparate information and data available to study the socio-ecological factors that influence the distribution of birds in two different regions: a) Kakamega Forest, a tropical rainforest in western Kenya, which has been well-studied with many published studies available and b) the Natura 2000 sites in Cyprus, a network of protected areas designated under the EU’s Habitats Directive, for which, although understudied, there are large amounts of unpublished relevant data. The first step of the approach is to create a path model for each study area by reviewing the current literature and identifying the most important variables that influence the species richness of the birds in those regions. Using the reported significant relationships between the variables of interest, two models are then created that illustrate the corresponding underlying ecological mechanisms. The models are then parameterized to extract important conservation conclusions.

A similar approach, called meta-analytic structural equation modelling (MASEM), has been implemented by social scientists where correlation values from multiple studies were used to estimate the effect size for each relationship studied and then pooled together to create a correlation matrix, which was analyzed using structural equation techniques to answer more complex questions than those addressed by the
individual studies (Cheung and Chan 2005, Furlow and Beretvas 2005). In this case the use of meta-analytic approach to calculate the effect size is not appropriate as the relationships used to develop the model have typically been reported in single studies only.

1.2 THE METHOD

1.2.1 Background information and description

Path analysis and SEM are multivariate statistical tools that can be used to study the effects of several independent variables on one or more response variable and also on one another (Schumacker and Lomax 2004, Grace 2006). Path analysis was developed by the population biologist Sewall Wright in the early 1920s (Wright 1921) to study “systems of multiple causality” (Petraitis et al. 1996, Almaraz 2005) using observational data (Mitchell 1992). Structural Equation Modelling (SEM) is considered to be the modern form of path analysis and is an extension of it that was developed later to include latent variables (Almaraz 2005, Grace 2006). Latent variables are factors that cannot be measured directly in the field or the lab and are therefore represented in the path model by one or more measurable indicator variables (Almaraz 2005).

Although the method was developed by a biologist it was only sporadically used in biology in the decades following its derivation. During this period it was used more frequently by social scientists who contributed substantially to its development (Grace 2008). Sociologists and econometricians have been using path analysis as an accepted statistical tool since the fifties while biologists started using it a little more frequently in the seventies (Almaraz 2005). The current form of SEM is a result of the LISREL (LInear Structural RELations) system developed by the Swedish mathematician Jöreskog (Grace 2006). Jöreskog and Sörbom are often referred to as the most noteworthy developers of this method (Malaeb et al. 2000, Iriondo et al. 2003, Grace 2006). Like path analysis, SEM has been also used widely in several research disciplines but its application in ecology has been recent and rather limited (Arhonditsis et al. 2006).
Path analysis and SEM can be thought of as a series of linear multiple equations used to explain the functioning of certain aspects of a system (Mitchell 1992, Wootton 1994), by testing an a priori hypothesized path model or be creating an exploratory model for the same purpose (Malaeb et al. 2000, Iriondo et al. 2003). The path model is essentially a hypothesized description of the casual relations of the variables included in the model (Iriondo et al. 2003) which can be used to study and understand the effect of each indirect and direct variable included in a path model and to also evaluate their impact on one or more response variables of interest (LeClair 1981, Rompre et al. 2007). The relative effect of each variable on the other is expressed with path coefficients (Iriondo et al. 2003). The path coefficients usually represent the standardized partial regression coefficient (beta weights) and they equal the effects of one variable on another when the rest are “held constant” (Mitchell 1992).

The causal model developed and tested through various appropriate tests of fit should be supported by a strong theoretical background developed through a synthesis of the previous studies and/or through scientific expertise (Grace 2006). A common test used for model fit is chi-square analysis (using the maximum likelihood methods) which assesses the similarity between the observed and the model covariance matrices (Marcot 2006). Assuming that the model fits the data, the researcher can conclude that the mechanism described is a plausible explanation of the observed variation (Grace 2006). It is important to acknowledge though, that a satisfactory fit of the model does not necessarily mean that the causal relationships included are also proven (Laughlin et al. 2007).

1.2.2 Steps involved in SEM analysis

The first step in SEM is the development of a causal model, usually based on available theory and/or expert knowledge. This information allows the researcher to hypothesize about potential direct and indirect relations between the endogenous and exogenous variables included in the model. Endogenous variables are variables that are explained by factors included in the model while exogenous ones are explained by other factors not included (Schumacker and Lomax 2004). The path diagram graphically represents the possible causal relations between the variables, and it also includes the residual errors for all dependent variables representing the unexplained
variation not accounted for by the exogenous variables (Schumacker and Lomax 2004).

Once the causal model and the path diagram have been developed, the relationships between the variables included are established by calculating the corresponding correlation or covariance matrix based on the available data (Schumacker and Lomax 2004). The matrix is then used to calculate the path coefficients which can be standardized or not. When the standardized coefficients are used they can be interpreted as estimates of how much variability in one variable can be explained by the variables that have a direct impact on it (Grace 2006). The standardized coefficients are calculated using multiple regression analysis and they represent the beta values (i.e. the multiple-regression coefficients) (Grace 2006). Standardized coefficients are usually preferred because they allow the comparison of variables in similar units (Grace 2008).

Once the coefficients have been calculated the researcher tests for the goodness of fit, i.e. how good the model fits the data. Although when originally developed path analysis did not include methods for testing model fit, such tools were developed later (Grace and Pugesek 1998). In this step the researcher tests whether the observed and estimated variance of the model are statistically similar. Several tests can be used to assess the goodness of fit such as the chi-square test, using maximum likelihood method, which minimizes the observed and estimated covariances (Mitchell 1992, Arhonditsis et al. 2006). It is important to note that maximum likelihood methods assume that data are “unbiased, consistent, efficient” and that “error distributions are stationary” (Almaraz 2005). For the cases where these assumptions are not met, and thus maximum likelihood is inappropriate, other methods can be used such as generalized least squares and weighted least squares (Almaraz 2005).

The final step of the analysis involves the interpretation of the results and the assessment of the validity of the model based on those results. It is important to note that a good fit does not necessarily mean the model is accurate, as several other models may also fit the data equally well (Grace 2006). Also, significant path coefficients do not prove causation. Depending on whether the researcher is using SEM analysis for exploratory or confirmatory purposes, he or she can use the
statistical results to refine and retest the model in order to improve model fit. Any re-
specification must not only be based on the results of the model but on a scientifically
sound theoretical background as well.

1.3. ADVANTAGES AND LIMITATIONS OF SEM

One of the main advantages of SEM is that it allows researchers to study both the
direct and indirect effects of predictor variables and also to compare their relative
importance using simultaneous regression equations (Malaeb et al. 2000). Unlike the
conventional multiple regression models, SEM can be used to simultaneously “test the
indirect effects of two explanatory variables mediated by another intermediary
variable” as part of the model (Arhonditsis et al. 2006). The ability to incorporate
indirect links between variables allows researchers to identify important relationships
that would be difficult to identify in other ways (examples of such cases are presented
in the selected studies section below), thus leading to a deeper understanding of the
systems studied. Also, along the same lines, through SEM we can compare alternative
models describing the same system and make useful inferences about missing
relationships that may need to be incorporated in future models (Mitchell 1992).

Another advantage of SEM is that unlike other linear models (where regression
analysis and analysis of variance are used) it does not assume that only the dependent
variable is subject to measurement error (Malaeb et al. 2000, Iriondo et al. 2003).
SEM takes into account that all variables are influenced by measurement error and
this is especially useful for observational studies where sources of error and
uncontrolled variation cannot be necessarily eliminated (Malaeb et al. 2000, Iriondo et
al. 2003). By taking into account measurement errors and their correlations SEM
gives a more realistic estimate of the parameters of the model compared to multiple
regression (Malaeb et al. 2000).

A third advantage is that traditional regression methods usually rely on the $R^2$
values to assess the amount of variation explained by the regression model. High $R^2$
values are interpreted by researchers as a satisfactory fit of the regression model to the data,
and usually the analysis concludes at that point (Mitchell 1992). However, in SEM a
model with some high $R^2$ values and significant path coefficients may also have an
unsatisfactory goodness of fit (because what is evaluated is the whole model) thus pointing to the need for further development of the model (Mitchell 1992).

SEM in its modern form can also be used to test the effects of variables that cannot be observed or measured (defined as latent variables) through the use of measurable indicators (Grace 2006). Latent variables may have important indirect and direct effects on the variables studied so their inclusion in the model is sometimes necessary (Grace 2006).

In general, SEM can be a very useful and informative tool when used properly and when the assumptions are met. Like other statistical tools, it is sensitive to a number of issues that may result from inappropriate evaluation and use of data (Grace 2006). Some of the basic assumptions of SEM are that the relationships between the variables are linear and additive, the data fit a random distribution, the sample size is large enough (LeClair 1981, Shipley 2000, Austin 2007) and the residuals of the variables are not correlated (LeClair 1981). Nevertheless, SEM is a continuously evolving method and strategies are being developed to deal with some of its limitations such as non-normality, non-linearity, collinearity, non independent samples, non-additivity, use of categorical variables and small sample sizes (Grace and Pugesek 1998, Grace 2006).

Despite its usefulness and potential, authors have raised concerns about the fact that because of its advanced nature and complexity the method can be used incorrectly and therefore lead to misleading conclusions (Petraitis et al. 1996, Smith et al. 1997). Petraitis et al. (1996) reviewed 26 studies in which 64 path analyses had been conducted by ecologists and they discovered that in many cases the method had been used wrongly. Some of the problems they came across were inadequate sample sizes, incorrect use of categorical data, lack of distinction between confirmatory and exploratory approaches, and problems with collinearity.

As in multiple regression analysis, collinearity, i.e. when two or more of the predictor variables are highly correlated can be problematic (Wootton 1994, Petraitis et al. 1996). In particular, collinearity “increases the standard errors of estimated partial regression coefficients” and thus results in inability to identify statistical significance
between estimated and observed variance leading to the acceptance of the model even when that might be wrong (Petraitis et al. 1996). However, collinearity can be relatively easily identified through pairwise calculations. Researchers usually address the problem of collinearity by dropping one of the highly correlated variables that they think is of lesser importance (e.g. may have a smaller path coefficient) but this is incorrect because it affects the rest of the predictors included in the model (Petraitis et al. 1996). It is also incorrect because the reason for choosing one variable over another is more driven by the data and the analysis and not necessarily by theoretical justification.

One should understand that the choice of casual links affects the strength of path coefficients (Petraitis et al. 1996). The omission of variables (either by deleting them or not incorporating them in the first place) can result in misleading conclusions on the validity of the model and the importance of variables included (Grace 2008). For example, a poor fit of the model does not necessarily mean that the mechanisms suggested are wrong but that they may be incomplete. For that reason, it is important to examine alternative models, even if a statistically satisfactory model has been identified, because we may find that other models also fit the data (Shipley 1997, Grace 2008).

Another issue associated with the use of path analysis and SEM is inappropriately distinguishing between “formal hypothesis testing” and “exploratory data analysis” (Petraitis et al. 1996). Hypothesis testing involves the development of an a priori model that is tested to examine whether it fits the data satisfactorily or not. Some authors even make the distinction between a “strictly confirmatory” hypothesis testing, in which a single model is developed and tested using the data and an “alternative models” hypothesis testing, in which a number of models are tested against a single set of data (Arhonditsis et al. 2006). Exploratory data analysis involves the development of an a posteriori model, though the modification of the original model tested (that is by adding and removing causal links) only after the data have been analysed and the results of the goodness of fit have been considered.

It is important to distinguish between the two approaches because the former is more driven by previous knowledge about the system studied while the latter is driven by
the statistical results of the analysis of the specific data used to test the model. Post-hoc models seem to be more common in ecological studies using path analysis (Petraitis et al. 1996) and it is likely that even in formal hypothesis testing researchers will modify the model to fit the data better. Some software packages used for SEM analysis (e.g. AMOS) calculate the “modification indices” which are values estimating how much the model fit can be improved through re-specification of the model. These values are of course determined by the data and not by the theory (Schumacker and Lomax 2004, Grace 2006) so they should be used with caution.

Shipley (1997) states that some researchers argue that exploratory analysis using SEM should be discouraged because it can lead to misleading results. On the other hand, exploratory analysis can be very useful if treated appropriately because biologists often do not have all the information necessary to develop an accurate model in advance (Shipley 1997). However, once the model is extensively refined and retested it should not be presented as an a priori model and ideally it should be tested using new data (Malaeb et al. 2000).

Other major problems associated with SEM result from the use of inappropriate samples (Petraitis et al. 1996, Schumacker and Lomax 2004). For example if the sample is not representative, for reasons such as incorrect sampling design or small sample size, any conclusions based on the model may not be applicable to the system studied. Petraitis et al. (1996) found that in most of the studies they reviewed the sample size was less than that required, which ideally should be five to twenty times the number of parameters that need to be estimated (Petraitis et al. 1996, Almaraz 2005, Grace 2008).

1.4 SELECTED STUDIES ILLUSTRATING THE ADVANTAGES OF SEM

In recent years ecologists have begun to appreciate the advantages and potential capabilities of path analysis and SEM, and the methods are increasingly used for conservation purposes (Iriondo et al. 2003, Grace 2006, Austin 2007) and to address issues at different scales ranging from local to national to global. Although they are usually used for observational studies (Marcot 2006) they have also been used for
experimental studies (Almaraz 2005). Below are some examples of how the methods have been applied in ecology so far.

Forester and Machlist (1996) used structural equation modelling to study the anthropogenic factors that influence biodiversity loss at the global level. By collecting available data from 107 countries they tested the effects of several socio-ecological variables on biodiversity loss and showed that human population densities and government policies have a particularly strong impact on it. The authors argue that although biodiversity loss can be usually linked and explained by biological factors, often the underlying causes are human actions which can be described using social, economic, and political factors, stressing therefore the importance of using an interdisciplinary approach to study biodiversity loss.

Maleab et al. (2000) showed the importance of SEM analysis in studying the indirect and direct effects of biological variables that cannot be directly measured (i.e. latent variables), to understand how they influence each other. The tested the impact of sediment contamination and natural variability on growth potential and biodiversity using eleven observed environmental variables collected from the US Environmental Protection Agency. Interestingly, the results showed that the impact of each variable was different when indirect links among them were considered. For example natural variability had an overall positive effect on growth potential, although the direct effect was negative, due to its strong positive effect on biodiversity. This study illustrates the potential of SEM to reveal relations between variables of interest that can not be identified through simple univariate or multivariate analyses that only examine the direct effects.

Laughlin et al. (2007) reached to similar conclusions when they used SEM to study the impact of soil composition on the understory plant richness in the Pinus ponderosa forests in the Southwestern United States. They showed that SEM can be used to study more effectively complex system interactions that cannot be analyzed using traditional statistical approaches. They demonstrated this by first examining the relations between the variables studied using bivariate tests which failed to reveal the significant relationships shown through SEM analysis.
Rompre et al. (2007), Kissling et al. (2007) and Kissling et al. (2008) used path analysis to examine the impacts of several environmental variables on bird species richness in lowland Panama rainforests, sub-Saharan Africa and Kenya respectively. By developing path and structural equation models they identified the factors that influence the species richness in those areas, both directly and indirectly, and assessed the relative importance of each variable included. Similarly, Becker et al. (2007) used SEM analysis to examine the habitat variables that influence the species richness of amphibians in twelve Brazilian Atlantic Forests. They found that habitat split (defined by the authors as: “human induced disconnections between habitats used by different life history stages of a species”) affected negatively the species with aquatic larvae but not those with “terrestrial development”; explaining thus the observed population decrease of the former and showing the conservation importance of restoring and protecting riparian vegetation.

In another example, Grace and Pugesek (1997) used SEM analysis to study the effects of three latent variables (disturbance, abiotic conditions, and plant biomass), represented by multiple indicators, on the species richness of plants in the lower Pearl River coastal wetlands in Louisiana, USA. Abiotic variables were found to be influencing plant richness, both indirectly and directly, while the effects of disturbance were indirect, through its impact on plant biomass. Interestingly, Weiher (2003) examined whether the Grace and Pugesek (1997) model could be confirmed in other systems by using data from oak savannas located in the floodplain of the Chippewa River in Wisconsin. The author concluded that the Grace and Pugesek (1997) model was not applicable mainly due to the lack of direct effects of disturbance on species richness. However, they showed that SEM can be useful in studying complex interactions because it revealed that the soil quality in the study area had a direct and opposite indirect effect on richness, a fact that could not be identified through the standard correlation or multiple regression methods.

Not all studies reviewed though were in favour of using path and SEM analyses. Smith et al. (1997) evaluated path analysis and its effectiveness in identifying and quantifying already known direct and indirect interactions in a natural system using experimental and observational data. Their model examined the impact of Kangaroo rats on two small granivorous rodent species, through competition and through the
effects on grass cover, in the Chihuahuan Desert in Arizona. For their analysis they used different data sets which included both “manipulated data” (where Kangaroo rats were excluded from the system) and “unmanipulated data”.

Their model was “just identified” (meaning that the number of known parameters was equal to the number of unknown) so it was difficult to evaluate its statistical significance because there were no degrees of freedom to conduct the appropriate goodness of fit tests (e.g. $\chi^2$) (Grace and Pugesek 1998). However, they were able to calculate the path coefficients and were surprised to find that the estimated values varied a lot depending on which dataset was used. Also the strong well-understood interaction between the rodents was not revealed when the manipulated data were excluded from the analysis. In addition, surprisingly to them, path analysis revealed an unknown strong relationship between one of the granivorous species and grass cover which was not identified in their previous studies.

The authors concluded that path analysis was ineffective because it failed to reveal the strong impact of kangaroo rats using only observational data. They argue therefore that researchers should be cautious about the results of path analysis since most of the analyses are conducted using observational data. According to the authors, in some cases experiments are necessary to reveal important interactions so it is a mistake to consider path analysis as an alternative. They also explained that the validity of the path analysis depends a lot on whether the most important factors are included in the model, but for many ecological studies it is not always possible to include all those factors. Considering all these issues they concluded that path analysis was no better than the statistical tools they used already (such as ANOVA) and in fact could even be misleading.

Grace and Pugesek (1998) reviewed the conclusions of Smith et al. (1997) and argued that the fact that path coefficients were different is not unexpected as the values of the coefficients depend on the datasets used. Therefore, it is reasonable to have different values with and without the presence of the Kangaroo species. It is important to realize that the generalizability and applicability of the results depend on whether the data used are representative of the system studied. If the model is not applicable it does not necessarily mean it is incorrect. Also the revealed unknown interaction does
not necessarily correspond to an error in the method but on the contrary, it shows that some interactions can only be identified using a multivariate approach of this type.

In conclusion, as highlighted in the selected studies, SEM is a very powerful but data demanding advanced statistical tool that if not used appropriately may result in misleading conclusions. Most problems associated with SEM have to do with incorrect model specification, errors in analysis and interpretation errors (Petraitis et al. 1996, Grace 2006, 2008). Incorrect model specification can result from poor understanding of the system studied and omission of important variables (Grace 2008). It is therefore important to consider alternative models which may also fit the data and it is also important to realize that model refinement and testing with new data may be necessary for meaningful and widely applicable conclusions (Grace 2008).

1.5 USING SEM TO LINK DISPARATE SOURCES OF DATA

Understanding complex biological interactions and revealing “hidden” relationships may require the use of advance statistical tools such as SEM. The key advantage of this method results from the fact that unlike basic univariate and multivariate methods it can simultaneously evaluate both the indirect and direct impacts of certain variables on other variables of interest. For this reason, if used appropriately, it can utilize and link disparate sources of information under one framework to give us a deeper understanding of the underlying dynamics of the system studied that otherwise would not be possible. Theoretically, it should be feasible to use the conclusions and the statistical estimates from individual studies to construct and parameterize a path model that describes the influence of the variables on each other extracted from different sources. However, for this to be possible it is necessary that certain assumptions are met in addition to those described in the limitations section, which are rooted in the statistical nature of the method:

1. The necessary information must be available either in the published literature, in an accessible data repository, or through the researchers who collected the data. This means that either the correlation values between the variables of interest or the raw data must be obtainable so that the corresponding matrix can be calculated and the model can be parameterized.
2. If the model is to be spatially valid there must be consistency in the sites from where the data have been collected. Although it is possible to estimate coefficients using correlations from different areas, when that is done the model may be inappropriate for drawing spatial conclusions.

If the required information is not available and the developed model cannot be parameterized then it is essentially only a logical description of the mechanism of how different factors are influencing the response variable of interest, based on the information extracted from the literature. In the following section I am using a hypothetical example to illustrate how SEM could be used to link together different studies and parameterize the resulting model.

1.6 A HYPOTHEtical EXAMPLE

Assume that the “Wildlife Service” is a governmental body responsible for managing the population densities of an important game bird species. The species is mainly found in grasslands and is a generalist, meaning that it feeds on a number of different plants. To effectively manage the population the wildlife service needs to understand which factors are mostly influencing its abundance. The management strategy usually followed involves the rotation of the areas available for hunting allowing thus the species to recover. For example, out of the 30 areas available a third of them are closed for 4 years and people are not allowed to hunt in those during that time period.

The hunting season is about to begin soon and the wildlife service will have to announce which areas will be closed for hunting. They realize that there have been several relevant studies in the region and they want to utilize the results of those studies to make an informed decision.

One of the studies examined the factors that influence the population densities of the birds. The authors found a strong positive correlation between population densities and plant species diversity, a variable thought to be an indicator of food availability (Table 1.1). They also found a significant positive correlation between the population densities and the size of the area.
In another study, researchers were interested in the impact that hunters may have on the biodiversity in the region. They examined the effects of the number of hunters on the plant species diversity and the population densities of birds and found strong negative impacts. The researchers were also interested in the factors that determine the number of hunters visiting an area so they studied the relationship between the number of hunters and the distance to the nearest village, which was thought to be an important factor influencing the choice of hunters on where to hunt.

In a third study, researchers interested in indentifying the factors that determine plant diversity examined the impact of the size of the area and of two abiotic factors (precipitation and nitrogen levels in the soil) on plants. Assuming all other conditions in those areas are equal, the management authority has three options for deciding which areas to close:

1. Areas that are smaller in size (possibly having lower bird population densities which are more vulnerable to hunting),
2. Areas closest to the villages (thus reducing the number of hunters in those areas) and
3. Areas with high precipitation or high nitrogen levels (to allow the plants to recover).

By linking the results of the studies together, a path diagram was developed (Figure 1.1) which illustrates the factors that influence the population densities of the birds in those areas. The corresponding path coefficients were calculated using the reported Pearson correlation values in Table 1.1. The path coefficient between the distance to the nearest village and the number of hunters is the bivariate correlation value between the two, since the number of hunters (which is an endogenous variable) is only influenced by the distance. The path coefficients explaining plant diversity, i.e. precipitation and nitrogen levels were calculated by estimating the first-order partial correlation coefficients using the formula below:
where $X$ is plant diversity, $Y$ is precipitation and $W$ is nitrogen levels; $r_{XY}$, $r_{XW}$, $r_{YW}$ are the corresponding bivariate correlation values and $r_{X\cdot W}$ is the first order partial correlation coefficient between precipitation and plant diversity when nitrogen levels are held constant.

The path coefficients for the links explaining population densities were calculated by estimating the second order partial correlation values since population densities is an endogenous variable explained by three other factors. Therefore the coefficients were calculated using the formula below by keeping constant two of the tree variables each time.

$$r_{X\cdot W\cdot O} = \frac{r_{XY} - (r_{XW} r_{YW})}{\sqrt{(1 - r^2_{XW})(1 - r^2_{YW})}}$$

where $X$ is plant diversity, $Y$ is the area, $W$ is the number of hunters and $O$ is the population densities. $r_{XY\cdot O}$, $r_{XW\cdot O}$, $r_{YW\cdot O}$ are the corresponding first-order partial coefficients and $r_{X\cdot W\cdot O}$ the second-order partial coefficient between plant diversity and area when the other two factors are held constant.

**Table 1.1 Pearson correlation coefficients between the variables examined in the three hypothetical studies**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Correlation Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Study</strong></td>
<td></td>
</tr>
<tr>
<td>Plant diversity : Population densities</td>
<td>0.794</td>
</tr>
<tr>
<td>Area : Population densities</td>
<td>0.667</td>
</tr>
<tr>
<td><strong>Second Study</strong></td>
<td></td>
</tr>
<tr>
<td>Hunters : Population densities</td>
<td>-0.831</td>
</tr>
<tr>
<td>Hunters : Distance to the nearest village</td>
<td>-0.180</td>
</tr>
<tr>
<td>Hunters : Plant diversity</td>
<td>-0.633</td>
</tr>
<tr>
<td><strong>Third Study</strong></td>
<td></td>
</tr>
<tr>
<td>Precipitation : Plant diversity</td>
<td>0.756</td>
</tr>
<tr>
<td>Nitrogen levels : Plant diversity</td>
<td>0.275</td>
</tr>
<tr>
<td>Area : Plant diversity</td>
<td>0.602</td>
</tr>
<tr>
<td>Precipitation : Nitrogen levels</td>
<td>0.188</td>
</tr>
</tbody>
</table>
Note that at the end of Table 1.1 the correlation value between precipitation and the nitrogen levels is also reported. Although this estimate was not required to answer the questions examined in the third study, it is important that it is available because otherwise the calculation of the path coefficients would not be possible.

For verification purposes, I used the “sem” package in R (Fox 2008, R Development Core Team 2007) to calculate the model’s path coefficients using the covariance matrix in Table 1.2. The matrix was calculated using the raw data available for each variable.

**Figure 1.1** The path diagram and coefficients of the analysis based on the hypothetical example (values in parentheses represent the $R^2$ for the endogenous variables)

The path coefficients estimated using the “sem” package were identical to the coefficients obtained through the correlation and partial correlation analysis indicating that with access to the corresponding correlation values, one should be able to parameterize the developed model. The numbers in parentheses in Figure 1.1, for the three endogenous variables, represent the $R^2$ values and thus describe how much variation is explained by the independent variables. The $R^2$ value for the population densities is high (0.81) showing that most of the variation is explained by the model.
The number of hunters has a strong negative impact on the population densities of the species. Also plant species diversity has a positive impact on densities while area does not affect densities much despite that the bivariate correlation between the two is high (0.667). The distance to the village does not seem to have a strong indirect impact on population densities (-0.18×-0.62=0.11). On the other hand, the indirect impact of precipitation is higher (0.38) but this is not true for nitrogen levels (0.07). When all three management options are considered, it seems more appropriate to close areas with high precipitation levels because the total effect on the population densities is stronger than the total effect of any of the other two options (despite the fact that the number of hunters has a stronger direct negative effect and the bivariate correlation value between area and densities is large).

1.7 THESIS OUTLINE

In this chapter I have presented the main statistical method (SEM) used to develop the path model for the two study areas of this project in order to identify and quantify the factors that influence the distribution of birds in those areas. In Chapter 2, I review the relevant available literature for Kakamega Forest and develop a path model by linking together the important relationships reported in the individual studies. I attempt to parameterize the model by extracting information found in the reviewed studies but the results show that the task is impracticable as several important pieces of information are missing. Therefore, I parameterize an updated version of the developed model, which was modified to include suggestions from experts working in the particular forest, using raw data obtained from researchers in the field and also necessary supplementary data I collected myself in 2009. By following an analogous methodological approach, in Chapter 3 I develop a path model which illustrates the possible mechanism influencing the species richness of birds in 38 protected areas in Cyprus. The model is modified based on the opinion of experts working on biodiversity related issues in the island and is parameterized by linking together different sources of available unpublished raw data. In Chapters 4 and 5, I use linear mixed model analyses to examine how specific species traits such as body length and clutch size are influencing the impacts of the habitat variables identified through SEM analysis, in Chapters 2 and 4. The species traits were extracted from various published sources such as relevant ornithological handbooks. The conclusions of the thesis
about: a) the usefulness of the methodological approach in linking disparate sources of information, b) the factors influencing the avifauna of the two regions and c) the importance of data availability are presented in Chapter 6.
CHAPTER 2

PARAMETERIZING THE MODEL FOR KAKAMEGA FOREST
2.1 ABSTRACT

Kakamega Forest, in western Kenya, is a highly degraded and threatened forest which nevertheless supports a rich bird community of conservation importance. The effective protection of the avifauna in the forest requires a deep understanding of the underlying ecological mechanisms influencing the birds in the area. In this Chapter, I review the literature on Kakamega Forest and develop a path model based on the reported significant relationships and on experts’ opinion, which illustrates the factors that influence the bird species richness and abundances in the forest. Results, using Structural Equation Modeling analysis (SEM), show that the birds are mostly influenced by the vegetation structure and community of the forest and also by the distance to the forest edge. Anthropogenic habitat disturbance also influences the birds indirectly by negatively affecting the forest’s vegetation.

2.2 INTRODUCTION

Many areas, especially areas of high conservation priority, are considered well studied, with numerous published papers available on different aspects of their biodiversity. For many of those areas, the available studies are disparate but the total amount of information and data available is enormous. Using the right tools, these sources of information can be brought together to give us a deeper insight into the underlying mechanisms and interactions influencing the biodiversity of the region. This deeper understanding may aid in developing appropriate management plans and conservation actions to tackle the causes of biodiversity loss.

In the first chapter of this thesis I explained how structural equation modelling (SEM) could be used as such a tool. Through SEM, a path diagram can be developed, illustrating the hypothesized causal relationships between the variables included in the literature (Grace 2006). This more inclusive and comprehensive approach can be useful for developing and implementing effective conservation measures for the protection of the rich avifauna of Kakamega Forest (Bennun and Njoroge 2000), and for stemming the negative effects of habitat disturbance and fragmentation. The forest’s biodiversity has been particularly affected by these two factors in the last few decades (Fashing et al. 2004, Lung and Schaab 2004, Farwig et al. 2006a, Farwig et
al. 2006b, Lung and Schaab 2006, Farwig et al. 2008a, Farwig et al. 2008b, Farwig et al. 2008d, Kirika et al. 2008a), mainly through commercial logging and subsistence exploitation, which have resulted into the reduction of the forest’s cover to half of its original size (Mitchell and Schaab 2008). Even more, these effects are likely to increase in the future, with potentially larger impacts on the bird communities, since the forest is found in one of the most densely populated areas of Kenya (Tsingalia 1990, Bleher et al. 2006, Lung and Schaab 2006).

The negative effects on birds will add an extra obstacle to the conservation efforts of the responsible management authorities since birds perform important ecosystem functions as dispersers and pollinators for example (Sekercioglu et al. 2004, Bleher and Bohning-Gaese 2006, Sekercioglu 2006). These functions are necessary for the persistence of the forest’s biodiversity in general so it is apparent that the protection of the avifauna of the forest, by understanding the factors that determine the bird species richness and abundances, is of vital importance. For that purpose, in this Chapter, I use available raw data and supplementary data I collected myself to parameterize the developed path model and identify the habitat variables that are influencing the bird communities in the forest.

2.3 METHODS

2.3.1 Study Area

The Kakamega Forest is located in Western Kenya and was once part of the Guineo-Congolian rainforest (Kokwaro 1988, Tsingalia 1990). Its official boundaries cover an area of about 240 km² but today less than half of that area contains indigenous forest (Bleher et al. 2006, Lung and Schaab 2006) (Figure 2.1). According to Bleher et al. (2006), the forest was first established as a protected area in 1933. Several years later, in 1967, the government of Kenya gazetted three additional areas within the forest as Nature Reserves (Isecheno, Kisere and Yala), and in 1986 two more areas in the northern part as Kakamega National Reserves (Buyangu and Kisere Forest). As a result today the forest comprises of areas with different protection status managed by two different authorities: the Forests Department in the southern part and the Kenya Wildlife Service (KWS) in the northern (Bleher et al. 2006).
The two departments follow different management strategies. The Forests Department allows local people to collect dead timber for fuel and various non-timber forest products such as mushrooms, fruits and plants for medicinal use (Althof 2005, Bleher et al. 2006, Guthiga et al. 2006, Guthiga 2007). It also grants limited permits for cattle grazing but not within the Nature Reserves (Bleher et al. 2006, Guthiga et al. 2006, Guthiga 2007). KWS on the other hand does not allow any harvesting or grazing within the forest (Bleher et al. 2006, Guthiga et al. 2006).

Figure 2.1 Map of Kakamega Forest (Courtesy of Professor Gertrud Schaab, BIOTA E02)

Despite its protection status, the forest is highly degraded and fragmented (especially in the southern part) although in some areas in the northeast the natural vegetation is
regenerating (Lung and Schaab 2004, Mitchell 2004). Scientists have documented through GIS analysis a loss of forest cover of more than 20% over the last few decades (Lung and Schaab 2004). The main cause of habitat destruction was past commercial logging which was banned in the early 1980s by presidential mandate, which prohibited the exploitation of indigenous trees (Mitchell 2004, Bleher et al. 2006).

Today, most of the disturbance results from the activities of the communities living next to the forest. The area is very densely populated with about 600 residents/km² (McGeoch et al. 2008). The majority of the inhabitants are farmers (Althof 2005, McGeoch et al. 2008) and the area is used intensely for subsistence agriculture (Guthiga 2007). The population numbers are growing (Guthiga 2007) and so does the pressure on the forest due to increasing demand for agricultural land within the official forest boundaries, documented through GIS analysis (Lung and Schaab 2004, Lung and Schaab 2006).

Moreover, the legal and illegal cattle grazing and the extraction of timber and non-timber products by the locals, who depend on the forest for their daily subsistence needs, puts additional pressure on the forest (Kokwaro 1988, Lung and Schaab 2004, Mitchell 2004, Onyango et al. 2004, Kambona 2005, Bleher et al. 2006). Studies have shown that in 1993 about 99% of the people living next to the forest relied on forest products for their daily needs and 78% were using them as a source of income (McGeoch et al. 2008).

There have been several attempts to protect and conserve the forest (some of which are ongoing) either through the participation of local grassroots organizations such as the Kakamega Environmental Education Program (KEEP) or through the enforcement of management plans by the appropriate authorities. The most referred plan was developed by the Kenya Indigenous Forest Conservation Programme (KIFCON), in the early 1990s, and proposed dividing the forest into zones for different uses (Bleher et al. 2006). Based on that plan, which was never implemented, three zones would have been created: 1) Protection zone 2) Rehabilitation zone for restoration purposes in areas where forest degradation was very high and 3) Subsistence zone in which local people could extract resources (Bleher et al. 2006).
Today, several scientists are working in Kakamega Forest trying to understand the complex ecological interactions between different aspects of its biodiversity so that they can contribute to the conservation of the forest. Most of them are members of BIOTA East Africa, an umbrella organization that consists of several German Universities that collaborate with the Kenya Wildlife Service and the National Museums of Kenya (NMK). One of the main reasons the forest has attracted so much attention is because it is Kenya’s only rainforest, with very rich biodiversity different from other areas in the rest of the country (Oyugi 1996). Ten to twenty percent of the species found in Kakamega are not found anywhere else in Kenya (Oyugi 1996). Kakamega’s avifauna is especially rich with more than 300 species (Zimmerman 1972) and the forest is recognized as an Important Bird Area (IBA), as defined by BirdLife international (Bennun and Njoroge 2000). Because of its uniqueness and the urgent need to conserve the forest (due to high and increasing human pressures), scientists have been studying extensively the impact of anthropogenic habitat disturbance on the forest’s biodiversity and as a result, there is now a large number of studies in this region.

2.3.2 Model Development

Using the information and the significant relationships reported in the rich literature on Kakamega Forest, I developed a path model (Figure 2.2) illustrating the possible mechanism influencing the bird communities in the area. I chose to focus on birds for the following reasons:

a) Birds are considered good biological indicators of the forest because they respond quickly and in detectable ways to habitat disturbance (Munyekenye et al. 2008).

b) Kakamega Forest is known for its rich bird diversity and is recognized as an Important Bird Area (IBA) (Bennun and Njoroge 2000). For that reason the conservation of its avifauna is of vital importance.

c) Birds perform crucial ecosystem functions, for example as dispersers (e.g. the Common Bulbul (*Pycnonotus barbatus*)) and pollinators, so their conservation is important for maintaining those functions (Schleunig et al. 2011, Sekercioğlu et al. 2004, Sekercioğlu 2006).
Based on the literature, I hypothesized that habitat disturbance will have an impact on the bird species richness mainly by negatively influencing the vegetation composition and structure of the forest. In the path diagram developed (Figure 2.2) anthropogenic habitat disturbance is represented by the number of trees logged according to the findings of Bleher et al. (2006), which suggest that the number of stumps is the most suitable and effective indicator of disturbance in the forest.

Several studies have shown that the management regime of the forest influences the levels of disturbance (Bleher et al. 2006, Lung and Schaab 2006, Mitchell and Schaab 2008) therefore the management strategy was included as an exogenous variable partly explaining anthropogenic habitat disturbance. GIS studies have mentioned that human population densities and access to the forest, in terms of distance to the roads and to the forest edge, also influence disturbance (Lung and Schaab 2006, Mitchell and Schaab 2008) therefore those two factors were included in the model.

Althof (2005) showed that habitat disturbance has a direct impact on plant diversity and it also influences the vertical vegetation profiles which vary considerably among sites in Kakamega Forest. Therefore those two variables were added into the model, being influenced by the habitat disturbance. Habitat heterogeneity measured as vertical vegetation variability appears to have an impact on bird species richness (Maina 2002) probably because more complex sites can support higher numbers of species by providing more ecological niches (Maina 2002, Eshiamwata et al. 2006). For that reason, vertical foliage diversity was included as a variable in the model, being influenced by disturbance and also influencing bird richness (Maina 2002).

Plant diversity, already included in the model, was hypothesized to have a direct impact on birds because it is an indicator of the forest’s vegetation composition, a factor known to influence birds (Maina 2002), and because areas with different plant diversity are likely to support different bird communities. One possible explanation for this is that higher plant diversity represents higher food availability. It is widely accepted that food abundance is considered an important limiting factor (Maina 2002). According to Brooks et al. (1999) fragment size also appears to have a direct impact on the bird communities so it was included as an exogenous variable.
explaining bird richness. All variables included in the model are listed in Table 2.1 and the corresponding path model is shown in Figure 2.2.

Table 2.1 List of variables included in the path model developed based on the available literature on Kakamega Forest

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management Strategy</td>
<td>Exogenous</td>
</tr>
<tr>
<td>Human Population Densities</td>
<td>Exogenous</td>
</tr>
<tr>
<td>Distance to the road</td>
<td>Exogenous</td>
</tr>
<tr>
<td>Number of trees logged</td>
<td>Endogenous</td>
</tr>
<tr>
<td>Foliage Height Diversity</td>
<td>Endogenous</td>
</tr>
<tr>
<td>Plant Species diversity</td>
<td>Endogenous</td>
</tr>
<tr>
<td>Fragment Size</td>
<td>Exogenous</td>
</tr>
<tr>
<td>Bird Diversities</td>
<td>Endogenous</td>
</tr>
</tbody>
</table>

Figure 2.2 The path diagram developed based on the available literature on Kakamega Forest

The attempt to parameterize the above model using only data from the published literature was unsuccessful so I contacted scientists working in Kakamega Forest, to ask for possible access to available raw data. In addition, during a meeting in Silwood Park, UK, in September 2008 I presented the above model to Dr. Katrin Böhning-Gaese and Dr. Matthias Schleuning (a member of professor’s Böhning-Gaese lab), both working on factors that influence birds communities in Kakamega Forest, and discussed the variables that were included in my original model identified from the literature. We decided to use data from 22 study sites in Kakamega Forest (Figure 2.3) for which most of the required information was available and were used for previous research conducted by Professor’s Böhning-Gaese group (Farwig et al. 2006b, Farwig et al. 2008c, Farwig et al. 2009).
Moreover, based on their expert opinion the model was modified to incorporate new factors or relationships that were thought to influence the bird distributions. During a second follow-up meeting in Mainz, Germany in November 2008, at which Dr. Gertrud Schaab also participated, the model was finalized by taking into account all participants’ comments.

Compared to the literature-based model (Figure 2.2), in the new model (Figure 2.4) the “distance to the road” variable has been replaced by the distance to the nearest forest edge. It was hypothesized that the distance to the edge of the forest is more likely to have a stronger impact on the levels of habitat disturbance since areas with shorter distances to the edge are more accessible to the local communities. It was also suggested that this variable would capture the edge effects on bird communities, a factor that according to the experts is often considered to negatively influence birds. Therefore a direct causal link was added between the new variable and “bird communities”.

Figure 2.3 Map of Kakamega Forest showing the 22 study sites from which data were collected to parameterize the path model (Dataset 1: Farwig et al. 2006b; Dataset 2: Farwig et al. 2008c)
A direct link between habitat disturbance and bird diversity was also added because it was proposed that disturbance not only influences birds indirectly through changes in vegetation, but it might also have a strong direct impact on the bird communities due to increased presence of humans which influences bird behaviour by causing for example birds to flee (Blumstein et al. 2005) and therefore interrupt important activities such as feeding and nesting (with potentially negative consequences to their fitness) (Burger and Gochfeld 1991, Gill 2007). In terms of the hypothesized direct influence of the vegetation structure on birds, two new variables were added in the model: canopy cover and stem densities. Although their importance has not been studied or reported in the literature in Kakamega Forest (and therefore were not considered in the original model) it has been shown in other areas that they are influencing the bird communities (Sekercioglu 2002, Thinh 2006, Meynard and Quinn 2008).

Correlation links were added between all four vegetation variables showing that these variables are likely to be influenced by common factors that were not measured or included in the model, such as climatic conditions and soil properties (Althof 2005). Lastly the fragmentation variable was removed as there would not have been enough replication in the 22 sites chosen for this analysis and because the fragmentation effects are likely to be captured by the distance to the forest edge variable.
2.3.3 Data Availability and Collection

Data on plant diversity and bird species richness and abundances were obtained from the datasets used for the Farwig et al., 2006b and 2008c (Corrigendum: Farwig et al. 2009). Data on human population densities and distance to the edge were calculated from previous GIS analyses carried out by Professor Schaab (unpublished data). Data on habitat disturbance and foliage height diversity were only available for half of the study sites and data on canopy cover and stem densities were not available at all. It was hence decided that I visit the 22 study sites and collect data on these variables during a fieldwork season (January to April 2009). Below is a detailed description of the data obtained and the methods used to collect them.

2.3.3.1 Data on plants and birds

The raw data used to calculate the plant diversity and bird abundance, species richness and diversity were provided by Professor Böhning-Gaese; they were collected in 2001/2002 (hereafter Dataset 1) and 2005/2006 (hereafter Dataset 2) in the 22 study sites in the forest (Figure 2.3). The tree data from the Dataset 1 were collected at 9 of
the sites, at which all trees with a diameter at breast height (DBH) of ≥20 cm were identified to the species level and recorded within a 1-ha plot. The data from Dataset 2 were collected at 15 study sites (two of these were the same as in the first dataset; Figure 2.3) which included five different types of habitats (primary forest, secondary forest, plantations of a mixture of indigenous species, monocultures of indigenous Maesopsis eminii and monocultures of exotic Bischoffia javanica) found in both the northern and the southern part of the forest. Again, at each study site all trees with a DBH of ≥20 cm were identified to the species level and recorded, in three 1-ha plots at each site.

The bird data were also collected at the same study sites in 2001/2002 (Dataset 1) and 2005/2006 (Dataset 2) using point counts all conducted between 07:00 and 8:30 am by the same local field assistant, Nixon Sajita. All species seen and heard within a radius of 20 meters were recorded for a total of 10 minutes at each time. For the first dataset, birds were sampled once a month at each site for a total of 12 months, while for the second dataset birds were sampled also once a month but for a total of 13 months. For a more detailed description of the methods see Farwig et al. 2006b and 2008c.

The tree and bird diversities were calculated using the Biodiversity Calculator (Danoff-Burg and Xu 2006), by estimating the Shannon-Wiener Index (Magurran 1988) for each site using the following formula:

\[ H' = -\sum_{i=1}^{S} (p_i \ln p_i) \]

where \( S \) is the number of species, and \( p_i \) is the relative abundance of each species calculated by dividing the abundance of each species by the total abundance in each site. The index takes into account not only the number of species but species evenness as well and it is one of the most commonly used so it allows for comparisons across different conservation studies.

In the case of the bird data, because there was more than one sample collected for each site, the diversities were calculated separately for each month and then averaged for each site to obtain an overall value per site. The same approach was used for the
In addition to calculating the above overall bird abundances, richness, and diversity the bird species were also categorized using two different classification criteria and the three measures were again calculated for each category. Categorization was based on:

a) birds’ forest use, i.e. forest specialists (FS), forest generalist (FG), and forest visitors (FV), following the classification in Bennun et al. (1996) where forest specialists are species that are mainly found in the forest interior and are more likely to be affected by forest disturbance, e.g. the Olive-green Camaroptera (*Camaroptera chloronata*); forest generalists are species that are also found in the forest interior but can use forest edges and fragments as well e.g. the Western Black-headed Oriole (*Oriolus brachyrhynchos*); and forest visitors are species that don’t depend on the forest for their survival e.g. the African Paradise Flycatcher (*Terpsiphone viridis*).

b) bird feeding guilds, i.e. frugivores and insectivores. To categorize the birds I used the comprehensive diet classification list of Sub-Saharan birds developed by Kissling et al. (2007) and provided by Dr. Katrin Böhning-Gaese. The list classifies birds according to the food items included in their diet (e.g. fruits, seeds, and insects). All birds that were recorded to rely on fruit as a major food item in their diet (both obligatory, e.g. the Yellow-rumped Tinkerbird (*Pogoniulus bilineatus*) and partially, e.g. the Joyful Greenbul (*Chlorocichla laetissima*) were classified as frugivores and similarly all birds that relied on insects (such as Black-faced Rufous Warbler (*Bahmocercus rufus*)) where classified as insectivores.

A total of 130 species were sampled in the forest. Table 2.2 shows the number of species for each of the bird categories used in the analysis.
### Table 2.2 Number of species in each bird guild examined

<table>
<thead>
<tr>
<th>Bird Guilds</th>
<th>Number of Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest Specialists</td>
<td>45</td>
</tr>
<tr>
<td>Forest Generalists</td>
<td>40</td>
</tr>
<tr>
<td>Forest Visitors</td>
<td>31</td>
</tr>
<tr>
<td>Frugivores</td>
<td>34</td>
</tr>
<tr>
<td>Insectivores</td>
<td>104</td>
</tr>
</tbody>
</table>

#### 2.3.3.2 Data on human densities, distance to the edge and management type

The human population densities were estimated using data from the 1999 national population census, conducted by the Kenyan Central Bureau of Statistics. The KCBS report lists the human populations at the sublocation level. Professor Schaab used GIS analysis to estimate the population densities by calculating the number of inhabitants per km² based on the populations reported and the total areas of each sublocation around the forest. Then for each of the twenty-two sites the population densities were estimated at the nearest forest entry from the centre of the study site. In addition, for each study site the distance to the forest edge (in meters) was calculated again from the centre of the site, using 2003 Landsat imagery and by considering forest glades as forest edge. The Management Type variable was created by distinguishing between the two management authorities in the forest assigning 1 to all sites managed by the Kenya Wildlife Service (KWS) and 2 to all sites managed by the Forests Department (FD).

#### 2.3.3.3 Number of trees logged and vegetation data

The data for the rest of the variables included in the model (i.e. canopy cover, stem densities, foliage height diversities (FHD), and anthropogenic habitat disturbance) were not available for all sites from previous studies and therefore were collected during a three-month field-season, from January to April 2009. At each of the twenty-two study sites, five 100m-long transects were established as closely as possible to the original sites where the aforementioned tree and bird data were collected. At each site, the number of trees logged was recorded along each transect and within a distance of 5m to each side. In addition, to ensure a sufficient sample size, another 300 meters of existing paths and trails were also surveyed covering in total an area of 0.8 ha. All trees with obvious signs of cuttings were recorded and their DBH measured.
At each site the canopy cover and foliage height diversity were measured at five different equidistant points along each of the five 100m-long transects. The canopy cover was estimated by taking a picture at each of the 25 points using a digital camera (Nikon Coolpix 2400) and a hemispherical lens (Nikon FC-E9) (Danson et al. 2007). Using the Gap Light Analyzer software (Version 2.0, Frazer et al. 2000) the percentage of canopy openness was calculated at each point which was then averaged to obtain an overall value for each site.

The FHD was calculated using the protocol described in Farwig et al. 2008c, by visually estimating the vegetation cover at 7 different height levels (0m, 1m, 2m, 4m, 8m, 16m, canopy). The Shannon-Weiner index was then used to estimate the FHD at each of the 25 points which was then averaged to calculate the overall FHD at each site. The stem densities were calculated by establishing ten 10m x 10m plots at each site, and measuring all stems with a DBH of more than 5cm in those plots. The total number of stems was divided by the total area covered (0.1 ha) to estimate the stem density per site. The location of these plots along each transect was chosen using a random number generator.

2.3.4 Statistical Analysis

The analysis was carried in AMOS 20.0.0, a statistical package developed especially for SEM analysis. All variables included in the model were tested, using SPSS 17.0 for collinearity and to check whether they meet the basic assumptions of SEM as described in Chapter 1. Canopy openness, distance to the forest edge and number of stumps were log-transformed and plant diversity were exponentially transformed to linearize the data. The fitness of the model was first assessed using the chi-square test which examines whether the model implied covariance matrix is not-significantly differently from the observed matrix, suggesting a good fit in that case (Grace 2006) and then by using the Normed Fit Index (NFI), the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA), which are three of the most common fitness tests used in SEM (Schermelleh-Engel et al. 2003). Based on these tests, the original model was simplified and retested to attain the most parsimonious model that fits the data satisfactorily. The two-stage False Discovery Rate (FDR) method, described by Pike (2011), based on the work of Benjamini and Hochberg
Parameterizing the model for Kakamega Forest (1995), was used to correct for possible inflated Type I errors (i.e. when the null hypothesis is rejected when is true) due to multiple comparisons in the model. In addition, the results of the final model were compared to the results of the analogous multiple regression analyses, carried in SPSS 20.0, and based on the mediation method suggested by Baron and Kenny (1986).

2.4 RESULTS

The first path model tested (Figure 2.4) included nine observed variables, with six of them being endogenous. Table 2.3 shows the minimum, maximum, mean and standard deviation of the eight continuous habitat variables included in the analysis. Results showed that this specific model did not fit the data well, since the observed covariances were significantly different from the covariances implied by the model. The chi-square value for this model is 39.228 with 17 degrees of freedom, and a p-value of 0.002. The lack of fit means that proposed causal model does not successfully represent the actual mechanisms of the system studied and consequently the model needs to be modified and re-tested.

A second simplified model was therefore tested (Figure 2.5), which excluded the “Human Population Densities” and “Management type” variables which only explained a small percentage of the habitat disturbance variance (0.12). The human population densities were removed because the variable was calculated using the 1999 census data which were collected long before the rest of the variables and at very coarse level. As mentioned in the methods section, the census data available from the Kenyan Central Bureau of Statistics reported human populations at the sublocation
level. Many of the study sites are located within the same sublocation so it was not possible to accurately estimate the correct human population densities at each site at a meaningful scale.

The categorical “Management Type” variable, distinguishing between the sites managed by the Forests Department and KWS, was also removed for two reasons: a) because the protection status of the sites does not only depend on which authority is managing the forest, but also varies with whether the site is located within a natural reserve or not (Bleher et al. 2006, Mitchell and Schaab 2008) thus confounding the results if the analysis is only based on management type and b) because the sample size is very small (n=22) so the use of categorical variables with only two levels can be problematic.

Figure 2.5 The resulting path diagram, after the “Human Population Densities” and the “Management Type” variables were excluded, showing the corresponding path coefficients and the R² values of each endogenous variable (top right corner)
The results showed that this second model (Figure 2.5) fits the data well with a chi-square value of 2.174 (df=4) and a p-value of 0.704 (NFI=0.957, CFI=1.000 and RMSEA=0.000). The AIC was 64.174 and the $R^2$ for the response variable 0.16, meaning that the model explains 16% of the variation in bird densities. Interestingly, when bird species richness and abundance were tested instead, the model explained higher percentage of their variance (40% and 70% respectively).

There were several not-significant path coefficients, especially for bird diversities, but this is not necessarily problematic as long as the values make sense theoretically. In this case, the only one that is not as originally hypothesized is the “Number of Stumps” to “Stem Densities” coefficient (0.42) which is the opposite direction to that expected implying that higher disturbance results in higher stem densities. Most likely the reason for this is because the two variables are interrelated. A site with more stems will probably also have more stumps just because there are more trees available in general. For that reason the SD variable was removed.

The resulting, and more parsimonious model which excludes SD (Figure 2.6) was then tested and results showed that it fits the data equally well with a chi-square value of 1.258 (df=3) and p-level of 0.739 (NFI=0.962, CFI=1.000 and RMSEA=0.000). The AIC is 49.258. This model was accepted as the final one, explaining however only a small percentage of the bird diversity and richness (0.07 and 0.26 respectively) but a high percentage of species abundance (0.61).

Differences between the three measures of the bird communities were also observed when the significance of each path coefficient was examined. As shown in Table 2.4, overall there are more significant path coefficients when abundance was tested compared to the other two.
Figure 2.6 The results of the SEM analysis when the final and more parsimonious model was tested; the path coefficients and the $R^2$ values (top right corner) of each endogenous variable are also displayed.

The distance to the forest edge has a strong direct impact on the overall species abundance (Table 2.4). The numbers of stumps seem to influence negatively all three of the vegetation variables examined which means that the more disturbed a site is the higher the canopy openness and the lower the plant and foliage height diversities are. There seems to be no significant direct impact of habitat disturbance on the bird communities and based on the results canopy openness and foliage height diversity do not influence significantly any of the three measures of the bird communities tested.
Table 2.4 The resulting path coefficients, based on the final model, for the three different measures of the bird communities in Kakamega Forest

<table>
<thead>
<tr>
<th></th>
<th>Diversity</th>
<th>Richness</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to forest edge</td>
<td>Number of Stumps</td>
<td>-.091</td>
<td>-.091</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>Canopy Openness</td>
<td>.637***</td>
<td>.637***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>FHD</td>
<td>-.669***</td>
<td>-.669***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>Plant Diversity</td>
<td>-.449*</td>
<td>-.449*</td>
</tr>
<tr>
<td>FHD</td>
<td>Bird Assemblages</td>
<td>.229</td>
<td>.204</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>Bird Assemblages</td>
<td>.053</td>
<td>.142</td>
</tr>
<tr>
<td>Canopy Openness</td>
<td>Bird Assemblages</td>
<td>.109</td>
<td>.108</td>
</tr>
<tr>
<td>Plant Diversity</td>
<td>Bird Assemblages</td>
<td>.102</td>
<td>.342</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>Bird Assemblages</td>
<td>.114</td>
<td>.327</td>
</tr>
</tbody>
</table>

Significance levels: *0.05, **0.01, ***0.001; Coefficients in blue colour indicate significance at the 0.5 level after correcting for False Discovery Rates (FDR).

When the different bird categories were tested the model consistently explained better the abundances of the species compared to the richness. Forest visitors were the only exception for which the model explained very well both their richness and abundances (0.84 and 0.81 respectively) (Table 2.5). The richness of the frugivores and insectivores were explained the least (0.32 and 0.34 respectively). The impact of each variable changed according to the guild examined (Table 2.6 and 2.7). Yet, the plant diversity and the distance to the forest edge were significant for all bird categories except the species richness of the frugivores, which were not influenced by the plant diversity, and the species richness of insectivores which were not influenced by the distance to the forest edge.

Table 2.5 The R-square values, for the different bird categories examined, showing the percentage of explained variance by the final path model

<table>
<thead>
<tr>
<th>Bird Guild</th>
<th>R-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialists Richness</td>
<td>0.58</td>
</tr>
<tr>
<td>Specialists Abundances</td>
<td>0.75</td>
</tr>
<tr>
<td>Generalists Richness</td>
<td>0.54</td>
</tr>
<tr>
<td>Generalists Abundances</td>
<td>0.72</td>
</tr>
<tr>
<td>Visitors Richness</td>
<td>0.84</td>
</tr>
<tr>
<td>Visitors Abundances</td>
<td>0.81</td>
</tr>
<tr>
<td>Frugivores Richness</td>
<td>0.32</td>
</tr>
<tr>
<td>Frugivores Abundances</td>
<td>0.70</td>
</tr>
<tr>
<td>Insectivores Richness</td>
<td>0.34</td>
</tr>
<tr>
<td>Insectivores Abundances</td>
<td>0.62</td>
</tr>
</tbody>
</table>
The rest of the variables were only significant for few of the categories, and canopy openness was not statistically significant for any of the guilds. As before, habitat disturbance negatively influences all three vegetation variables examined. Forest specialists and generalists are positively influenced by higher plant diversity and larger distances to the forest edge, while forest visitors are negatively influenced by these two variables and by decreasing habitat disturbance (Table 2.6).

Table 2.6 The resulting path coefficients, based on the final model for the forest specialists, forest generalists and forest visitors

<table>
<thead>
<tr>
<th>Path</th>
<th>Specialists</th>
<th>Generalists</th>
<th>Visitors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>N</td>
<td>S</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>-.091</td>
<td>-.091</td>
<td>-.091</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>.637***</td>
<td>.637***</td>
<td>.637***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.669***</td>
<td>-.669***</td>
<td>-.669***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.449*</td>
<td>-.449*</td>
<td>-.449*</td>
</tr>
<tr>
<td>FHD</td>
<td>.253</td>
<td>.306*</td>
<td>.107</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.064</td>
<td>-.012</td>
<td>.019</td>
</tr>
<tr>
<td>Canopy Openness</td>
<td>-.043</td>
<td>-.013</td>
<td>.142</td>
</tr>
<tr>
<td>Plant Diversity</td>
<td>.468**</td>
<td>.550***</td>
<td>.411*</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>.314*</td>
<td>.408***</td>
<td>.573***</td>
</tr>
</tbody>
</table>

Significance levels: *0.05, **0.01, ***0.001; S: Richness, N: Abundance; Coefficients in blue colour indicate significance at the 0.5 level after correcting for False Discovery Rates (FDR).

Species abundances of frugivores and insectivores are positively influenced by higher levels of plant diversity and larger distances to the forest edge. In addition, the abundance of the insectivores is influenced positively by higher levels of foliage height diversity. With regards to the richness of the two feeding guilds only a small percentage was explained but the effects are in the same direction (Table 2.5, 2.7).
Table 2.7 The resulting path coefficients, based on the final model, for frugivores and insectivores

<table>
<thead>
<tr>
<th>Direct Effects</th>
<th></th>
<th>Frugivores</th>
<th></th>
<th>Insectivores</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>N</td>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>→</td>
<td>Number of Stumps</td>
<td>.637***</td>
<td>.637***</td>
<td>.637***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>→</td>
<td>Canopy Openness</td>
<td>-.669***</td>
<td>-.669***</td>
<td>-.669***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>→</td>
<td>FHD</td>
<td>-.449*</td>
<td>-.449*</td>
<td>-.449*</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>→</td>
<td>Plant Diversity</td>
<td>.056</td>
<td>.152</td>
<td>.202</td>
</tr>
<tr>
<td>FHD</td>
<td>→</td>
<td>Bird Communities</td>
<td>-.181</td>
<td>.233</td>
<td>.139</td>
</tr>
<tr>
<td>Canopy Openness</td>
<td>→</td>
<td>Bird Communities</td>
<td>.331</td>
<td>.251</td>
<td>.046</td>
</tr>
<tr>
<td>Plant Diversity</td>
<td>→</td>
<td>Bird Communities</td>
<td>.037</td>
<td>.288*</td>
<td>.468*</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>→</td>
<td>Bird Communities</td>
<td>.498**</td>
<td>.745***</td>
<td>.277</td>
</tr>
</tbody>
</table>

Significance levels: *0.05, **0.01, ***0.001; S: Richness, N: Abundance; Coefficients in blue colour indicate significance at the 0.05 level after correcting for False Discovery Rates (FDR)

Table 2.8 shows the cumulative indirect effect of habitat disturbance (number of tree stumps) on the species richness and abundances of the various bird guilds, by taking into account only the path coefficients that were significant at the 0.05 level (after correcting for False Discovery Rates). The abundances of insectivores and specialists are influenced the most (negatively), while habitat disturbance has a positive indirect effect on the species richness and abundances of visitors.

Table 2.8 Path coefficients, based on the final model, showing the cumulative indirect effects of habitat disturbance on all the bird guilds examined

<table>
<thead>
<tr>
<th>Bird Guild</th>
<th>Indirect effect of habitat disturbance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialists Richness</td>
<td>-0.234</td>
</tr>
<tr>
<td>Specialists Abundances</td>
<td>-0.479</td>
</tr>
<tr>
<td>Generalists Richness</td>
<td>-0.205</td>
</tr>
<tr>
<td>Generalists Abundances</td>
<td>-0.191</td>
</tr>
<tr>
<td>Visitors Richness</td>
<td>0.216</td>
</tr>
<tr>
<td>Visitors Abundances</td>
<td>0.213</td>
</tr>
<tr>
<td>Frugivores Richness</td>
<td>no effect</td>
</tr>
<tr>
<td>Frugivores Abundances</td>
<td>-0.144</td>
</tr>
<tr>
<td>Insectivores Richness</td>
<td>-0.234</td>
</tr>
<tr>
<td>Insectivores Abundances</td>
<td>-0.557</td>
</tr>
</tbody>
</table>
Overall, four clear patterns result from the analysis:

a) The model explains the abundances of species better than species richness and diversities.

b) The impact of the variables depends on the bird guild examined.

c) Habitat disturbance has a strong direct negative impact on the vegetation structure and composition of the forest, and a significant indirect impact on all guilds examined, with the richness of the frugivores being the only exception (Table 2.8).

d) Plant diversity and distance to the forest edge are the two most important factors consistently influencing the bird communities of the different guilds examined.

To examine how the above results of the SEM analysis compare to the results of analogous multiple regression analyses, the path model (in Figure 2.6) was decoupled into the following five regression equations, based on the Baron and Kenny mediation method (1986):

Equation 1: \[ \text{Stumps} = \alpha_1 + \beta_1 \text{DFE} + \varepsilon_1 \]
Equation 2: \[ \text{CO} = \alpha_2 + \beta_2 \text{Stumps} + \varepsilon_2 \]
Equation 3: \[ \text{FHD} = \alpha_3 + \beta_3 \text{Stumps} + \varepsilon_3 \]
Equation 4: \[ \text{PD} = \alpha_4 + \beta_4 \text{Stumps} + \varepsilon_4 \]
Equation 5: \[ \text{BA} = \alpha_5 + \beta_5 \text{FHD} + \beta_6 \text{Stumps} + \beta_7 \text{CO} + \beta_8 \text{PD} + \beta_9 \text{DFE} + \varepsilon_5 \]

where: Stumps=Number of stumps, DFE= Distance to forest edge, CO= Canopy Openness, FHD=Foliage Height Diversity, PD= Plant Diversity, BA=Bird Assemblages.

The resulting beta coefficients (Table 2.9) are very similar to corresponding standardized path coefficients (Table 2.7) and the minor differences can be probably attributed to the fact that, unlike SEM analysis, multiple regression does not take into account that dependent variables may also be subjected to measurement error (Grace, 2006) and thus the correlations between the measurement errors, which were included in the path model (Figure 2.6), could not be incorporated in the regression analyses. Yet, it is important to note that despite these minor differences the conclusions about
the impact and the statistical significance of the predictors examined remain the same for all bird categories tested (please see Tables A.1 and A.2 in Appendix I for the results on the multiple regression analyses on the various feeding guilds).

Table 2.9 Results of multiple regression analyses (standardized beta coefficients) for the overall bird species diversity, richness and abundance

<table>
<thead>
<tr>
<th>Equation</th>
<th>Predictors</th>
<th>Diversity</th>
<th>Richness</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eq 1</td>
<td>Distance to forest edge → Number of Stumps</td>
<td>-0.091</td>
<td>-0.091</td>
<td>-0.091</td>
</tr>
<tr>
<td>Eq 2</td>
<td>Number of Stumps → Canopy Openness</td>
<td>0.637***</td>
<td>0.637***</td>
<td>0.637***</td>
</tr>
<tr>
<td>Eq 3</td>
<td>Number of Stumps → FHD</td>
<td>-0.669***</td>
<td>-0.669***</td>
<td>-0.669***</td>
</tr>
<tr>
<td>Eq 4</td>
<td>Number of Stumps → Plant Diversity</td>
<td>-0.449*</td>
<td>-0.449*</td>
<td>-0.449*</td>
</tr>
<tr>
<td>Eq 5</td>
<td>FHD → Bird Assemblages</td>
<td>0.228</td>
<td>0.202</td>
<td>0.323</td>
</tr>
<tr>
<td></td>
<td>Number of Stumps → Bird Assemblages</td>
<td>0.053</td>
<td>0.141</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td>Canopy Openness → Bird Assemblages</td>
<td>0.109</td>
<td>0.107</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td>Plant Diversity → Bird Assemblages</td>
<td>0.102</td>
<td>0.339</td>
<td>0.415*</td>
</tr>
<tr>
<td></td>
<td>Distance to forest edge → Bird Assemblages</td>
<td>0.113</td>
<td>0.325</td>
<td>0.572**</td>
</tr>
</tbody>
</table>

Significance levels: *0.05, **0.01, ***0.001

2.5 DISCUSSION

The fact that the final path model failed to adequately explain species diversity of birds, but performed better at explaining species richness and abundances suggest that the latter two quantities might be more appropriate indicators of the effects of anthropogenic habitat disturbance on bird communities. Diversity measures, which also take into account species evenness (Magurran 1988) are likely to be influenced by other more complicated factors not captured in this analysis.

The varying impact of the variables included in the model depending on the guild examined has been also documented in other studies (Dale et al. 2000, Sekercioglu 2002, Waltert et al. 2005, Gray et al. 2007, Lee and Marsden 2008, Clough et al. 2009) and was expected because the various guilds considered have different ecological requirements. Yet, distance to the forest edge consistently positively influenced all guilds examined (expect forest visitors) supporting results of previous studies showing that although some tropical birds might be edge tolerant or even show preference in edges in some cases e.g. the sallying insectivores in Budongo Forest, Uganda (Dale et al. 2000), in general there are less species closer at the edge (Dale et
This shows that anthropogenic habitat disturbance and fragmentation can have significant negative impacts on the overall forest bird community. As predicted forest visitors, which according to Bennun et al. (1996) usually depend on non-forest habitats for their survival and their presence may be a sign of habitat disturbance, benefited from shorter distances to the edge and interestingly it was the only guild for which anthropogenic habitat disturbance had a statistically significant direct positive effect.

Habitat disturbance also influenced all vegetation variables included in the model and as shown by previous studies in Kakamega Forest and other tropical ecosystems, it can have a strong negative effect on forest’s vegetation (Gillespie and Walter 2001, Sekercioglu 2002, Althof 2005); a factor that is known to be a strong determinant of the tropical bird communities (Maina 2002, Naidoo 2004). This was shown in this study through the strong impact of plant diversity on species richness and abundance. Interestingly, the effects of the other two measures of vegetation included in the model, canopy openness and foliage height diversity, were not as strong as hypothesized although in other cases they have been shown to influence tropical birds (Seymour and Dean 2010). Even more, the effect of canopy openness, although non-significant, was positive for all the guilds except the forest specialists. These results must be interpreted by keeping in mind that the levels of canopy openness recorded in the forest were fairly low (ranging from 6 to 13%) probably preventing us from capturing any strong effects. If that is true, it is interesting that specialists responded negatively to even such low levels.

Regarding the effects of foliage height diversity, this variable was only significant for the abundances of specialists and insectivores. Among all the guilds examined these two are generally considered the most sensitive to habitat disturbance (Sekercioglu et al. 2002, Sodhi et al. 2004, Lees and Peres 2006, Lindell et al. 2007, Sekercioglu and Sodhi 2007). The reason they were affected by FHD could be due to the possibility that lower levels of vertical diversity are likely to result into less rich microhabitats in the forest on which both guilds depend for their foraging and nesting requirements (Sodhi et al. 2004). This is only a hypothesis though and to reveal the exact mechanisms for this particular relationship more detailed research is needed.
In conclusion, the results suggest that fragmentation (e.g. through expansion of subsistence agricultural practices which results into more forest edges) and habitat disturbance (through extraction of firewood) have a negative impact on the bird communities of Kakamega Forest and need to be addressed by the management authorities if the avifauna is to be effectively protected. It is worth noting that final model used to test the effects of the depended variables on the species richness and abundances of the various bird gilds was refined based on the results of the overall species richness and abundances. However, it is unlikely that the factors excluded (i.e. Management Type, Human Population Densities and Stem Densities) would have influenced model re-specification if each guild was examined independently since the first two variables were removed for not having an effect on the number of stumps, and stems densities for being interrelated with habitat disturbance, i.e. for reasons that would be valid for all guild models.

The results of the analysis support the original conceptual model which was developed using the available literature and hypothesized that anthropogenic habitat disturbance influences indirectly the bird communities of the forest by influencing its vegetation (both the structure and community). It should be noted, though, that the important direct impact of edge effects on birds was missed when the literature of Kakamega Forest was reviewed, but was correctly suggested by the experts working in the field and added to the path model.

The modelling approach chosen to link the different sources of information available for Kakamega Forest was partly successful in the sense that with the help of experts working in field available information was utilized and synthesized to create a path model that describes to a satisfactory extent the factors that influence bird communities. Parameterizing a model solely from the literature though is impracticable if the required information is not available for the reasons explained. Even then, there might be cases where additional data still must be collected because important variables (that must be included in a SEM analysis) may have not been studied before by other researchers. Nevertheless, the modelling approach introduced in this study shows that there is potential for at least in some cases to take advantage of the enormous amount of information and data already available to link factors from
different studies and understand how they relate to one another; even if that means that the existing information must be complemented with some new additional data.
CHAPTER 3

DEVELOPING AND TESTING THE PATH MODEL FOR CYPRUS
3.1 ABSTRACT

Following the same methodological approach used for Kakamega Forest, I developed a path model that illustrates the mechanisms influencing species richness of birds in thirty-eight Natura 2000 sites in Cyprus. The model is parameterized by bringing together various sources of unpublished data. Results show that habitat diversity, the size of the area and the road density within the protected areas are the most important factors determining species richness, with the latter having a negative effect on all bird guilds examined except forest-dependent birds which are more influenced by vegetation cover.

3.2 INTRODUCTION

Despite the ongoing efforts of scientists, conservationists and policy makers to stop biodiversity loss, ecosystems and species are negatively impacted at high rates due to a number of anthropogenic activities including habitat destruction, fragmentation and land modification (Hoffmann et al. 2010). Within Europe the efforts to halt biodiversity loss by 2010 have been largely unsuccessful, with many species still threatened by human activities including changes in land use (Young et al. 2005, Verboom et al. 2007, Santos et al. 2008, Hoffmann et al. 2010).

The need to protect biodiversity been long acknowledged and efforts have been made, especially within the European Union, to create and implement policies for biodiversity protection. An attempt to protect EU avifauna dates back to 1979 when the Birds Directive (Council Directive 2009/147/EC on the conservation of wild birds) was passed with the aim of protecting species and their habitats by setting up a network of protected areas termed Special Protection Areas (SPA) (Young et al. 2005). The directive was later incorporated into a new policy called the Habitats Directive (officially known as Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora) which aimed at protecting important habitats and species of Europe through a complementary network of protected areas known as Special Areas of Conservation (Young et al. 2005). The SACs together with the SPAs make up the Natura 2000 network of protected areas (Young et al. 2005) and so far a total of 5,315 SPAs and 22,529 SACs.
areas have been designated covering 11.4% and 13.7% of EU’s total area respectively (EC 2011).

The designation of these sites alone has not been sufficient to protect the biodiversity found within them (EC 2000). Based on the provisions of Article 6 of the Habitats Directive, EU member states have been asked to develop management plans for each of these areas (EC 2000). For some countries this process is likely to take some years and as a result many regions still remain significantly understudied and potentially poorly managed. This is especially true for Cyprus where, despite its rich biodiversity and its importance as a biodiversity hotspot (Myers et al. 2000) and an endemic bird area (Stattersfield et al. 1998), little is known about the factors that influence the bird richness of the island.

The goal of this study is to use available data to identify and study the factors that influence the species richness of birds found within the network of Natura 2000 sites in Cyprus. The island is an important migratory corridor for birds travelling between Europe and Africa, since it is located on one of the eight “principal migration routes”. Cyprus supports a rich avifauna of 370 species (Iezekiel et al. 2004). Consequently, it is of vital importance for conservation purposes to have a good understanding of the factors that affect the presence and the distribution of bird species on the island.

Following an approach similar to that used for Kakamega Forest (Chapters 2), I use structural equation modeling (SEM) analysis to link sources of available information to study the impact of socio-ecological factors on the species richness of birds in the region. As explained in Chapter 1, SEM is a multivariate statistical tool that can be used to simultaneously study the effects of independent variables on more than one dependent variable and also on one another (Grace 2006). It can be used to confirm or reject an a priori hypothesized multivariate model developed to explain the functioning of certain aspects of a system (Malaeb et al. 2000, Iriondo et al. 2003). Here I use SEM to investigate the underlying processes that determine bird species richness in protected areas.

Through SEM one can study the individual contribution of each indirect and direct variable included in a causal model, and also evaluate their impact on the response
Developing and testing the path model for Cyprus

variable(s) of interest (Schumacker and Lomax 2004). The end result is a path model (in which each path has its own coefficient indicating the strength of that link) which has been tested using appropriate tests of goodness of fit (Grace 2006). The results of Chapter 2 have shown that once the right data are available, SEM can be a powerful tool which can be used to effectively link the different sources of information allowing us to extract important conclusions at a broader and more inclusive level.

3.3 METHODS

3.3.1 Study Area

Cyprus is located in the eastern end of the Mediterranean Sea, south of Turkey and northwest of Lebanon (Figure 4.1). It covers an area of 9251 km$^2$ and its climate can be described as a typical Mediterranean climate with dry summers and wet winters (Iezekiel et al. 2004, Paralikidis et al. 2010). The annual precipitation of the island is 480mm (Iezekiel et al. 2004). The flora of the island is considered particularly rich with about 1865 species and subspecies, out of which 131 are endemic (Iezekiel et al. 2004). It has also rich bird diversity with 370 species recorded, of which 53 are permanent residents and 31% are known to have bred on the island at least once (Iezekiel et al. 2004).

Cyprus has been a full member of the European Union since May 2004 and following the obligations of the EU’s Birds and Habitats directives, has already designated 40 SACs and 29 SPAs covering an area of 883 km$^2$ and 1593 km$^2$ respectively (EC 2011). The sites are all located in the southern part of the island since currently the Government of Cyprus does not exercise control in the northern part due to a military invasion by Turkey in 1974 and the illegal occupation of that area since then.
Developing and testing the path model for Cyprus

Figure 3.1 Map of eastern Mediterranean Sea showing the location of Cyprus (Data source: ESRI)

3.3.2 Model Development
Following an approach similar to that used for the development of the model for Kakamega Forest (Chapters 2), a path diagram was developed from literature illustrating the factors that might influence bird species richness (Figure 3.2). Since the biodiversity of the island is poorly studied, in contrast to the case of Kakamega Forest, I had to rely on studies conducted elsewhere in the Mediterranean region and focus on those covering a similar type of habitats found in Cyprus.

Figure 3.2 Path model for the Natura 2000 sites developed based on available literature
It was hypothesized that bird species richness was directly influenced by the area of the site (Freemark and Merriam 1986, Donázar et al. 2005, Gil-Tena et al. 2007), the vegetation cover (Jetz and Rahbek 2002, Luoto et al. 2004, Coreau and Martin 2007, Bino et al. 2008), habitat diversity (Jetz and Rahbek 2002, Luoto et al. 2004), distance from migratory corridors, and anthropogenic habitat disturbance described by road density within the areas and possibly other anthropogenic land uses (Brotons and Herrando 2001, Santos et al. 2008). Several studies have shown that roads and linear structures have a significant impact on birds although the impact is not uniform across species (Brotons and Herrando 2001, Benítez-López et al. 2010). Also, it has been shown that other anthropogenic factors might influence birds including agricultural intensification and land abandonment (Kati et al. 2009).

The levels of habitat disturbance were hypothesized to be influenced by human population densities around each area as more densely populated areas are likely to have a higher density of roads. In addition, a direct link was added between anthropogenic habitat destruction and the two vegetation variables (cover and habitat diversity) since disturbance is likely to affect both. Correlation links were added between area, vegetation cover and habitat diversities (Figure 3.2) because all three factors are probably related in the sense that larger areas will have higher vegetation cover and higher diversities and also areas with higher habitat diversities are likely to have higher vegetation cover.

Acknowledging that other important factors might be influencing birds and may have been missed, I presented the above model to a number of experts working on biodiversity issues on the island for their comments (see Appendix II for the list of experts). Based on their feedback a new improved model was developed which included three additional variables: distance to water sources, hunting pressure and altitude (Figure 3.3). Some of the experts suggested that availability of water might be an important factor determining the distribution of birds. Similarly, because hunting is very popular in Cyprus, the hunting pressure might influence the presence of birds. Lastly, according to the experts it is widely accepted that altitude has a negative influence on bird species richness.
3.3.3 Data Collection

All the necessary data were collected from sources that were already available. I chose to include in the model only the SACs (Figure 3.4), i.e. the areas designated under the Habitats directive, for which all the required information was available. This was because in the case of the SPAs (designated under the Birds Directive) important information was missing including the number and extent of each habitat type and the percentage of hunting reserves (i.e. areas where hunting is prohibited).

In addition, I excluded from the dataset the two solely marine sites (CY3000006-Thalassia Periochi Nisia and CY4000006-Thalassia Periochi Moulia) since birds recorded in these areas are likely to be influenced by other factors not included in the model. A total of 38 sites was used for the analysis (see Appendix III for a list of the sites included).
The main source of data was BioCyprus (version Sep. 2010), which is a database created and submitted to the EU (under the obligations of the Habitats Directive) by the Environment Department of the Ministry of Agriculture, Natural Resources and Environment of Cyprus. The database is freely available for download at the EIONET Central Data Repository (http://cdr.eionet.europa.eu/cy/eu/n2000/envtkl3sg).

The database was used to extract the overall species richness of birds, which included mostly migratory birds (96%), the species richness of birds listed in Annex I of the Birds Directive, the area of each SCA, the mean altitude, and the percent of hunting reserves in each site. The latter was used as an indicator of hunting pressure since no data were available on the number of hunters in each protected area.

The species richness of birds was further categorized according to the habitat preference of each species by using the corresponding lists reported in Tucker and Evans (1997):

1. Mediterranean forest, shrubland and rocky habitat species (e.g. Long-legged Buzzard (Buteo rufinus))
2. Agricultural species (e.g. Barn Swallow (Hirundo rustica)).
3. Wetland species (e.g. Purple Heron (*Ardea purpurea*))

The data on vegetation and agricultural cover was extracted using ArcMap 9.2 (ESRI 2009) from the CORINE Land Cover 2000 database (version 13). The CLC 2000 is an EU wide project which maps land cover uses in all member states and is freely available for download at the European Environment Agency’s website ([http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-raster](http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-raster)). CLC 2000 uses 44 different land use classifications which include various types of vegetation, agricultural areas, artificial structures and water bodies. The land use categories are grouped at three different levels with the 3rd one containing the most detailed types of categories. For the purposes of this project I used level 1 which only distinguishes between the following five general categories: artificial surfaces, agricultural areas, forest and semi-natural areas, wetlands and water bodies. The total area of the two categories of interest (i.e. agricultural areas and forest and semi-natural areas) was calculated for each of the 38 sites and then based on the total area of each site the percentage of cover of each category was estimated.

The human population densities were calculated using the census of 2000 (at the village level) obtained from the Statistical Service of Cyprus (Ministry of Finance). Using ArcMap 9.2 I first identified the villages of which their boundaries fall within the protected areas. The number of residents for each of those villages was divided by its area to obtain the density. Then densities were added to obtain the total number for each protected site. Similarly, the road density was calculated using ArcMap 9.2 by extracting the total length of main roads within each site and by dividing that with the total area of each protected area.

To quantify migratory corridors, a map was obtained from the GameFund of Cyprus (Ministry of Interior) which showed the main migratory routes on the island. Since a lot of those routes were within the SACs, instead of calculating the distance to the corridors I calculated the percentage of overlap by using ArcMap 9.2 to estimate the total area of the corridor within each site and then divided that number by the total area of the protected site.
Habitat diversity was calculated from the habitats’ map provided by the Forests Departments of Cyprus (of the Ministry of Agriculture, Natural Resources and Environment) which shows the extent of all habitats within each site. Again, using ArcMap 9.2, the area of each habitat was calculated and the overall habitat diversity was estimated using the Shannon-Wiener Index (Magurran 1988):

\[ H' = -\sum_{i=1}^{S} (p_i \ln p_i) \]

where \( S \) is the number of habitats, and \( p_i \) is the relative cover of each habitat calculated by dividing the cover of each habitat by the total cover in each site.

The distance to water sources was not included in the analysis because it was impossible to calculate accurately the availability of water resources for two main reasons:

a. the map data provided by the Water Development Department of Cyprus (of the Ministry of Agriculture, Natural Resources and Environment), which showed the location of the rivers and other water bodies, did not include any information on the actual water flow and water availability. According to the department both the flow and the availability varies considerably spatially and temporally and therefore any variable based on just the distance to the nearest water body would have been misleading.

b. in addition, according to the GameFund in Cyprus (the official body responsible for managing the avifauna of the island) many water points have been installed throughout the protected areas (and the island in general) for which no spatial data are available. The purpose of these water points is to increase the access of birds (especially game species) to water particularly in water scarce areas. Obviously the presence of these points and the lack of spatial data on their location would have confounded any results based on just the distance to the nearest river or water body.
3.3.4 Statistical Analysis

The SEM analysis was conducted using AMOS 20.0. All variables were first checked for linearity and normality and collinearity using SPSS 17.0 and then transformed using logarithmic transformations when appropriate. The only exception were the variables extracted from CLC 2000 on vegetation and agricultural cover, which were linearized using probit transformation because of their sigmoid distribution. Probit transformation does not allow for variables to be either 100 or 0 therefore in the few cases where the cover was either entire or none the value was first transformed to 99.9999 or 0.00001 accordingly. This will not influence results because in effect the actual area changed is very small and realistically no area is fully covered by vegetation or agriculture or has none of them at all. The only reason we get these extreme values in the first place is because of the coarse scale at which the CORINE data are available.

As in Chapter 2, the two-stage False Discovery Rate (FDR) method (Pike 2011) was used to correct for inflated Type I errors due to multiple comparisons. In addition, the results of the final model were compared to the results of the analogous multiple regression analyses based on the method suggested by Baron and Kenny (1986).

3.4 RESULTS

The average area of the SACs was 2312.2 hectares with a standard deviation of 4253.4. The minimum area was 4.9 ha and the maximum 18257.7 ha. Species richness of birds varied across sites and categories. Table 3.1 shows the minimum, maximum, and average number of species in the protected areas for each bird category.

| Table 3.1 Basic descriptive statistics for each species richness category in the Natura 2000 sites |
|-----------------------------------------------|-------|------|------|------|------|
|                                              | Total number of species | Minimum | Maximum | Mean | Std. Deviation |
| Overall Bird Richness                        | 248   | 6    | 99    | 48.45| 18.376 |
| Birds of Annex I                             | 94    | 1    | 67    | 21.87| 12.276 |
| Forest Birds                                 | 54    | 9    | 46    | 26.71| 8.463  |
| Agricultural Birds                           | 84    | 8    | 58    | 26.16| 10.417 |
| Waterbirds                                   | 79    | 1    | 53    | 11.53| 13.770 |
Results of the path analysis showed that the experts’ model (Figure 3.3) did not fit the data well for any of the bird categories examined. The chi-square value for the overall bird richness was 46.970 with 23 degrees of freedom and a p-value of 0.002. Based on these results and on the modification indices provided by AMOS 20.0 the model was re-specified and simplified by removing the following links for which their path coefficients were highly insignificant:

- Mean Altitude $\rightarrow$ Species Richness (p-value=0.787)
- Road density $\rightarrow$ Habitat Diversity (p-value= 0.672)
- Road density $\rightarrow$ Vegetation Cover (p-value = 0.553)

In addition, as suggested by the modification indices, a link was added between human population densities and area, which is also theoretically valid since larger areas have more villages within their boundaries and consequently potentially higher population densities (Figure 3.5).
The chi-square values for the resulting refined model (Figure 3.5) showed that it was statistically significant for all bird categories but the rest of the goodness of fit tests used (e.g. RMSEA and NFI) showed a relatively poor fit. Table 3.2 shows the goodness of fit results for all different bird categories.

Table 3.2 Results of the goodness of fit tests for the experts’ model

<table>
<thead>
<tr>
<th>Bird Richness</th>
<th>Chi-square (df=17, p=0.145)</th>
<th>( R^2 )</th>
<th>RMSEA</th>
<th>NFI</th>
<th>CFI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annex I</td>
<td>24.252 (df=17, p=0.113)</td>
<td>0.59</td>
<td>0.107</td>
<td>0.675</td>
<td>0.845</td>
<td>62.252</td>
</tr>
<tr>
<td>Forest</td>
<td>24.339 (df=17, p=0.111)</td>
<td>0.46</td>
<td>0.108</td>
<td>0.620</td>
<td>0.796</td>
<td>62.339</td>
</tr>
<tr>
<td>Agricultural</td>
<td>23.366 (df=17, p=0.138)</td>
<td>0.50</td>
<td>0.101</td>
<td>0.645</td>
<td>0.831</td>
<td>61.366</td>
</tr>
<tr>
<td>Water Birds</td>
<td>23.332 (df=17, p=0.139)</td>
<td>0.45</td>
<td>0.100</td>
<td>0.639</td>
<td>0.838</td>
<td>61.332</td>
</tr>
</tbody>
</table>

The path coefficients and their significance varied based on the category of the birds examined. This was expected given that different categories have different habitat requirements. Table 3.3 shows the p-values for each path coefficient and each bird category. What is obvious from the results however is that habitat diversity and area were always significant for the species richness of all bird categories.

Table 3.3 P-values of each path coefficient for each species richness category tested using the refined model

<table>
<thead>
<tr>
<th>Human Densities</th>
<th>Road Densities</th>
<th>Overall</th>
<th>Annex I</th>
<th>Forest</th>
<th>Agricultural</th>
<th>Water Birds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habitat Diversities</td>
<td>Species Richness</td>
<td>***</td>
<td>.009</td>
<td>.028</td>
<td>***</td>
<td>.046</td>
</tr>
<tr>
<td>Vegetation Cover</td>
<td>Species Richness</td>
<td>.448</td>
<td>.049</td>
<td>.006</td>
<td>.921</td>
<td>.005</td>
</tr>
<tr>
<td>Area</td>
<td>Species Richness</td>
<td>***</td>
<td>***</td>
<td>.008</td>
<td>.005</td>
<td>.004</td>
</tr>
<tr>
<td>Corridors</td>
<td>Species Richness</td>
<td>.262</td>
<td>.163</td>
<td>.801</td>
<td>.693</td>
<td>.594</td>
</tr>
<tr>
<td>Game Reserves</td>
<td>Species Richness</td>
<td>.206</td>
<td>.144</td>
<td>.640</td>
<td>.519</td>
<td>.031</td>
</tr>
<tr>
<td>Road Densities</td>
<td>Species Richness</td>
<td>.010</td>
<td>***</td>
<td>.102</td>
<td>***</td>
<td>.003</td>
</tr>
</tbody>
</table>

Coefficients in blue colour indicate significance at the 0.5 level after correcting for False Discovery Rates (FDR)
Developing and testing the path model for Cyprus

The same is true for the road densities which consistently had a negative impact on the species richness of birds with the only exception forest birds for which vegetation cover seem to be more important than road density. Based on these results the path diagram was further simplified and tested for goodness of fit by using only the three most important variables mentioned above, i.e. habitat diversity, area, and road density (and the vegetation cover for the forest bird category). A correlation link between road densities and areas was also added since larger areas are likely to have more roads. The model fit was greatly improved reaching satisfactory levels while explaining a nearly equivalent amount of variance (Table 3.4, Figure 3.6).

| Table 3.4 Results of the goodness of fit tests for the final model |
|-----------------------------------------------|--------|--------|--------|--------|--------|--------|
| Bird Richness                               | 0.457  | 0.42   | 0.000  | 0.982  | 1.000  | 18.457 |
| Annex I                                      | 0.457  | 0.51   | 0.000  | 0.985  | 1.000  | 18.457 |
| Forest Birds                                 | 1.164  | 0.42   | 0.067  | 0.959  | 0.993  | 19.164 |
| Agricultural Birds                           | 0.457  | 0.47   | 0.000  | 0.984  | 1.000  | 18.457 |
| Water Birds                                  | 0.457  | 0.30   | 0.000  | 0.974  | 1.000  | 18.457 |

As before, the impact of each variable included in the model changes based on the bird category used (Figure 3.6). When all species are included, the area of the SAC seems to be the most important factor influencing species richness (0.52). The same is true for the birds listed in Annex I (0.51) although they seem to be more vulnerable to road densities compared to the other categories (-0.48). Species that use mostly forests, shrublands and rocky areas are influenced more by the percentage of vegetation cover within the protected area (0.35) while for agricultural birds the diversity of the habitats (0.47) seem to be the most important factor with road densities still having a strong negative impact (-0.41). The water birds are influenced less by the three variables included (compared to the rest of the species) but road densities seem to have a strong negative impact (-0.38).
Figure 3.6 Results of the SEM analyses for each species richness category based on the final model, showing the corresponding path coefficients and the $R^2$ for the endogenous variables.
To examine how the results of the SEM analyses compare to the equivalent multiple regressions I used the following equation to estimate the beta coefficients for each of the three variables included in the path model for each bird category (Figure 3.6):

\[ Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon_1 \]

where \( Y \) = bird species richness and \( X_{1,2,3} \) = the three predictors included in each model.

The resulting beta coefficients (Table 3.5) were very similar to the path coefficients estimated through SEM (Figure 3.6) but in the case of forest birds and waterbirds some of those were not significant. As explained in Chapter 2, these differences are probably due to the fact that in multiple regression one cannot take into account the correlation of the measurement errors of the independent variables.

### Table 3.5 The resulting beta coefficients, estimated through multiple regression analyses, for each bird category

<table>
<thead>
<tr>
<th>Habitat Diversities</th>
<th>Overall</th>
<th>Annex I</th>
<th>Forest</th>
<th>Agricultural</th>
<th>WaterBirds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species Richness</td>
<td>0.340*</td>
<td>0.389**</td>
<td>0.275</td>
<td>0.475***</td>
<td>0.345*</td>
</tr>
<tr>
<td>Area</td>
<td>0.526***</td>
<td>0.520***</td>
<td>0.305</td>
<td>0.395**</td>
<td>0.310</td>
</tr>
<tr>
<td>Species Richness</td>
<td>-0.330*</td>
<td>-0.483***</td>
<td>--</td>
<td>-0.418**</td>
<td>-0.385*</td>
</tr>
<tr>
<td>Road Densities</td>
<td>--</td>
<td>--</td>
<td>0.365*</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Vegetation Cover</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Significance levels: *0.05, **0.01, ***0.001

### 3.5 DISCUSSION

Despite the temporal variation in the data and the coarse scale at which some of the variables were recorded, the model adequately quantifies the impact of three variables on the distribution of birds. The importance of those variables has been documented previously, suggesting the conclusions are consistent with studies conducted elsewhere.

The area of the site has a strong impact because larger areas are able to support more species (Freemark and Merriam 1986, Honkanen et al. 2010, Proenca et al. 2010). The
Developing and testing the path model for Cyprus

proposed explanation for this well known relationship between species richness and area dates back to the novel work of Mc Arthur and Wilson, on the equilibrium theory of Island Biogeography, in which they developed a model to explain the presence of species on islands, through the extinction and colonization rates, which are influenced by among other things the degree of isolation and the area of the island (MacArthur and Wilson 1963, MacArthur and Wilson 1967). Since then, the concept of the theory of island biogeography has been widely used and expanded to include other types of landscapes, with important implications in conservation biology and nature reserve design (Diamond 1975, 1976, Simberloff and Abele 1976), such as fragmented regions and generally areas which are surrounded by unsuitable habitat and therefore could be considered as isolated islands (Higgs 1981, William F 2008).

Habitat diversity has been also shown to positively influence birds in European landscapes (Boecklen 1986, Freemark and Merriam 1986, Atauri and de Lucio 2001, Honkanen et al. 2010). Higher diversity results in a more complex ecosystem with larger heterogeneity and more niches for feeding and nesting. The effect of road densities has been also documented by studies showing that birds respond negatively to the presence of roads and high traffic density (Whited et al. 2000, Brotons and Herrando 2001, Benítez-López et al. 2010, Kociolek et al. 2011) due to a number of direct and indirect effects such as road mortality, traffic noise and habitat destruction (van der Zande et al. 1980, Reijnen et al. 1997, Kuitunen et al. 1998, Bautista et al. 2004, Kociolek et al. 2011, Summers et al. 2011).

Although the models explained a satisfactory degree of the observed variance of the species richness, it is obvious that other factors might be influencing the birds which were not included in these models. This is especially true for the waterbirds for example, such as the Black-tailed Godwit (Limosa limosa) and the Greater Flamingo (Phoenicopterus ruber) for which variables related to water bodies (such as water depth and type of sediment) are important for their distribution (Granadeiro et al. 2007, Bolduc and Afton 2008).

It should be also noted that as in the case of Kakamega Forest, the model used to test the effects of the various predictors on the different bird categories was refined using the overall species richness. Although the removal of the effects of road densities on
Developing and testing the path model for Cyprus

vegetation cover and habitat diversity would not have influenced model re-specification even if every category was examined independently, mean altitude may have had an effect on some of the bird categories tested, in addition to the three important variables that were identified.

Overall, the results of this analysis have important conservation implications for the bird species of the island. The strong negative impact of road densities on certain species categories should be taken into account by policy makers when designing or upgrading the road network in the island. For example, based on the results, the new major road which is planned to be built adjacent the narrow strip of land of the Periochi Polis-Gialia (CY4000001), one of the SACs used in this analysis, is likely to have a negative impact on the bird communities of the site which is rich both in overall bird species (n=46) and species included in the Annex I of the Birds Directive (n=21).

The importance of the diversity of habitats in the models shows that is necessary for the management authorities to monitor the quality of the habitats within the sites and ensure that they are not negatively affected by human activities. Intensification of agriculture or land abandonment for example (a common problem in Cyprus due to increasing rates of urbanization) have been shown to contribute to the homogenization of habitats, influencing diversity with potentially negatively effects on bird species richness (Sirami et al. 2008, Kati et al. 2009, Geri et al. 2010). The size of the area is of lesser importance for management purposes, as areas of the protected sites are not expected to change drastically but nevertheless it is an important factor to keep in mind for the designation of future areas.

Overall, the approach used in this study was useful in identifying and quantifying some of the major variables influencing the bird communities in the Special Areas of Conservation (SACs) in Cyprus. The results of this study are important not only for the avifauna of Cyprus but also of the wider region considering that the majority of the species included in the analysis are migratory birds.

Moreover, as in the case of Kakamega Forest, it was shown that SEM can be successfully used to link disparate sources of data already available in order to extract
useful wider ecosystem conclusions. This is especially important for areas such as Cyprus for which considerable amount of data is available but the wildlife of the region remains still mostly understudied.
CHAPTER 4

THE INFLUENCE OF SPECIES TRAITS ON THE PRESENCE OF BIRDS IN NATURA 2000 SITES IN CYRPUSS
4.1 ABSTRACT

The effects of the various habitat variables on species richness and presence of birds are likely to be influenced by specific species-traits such as body size and habitat specialization, which possibly affect the birds’ responses to changes in those habitat characteristics. In this chapter I use generalized mixed model analysis to identify how specific traits influence the presence of birds in the Natura 2000 sites in Cyprus, in relation with the habitat characteristics identified in Chapter 3. Species with higher regional population numbers and wider habitat breadth are generally more present while species with smaller clutch size and longer incubation periods are usually found in larger areas. Species with large body size are found in areas with higher habitat diversity showing that certain species traits may make birds more vulnerable to changes in those habitat characteristics.

4.2 INTRODUCTION

The results of the structural equation modelling (SEM) analysis of Chapter 3 suggest that overall bird species richness in Natura 2000 sites in Cyprus is influenced by the size of the area, the habitat diversity, and the road density within each protected area. The impact of these three habitat variables changes as a function of the categories of the birds examined. For example, while overall richness is influenced more by the size of the area, the richness of the waterbirds is influenced more by road density.

While these results give us important information on some of the main factors that are influencing the species richness of various bird guilds in the protected areas in Cyprus, we must acknowledge that richness is just one of the possible measures of diversity, which can be used to describe the bird communities in an area. From these results it is difficult to identify which particular species are influenced most and impossible to identify how species traits influence birds’ susceptibility to changes in habitat variables. The latter can be of considerable ecological and conservation importance since this complementary information would allow us to better understand the underlying ecological mechanisms influencing bird communities and would allow managers to shape their conservation actions accordingly to protect species at higher risk.
The influence of species traits on the presence of birds in Natura 2000 sites

The importance of identifying how species traits influence the responses of birds to certain habitat characteristics and how they affect their vulnerability to anthropogenic disturbance has been the topic of several studies. Researchers have been especially interested in identifying species traits that influence species presence, population growth, vulnerability to habitat disturbance and extinction risk (e.g. Bennett and Owens 1997, Owens et al. 1999, Owens and Bennett 2000, Sæther et al. 2004, Jiguet et al. 2007, Reif et al. 2010, Sæther et al. 2011).

We know from previous studies that extinction risk and vulnerability to disturbance is not randomly distributed among the extant bird species but, on the contrary, certain traits make birds more susceptible to extinction (Bennett and Owens 1997, Amano and Yamaura 2007, Gray et al. 2007). Although the exact mechanisms are not always fully understood yet (Beissinger 2000), factors such as body size, population size, range and fecundity rates have been shown to have a strong impact (Bennett and Owens 1997).

In general, studies have shown that life-history characteristics and other species traits could be used to describe bird communities in certain areas. For example, Galván and Benayas (2010) found that the bird densities in pine plantations and woodlands under secondary succession in central Spain were explained by a number of behavioral, life-history and phylogenetic variables such us territoriality, sexual dichromatism, and egg mass. Similarly, Hansen and Urban (1992) showed that species characteristics such as number of eggs per year and nest type of birds found in Eastern Deciduous and Pacific Northwestern forests, in USA, can be used to understand their response to landscape changes and to assess their sensitivity to forest fragmentation.

Foraging behavior and nest placement have been also shown to explain changes in bird communities and are helpful in revealing the mechanisms by which the birds are influenced (Bishop and Myers 2005, Davies et al. 2010). Species-trait analysis has shown that the expansion of the artificial water network in arid zones in Australia (for enhancing livestock grazing) significantly affected birds that forage and nest on the ground because of increased trampling and a decrease in vegetation cover (Davies et al. 2010). Declines in population changes of birds in Netherlands were influenced by...
nest height and migratory behavior with late arrival birds being more vulnerable to landscape changes (Van Turnhout et al. 2010).

Other species traits, such as habitat preference have also been shown to affect the response of birds to disturbance. Seoane et al. (2011) found that variation in threat of birds in the Canary Islands was explained to a large extent by habitat preference and several phylogenetic characteristics. Reed (1999) explains that habitat generalists and species with large distributions and population densities are less vulnerable to extinction pressures.

Studying how various species characteristics affect the influence of habitat variables on bird communities is important for: a) better understanding the underlying ecological processes that influence the bird communities in an area and b) developing more effective conservation measures since management decisions can be based on the species characteristics that are most important.

Using generalized linear mixed modelling I examine the impact of several morphological, behavioral and life-history characteristics on the presence/absence of birds in protected areas in Cyprus. I studied the interactions between those traits and the three important habitat variables already identified in Chapter 4 to understand how those characteristics might be influencing the distribution of birds in those sites.

4.3 METHODS

4.3.1 Study Area

For the purposes of this analysis I used the same thirty-eight Natura 2000 sites found in Cyprus that were used for the analysis in Chapter 3 (Figure 3.4). A detailed description of the study area and the study sites can be found in Chapter 3 under the Methods section.
4.3.2 Data on habitat variables and species traits

Data on the presence/absence of birds in the protected areas were extracted from the BioCyprus Database as explained in Chapter 3. In total 248 different species were recorded in the thirty-eight protected areas. For each of the study sites a list of all the species was included assigning 1 or 0 to each species depending on whether they were recorded in the particular site or not. For the analysis, I used the three main habitat variables that were identified as important determinants of species richness observed in the Natura 2000 Sites. All three variables were extracted through GIS analysis using ArcMap 9.2 (ESRI 2009) as explained in Chapter 3.

Data on body length, clutch size and incubation periods were collected from Snow and Perrins (1998) and supplemented where necessary using relevant data from Cramp (1977–1994). I only used the body length measurements of males for which complete information was available. For most species, a range of body length, clutch size, and incubation periods was reported in the literature which was averaged to obtain the mean value for each trait for each species.

In addition to these traits I collected data on the number of breeding pairs within the European Union (EU25) using the information found in BirdLife International (2004). The number of breeding pairs was used as an indicator of the regional population abundance of species which are likely to affect the presence/absence of birds in the protected sites. Using the habitat lists in Tucker and Evans (1997), and complementary information from Cramp (1977–1994) when necessary, I estimated the Habitat Breadth of each species by adding the number of habitats they were reported to be found in using the eight different reported categories: 1. Marine Habitats, 2. Coastal Habitats, 3. Inland Wetlands, 4. Tundra, Mires and Moorland, 5. Lowland Atlantic Heathland, 6. Boreal and Temperate Forests, 7. Mediterranean Forest, Shrubland and Rocky Habitats, 8. Agricultural and Grassland Habitats. Habitat breadth for each species ranged from 1 (i.e. the species was only found in one habitat therefore has narrow breadth) to 5 (wide breadth). All variables besides habitat diversity and habitat breadth were log transformed to linearize data. Twenty species, for which there was not enough information on one or more of trait characteristics, were excluded from the analysis.
4.3.3 Statistical Analysis

Data were analyzed using generalized linear mixed modelling (GLMM) with binomial error structure (logit link) with parameters using Maximum Likelihood. Analysis was conducted in R (R Development Core Team, 2011) using the “lme4” package (Bates et al. 2011). This type of analysis is suitable for longitudinal data on presence/absence of species for which random effects are present (Zuur et al. 2009). The study sites and the species were treated as random effects while the three habitat variables along with the species traits were treated as fixed. The interaction effects between all the response variables were examined and the original model was simplified in stepwise fashion by removing all non-significant interactions to obtain the minimal adequate model (Crawley 2007). At each step, the subsequent models were compared using ANOVA to assess whether simplification was justified (Crawley 2007). Results were plotted using the plotLMER function of package “languageR” (Baayen 2011).

To get an estimate of how much variance was explained by the fixed effects in the two models I used the formula suggested by Xu (2003) to calculate the Ω² value for each model by dividing the model’s variance against the total variance of the corresponding null model (which includes only the intercept and the two random effects) and subtracting that from 1.

\[ Ω^2 = 1 - \frac{\sigma^2}{\sigma_0^2} \]

4.4 RESULTS

GLMM results (Tables 4.1 and 4.2) confirm the findings of the structural equation modelling analysis showing that habitat diversity, road density and the size of the protected site influence the bird communities in Natura 2000 in Cyprus. The Ω² value is 0.41 showing that a considerable percentage of variance is explained by the model. The probability of presence is always higher in larger areas and as expected road densities negatively influence the presence of the species examined (Figure 4.1 a,b,c). There were significant two-way interactions between the three habitat variables considered. Interestingly, Figures 4.1a and 4.1b show that the positive impacts of
habitat diversity are reversed in sites where road densities are high or have small areas indicating that those two variables have a stronger impact on the bird communities. Moreover, the negative impact of road density on birds is stronger in smaller areas compared to larger (4.1c).

Table 4.1 Results of the GLMM analysis showing the variance explained by the random effects

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species (n=228)</td>
<td>2.74292</td>
<td>1.65618</td>
</tr>
<tr>
<td>Sites (n=38)</td>
<td>0.28585</td>
<td>0.53464</td>
</tr>
</tbody>
</table>

AIC: 6978, BIC: 7112, Deviance: 6940; Residual deviance for binomial set at 1.0

Table 4.2 Results of the GLMM analysis showing the importance of the independent variables and their interactions on the presence/absence of species

| Fixed Effects                      | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------------------------|----------|------------|---------|----------|
| (Intercept)                        | 4.38     | 2.29       | 1.92    | 0.055519 |
| Habitat Diversity                  | -2.41    | 0.63       | -3.81   | 0.000141 *** |
| Road Density                       | 2.83     | 1.19       | 2.38    | 0.017383 * |
| Area                               | -1.30    | 0.54       | -2.41   | 0.015765 * |
| Incubation Period                  | -1.71    | 0.65       | -2.61   | 0.008960 ** |
| Clutch Size                        | 0.63     | 0.39       | 1.61    | 0.107046 |
| # of EU(25) Pairs                  | 0.03     | 0.09       | 0.33    | 0.738306 |
| Body Length                        | -0.41    | 0.35       | -1.16   | 0.245042 |
| Habitat Breadth                    | -0.99    | 0.49       | -2.01   | 0.044928 * |
| Habitat Div:Road Density           | -1.71    | 0.47       | -3.61   | 0.000305 *** |
| Habitat Div:Area                   | 0.89     | 0.22       | 4.10    | 4.06e-05 *** |
| Habitat Div:Body Length            | 0.43     | 0.09       | 4.63    | 3.67e-06 *** |
| Road Density:Area                  | 0.62     | 0.31       | 1.98    | 0.047356 * |
| Road:Incubation Period             | -1.30    | 0.27       | -4.73   | 2.20e-06 *** |
| Area:Incubation Period             | 0.29     | 0.12       | 2.40    | 0.016428 * |
| Area:Clutch Size                   | -0.20    | 0.09       | -2.11   | 0.034529 * |
| EU Pairs:Habitat Breadth           | 0.10     | 0.04       | 2.49    | 0.012945 * |

Significance levels: *** 0.001, ** 0.01, * 0.05
The influence of species traits on the presence of birds in Natura 2000 sites

Figure 4.1 Graphs depicting the interaction effects between the three habitat variables and between the number of EU(25) pairs and habitat breadth.

The probability of presence of a species is also influenced by incubation period, clutch size, body length, habitat breadth and number of breeding pairs in EU(25). Incubation period interacts with two of the three habitat variables considered. Figure 4.2a shows that although species with longer incubation periods are less likely to be present in general, the probability of absence is higher in areas where road densities are higher. Also, higher road densities seem to impact more adversely the species with longer incubation periods (steeper slope). Species with longer incubation periods are more likely to be present in sites with larger areas (Figure 4.2b). Likewise, species with smaller clutch sizes are more likely to be present in larger sites (Figure 4.2c).
Large bodied species are less likely to be present in areas with low habitat diversity, indicating that larger species may require high levels of diversity (Figure 4.2d).

The only significant two-way interaction among the species traits included in the analysis was between the numbers of EU(25) breeding pairs and the habitat breadth showing that species with more pairs and wider breadth are more likely to be present in the study sites (4.1d).

4.5 DISCUSSION

The results of this analysis are in accordance to the findings of the structural equation modelling analysis, thus confirming the impact of the three habitat variables.
identified, and allow us to better understand how these specific variables affect species presence by identifying the species traits that determine presence/absence. An additional important conclusion that was not revealed through structural equation modelling (but is clear by examining the interaction effects between the three habitat variables) is that the negative effects of high road densities and small size of the areas are strong enough to offset any positive effects on species presence from increased habitat diversity (Figure 4.1 a,b). Along the same lines, the impact of roads seems to be more adverse in smaller areas (Figure 4.1c). This shows that birds found in smaller areas with higher road densities might be at higher risk, something that must be taken into account by the authorities responsible for managing the protected areas in Cyprus.

Regarding the influence of various species traits on the bird species presence, it was expected based on previous literature that species with small regional population sizes, narrow habitat breadth and lower increase rates (i.e. longer incubation rates, smaller clutch sizes and larger bodies) would be more influenced by road densities, smaller areas and possibly benefit more from higher habitat diversity. It should be noted though that often the exact effects of a trait might be case and species specific and therefore not widely generalizable.

For example in the case of body size it has been shown by Owens and Bennett (2000) that large body size is associated with higher extinction risk in cases where birds are threatened by harvesting, egg collecting and trade but was also associated with lower extinction risk in cases of habitat loss. Also, in another study large species were shown to be more vulnerable to extinction risk at low population sizes but smaller species more vulnerable at high population sizes, because large body size is usually associated with slow increase rates which make birds more vulnerable to disturbances but also with higher longevity which decreases extinction risk (Pimm et al. 1988).

These complicated mechanisms make it difficult in some cases to disentangle the exact effect of species traits and to identify clear and useful trends because of simultaneous effects in the opposite direction. This might be more difficult through the type of analysis used in this chapter in which species with different habitat, foraging, nesting and other requirements are grouped together, possibly confounding
the effects of particular traits. Yet, for conservation purposes and for more effective management of the specific protected areas it is important to examine whether it is possible to identify general trends that indicate how certain species characteristics might be influencing overall species richness.

One such obvious trend is the observed positive interaction between the number of breeding pairs in the EU(25) (which was used as an indicator of regional population size) and habitat breadth, which both are increasing the probability of presence of bird species in Natura 2000 sites. This was expected as it has been shown in other studies that: a) regional abundance influences local population sizes, presence, and extinction risk (O'Grady et al. 2004, Freckleton et al. 2006, Sæther et al. 2011) and b) wider habitat breadth positively influences the bird communities and reduces their extinction risk (Julliard et al. 2004, Norris and Harper 2004).

Another obvious trend is that species with longer incubation periods are less likely to be present in Natura sites and interestingly these species seem to be more sensitive to increases in road densities (Figure 4.2 a,b). This can be explained by the fact that species with longer incubation rates usually have slower increase rates making them more vulnerable to habitat disturbance (Pimm et al. 1988). Although this effect was not observed for species with small clutch sizes, which for the same reason (i.e. slow increase rates) are usually considered more prone to extinction risk (Pimm et al. 1988, Bennett and Owens 1997), results showed that species with small clutch sizes and longer incubation periods are less likely to be present in smaller areas (Figure 4.2 b, c) indicating that these species need larger areas.

The lack of interaction between clutch sizes and road densities could be due to the fact that other factors influence this particular species trait such as diet for example, nest type (Jetz et al. 2008) and migratory behaviour (Böhning-Gaese et al. 2000). This may cause certain species with similar clutch sizes to respond differently to habitat disturbance (Gray et al. 2007) thus confounding the effects of this trait.

One would expect that large bodied species would also need larger areas and be more negatively affected by road densities as they are usually associated with larger ranges (Bennett and Owens 1997) and also slower increase rates (Pimm et al. 1988).
However, the results of this analysis (and specifically the lack of interaction between body size and area and road densities) show that all sized birds are influenced similarly by these two factors (negatively and positively respectively). A possible reason for why larger bodied birds do not necessarily require larger areas and less road disturbance, as one would expect, is because many of those species are waterbirds, such as the Great White Pelican (*Pelecanus onocrotalus*), the Whooper Swan (*Cygnus cygnus*) and the Greater Flamingo (*Phoenicopterus ruber*), which are mostly found in small wetlands. Nevertheless, based on the results larger bodied species are less likely to be present in areas where diversity is low (Figure 4.2d). Whether this is because higher diversities offer more resources or just because they indicate higher quality habitats it is not clear. Nevertheless, the importance of maintaining high diversity for larger bodied species is something that should be taken into account when management plans are implemented for these areas.

### 4.4.1 Conclusions

To summarize, trait-based research has been used extensively in last years to better understand the factors that influence bird communities and it can be useful in informing conservation decisions (Pocock 2011). Despite some limitations in its potential to fully disentangle the underlying ecological mechanisms and despite the fact that some of these mechanisms still remain unclear (Beissinger 2000, Pocock 2011) several conclusions with important conservation potential can be drawn from the results of such studies. In this case it is evident that:

a. The negative effects of road densities are stronger in smaller areas compared to larger.

b. The positive effects of increased habitat diversity are reduced and even reversed in sites with high road densities and small area size.

c. High regional population sizes and wide habitat breadth increases the probability of presence of species.

d. Species with long incubation periods and small clutch sizes need larger areas; with the former being more influenced by increased road densities.

e. Species with large body lengths require higher habitat diversity.
Any management actions aiming at conserving the avifauna in the protected areas in Cyprus should focus on the impact of road densities in protected areas especially in smaller sites and also on retaining habitat quality by preserving habitat diversity. It should be taken into account that species with long incubation periods, small clutch and large bodies might be more sensitive to these changes, as well as species with narrow habitat breadth and low regional population sizes.
CHAPTER 5

THE INFLUENCE OF SPECIES TRAITS ON THE PRESENCE AND ABUNDANCES OF BIRDS IN KAKAMEGA FOREST
5.1 ABSTRACT

As shown in Chapter 4, the effects of habitat variables on the presence of birds are influenced by species-specific traits. In this chapter I use linear mixed models to examine how traits such as body length, clutch size and degree of forest specialization interact with habitat disturbance and the vegetation characteristics identified through SEM analysis to determine bird species richness and abundance in Kakamega Forest. Results show that although overall the patterns observed in Chapter 2 are supported, the interspecific variation in presence and abundance is too high to be able to extract meaningful conclusions about the influence of the particular traits examined and the impact of the habitat variables on individual bird species in Kakamega.

5.2 INTRODUCTION

Understanding which factors influence the bird communities of the forest and in what ways is important for the conservation of its rich avifauna which not only performs important ecosystem services necessary for the preservation of the forest, such as seed dispersion and plant pollination (Sekercioglu et al. 2002, Sekercioglu et al. 2004), but it also has an important economic value and high potential for ecotourism since the forest is a well-known hot spot for bird watching and it is considered an Important Bird Area as defined by BirdLife International (Bennun and Njoroge 2000).

Results of the structural equation modelling (SEM) analysis in Chapter 2 showed that although the effects of the variables considered are guild specific, overall the bird communities in Kakamega Forest are strongly influenced by the vegetation structure and composition of the forest and the distance to the forest edge. Birds are also influenced, mostly indirectly, by the habitat disturbance measured as the number of tree stumps which has a strong negative effect on the vegetation of the forest. The distance to the forest edge and the plant diversity have a positive effect on all the bird guilds examined except the forest visitors, which according to Bennun et al. (1996) they are usually non-forest birds and their presence could be due to habitat disturbance. Indeed, SEM results showed that the richness and abundances of forest visitors were higher in areas found closer to the forest edge and in areas with lower plant diversity and higher levels of anthropogenic habitat disturbance.
To disentangle the exact effects of these factors on the bird communities and to better understand how they are affecting the species richness and abundances I used linear mixed effect model analysis, as in Chapter 4, to study how different species traits interact with the habitat variables previously identified to influence birds in Kakamega Forest. This type of analysis is important for better understanding the factors that determine the bird distributions in the forest, as we know from the results of Chapter 4, and also from the existing literature that the response of birds to land changes and habitat characteristics are likely to be influenced by specific species traits such as body size, mobility and specialization (Gray et al. 2007, Sekercioglu and Sodhi 2007).

In addition, results of Chapter 4 showed that this kind of analysis can give us not only important insights on which traits are important for understanding the observed patterns in bird distributions but it can also provide us with useful information on how the various habitat variables identified through structural equation modelling influence each other, therefore understanding better how the underlying ecosystem mechanisms work.

Since in the case of Kakamega Forest there were data on both richness and abundances of birds in each site, I used two different linear mixed models to study how these factors are influencing a) the probability of presence of birds and b) their abundances.

5.3 METHODS

5.3.1 Study Area

For the purposes of this analysis I used the same twenty-two sites in Kakamega Forest as in Chapter 2. For a full description of the forest and the study sites see Chapter 2.

5.3.2 Data habitat variables and species traits

Individual species abundances at each site were calculated using the same two datasets as in Chapter 2 collected at the various study sites in Kakamega Forest for the
purposes of the Farwig et al. 2006b and 2008c studies. The first dataset (Dataset 1) includes bird counts from 9 areas all sampled once a month for a total of 12 months in 2000/2001. The second dataset (Dataset 2) includes bird counts from 15 areas all sampled once a month for a total of 13 months in 2005/2006. As mentioned in Chapter 2 two of the study sites (Buyangu Hill and Colobus) were common in both datasets.

Following the same method as in Chapter 2, species abundances were estimated for each sampling month and then averaged to obtain an overall value for each species at each site. For the probability of presence analysis all species observed at each site during any of the sampling months were considered as present. If a species was not sighted in any of the sampling months it was recorded as absent from that site. It should be noted that compared to the species richness values used in SEM analysis in Chapter 2 (where the average number of the total species richness was estimated for each site) this approach yields a higher number of the species per site because it estimates the cumulative number of birds that were recorded compared to the average number that one would expect to be present at any given instance.

For the two sites that were common in both datasets the abundances of each species were averaged as in Chapter 2. Since this was not possible for the presence/absence analysis, and to avoid overestimating presence by combining the two datasets, I used the sample function in R (R Development Core Team, 2011) to randomly select one of the two datasets for each site. Then I rerun the analysis using the other dataset to ensure that results were in accordance with each other (see Results section). The first analysis was done using the presence and absence from Dataset 2 for both sites, while the second one using Dataset 1 again for both sites.

For the two linear mixed effects models (presence/absence and abundance) I used the same five habitat variables that were used in SEM. Those variables were: canopy openness, foliage height diversity, plant diversity, distance to the forest edge and number of stumps. For a detailed description on how these data were collected see Chapter 2. As in the case of the path analysis the variables were transformed to linearize the data. All variables were log transformed except the plant diversity variable which was exponentially transformed.
Data on the individual species traits were collected using the ornithological handbook “The Birds of Africa” (Brown et al. 1982, Urban et al. 1986, Fry et al. 1988, Keith et al. 1992, Urban et al. 1997, Fry and Keith 2000, Fry et al. 2004) and were supplemented when necessary using the “Handbook of the birds of the World” (Del Hoyo et al. 1992-2002). Data collected included body length, clutch size, incubation periods, fledging periods and nest height. Unfortunately, for the latter three traits there was not enough information for all the birds to run the analysis since data were not available for 24 to 76 species depending on the particular trait. For that reason, in the final analysis only body length and clutch size were included for a total of 108 bird species resulting into 2376 observations.

The degree of forest specialization (i.e. Forest Specialists (FS), Forest Generalists (FG) and Forest Visitors (FV)) based on the Bennun et al. (1996) classification, was also included since we know from the SEM analysis that birds are influenced by the habitat variables in different ways based on their guild, sometimes even in opposite direction. Twelve species of birds that were not included in the Bennun et al. (1996) classification, such as the Martial Eagle (*Polemaetus bellicosus*) were listed as non-forest (NF) birds following the classification scheme of Mulwa (unpublished report).

Table 5.1 shows the number of species for each forest guild included.

<table>
<thead>
<tr>
<th>Bird Guilds</th>
<th># Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest Specialists (FS)</td>
<td>33</td>
</tr>
<tr>
<td>Forest Generalists (FG)</td>
<td>34</td>
</tr>
<tr>
<td>Forest Visitors (FV)</td>
<td>29</td>
</tr>
<tr>
<td>Non- Forest Birds (NF)</td>
<td>12</td>
</tr>
</tbody>
</table>

5.3.3 Statistical Analysis

As in Chapter 4, data on presence and absence were analyzed using a generalized linear mixed model (GLMM) with binomial error structure (logit link), with parameters using Maximum Likelihood (ML). Species abundances were analyzed using a linear mixed model (using ML), by including all species with abundance of more than zero and log transforming the response variable to approximate normal distribution, resulting into 904 observations in total. All analyses were conducted in R (R Development Core Team, 2011) using the “lme4” package (Bates et al. 2011).
The study sites and the species were treated as random effects while the habitat variables along with the species traits were treated as fixed. The interaction effects between all the response variables were examined and the original model was simplified in stepwise fashion by removing all non-significant interactions to obtain the minimal adequate model (Crawley 2007). At each step, the subsequent models were compared using ANOVA to assess whether simplification was justified (Crawley 2007). Results were plotted using the plotLMER function of package “languageR” (Baayen 2011). The p-values reported for the linear mixed model analysis of species abundances were estimated through a “Markov chain Monte Carlo (MCMC) sampling from the posterior distribution of the parameters”, using the pvals function in “languageR” package as suggested by Baayen et al. (2008).

To estimate the proportion of variance explained by the fixed effects in the two models I used the formula suggested by Xu (2003) to calculate the $\Omega^2$ value for each model:

$$\Omega^2 = 1 - \frac{\sigma^2}{\sigma_o^2}$$

where $\sigma^2$ is the model’s variance and $\sigma_o^2$ the variance of the null model (which includes only the intercept and the random effects).

Additionally, for the abundance model I plotted the observed abundances against the fitted and predicted abundance values and calculated the resulting $R^2$ value against the 1:1 line (slope=1, intercept=0) to estimate the percentage of variation of the observed abundance explained by the model (Piñeiro et al. 2008). The fitted values were obtained using the “fitted” function in “lme4”, which also takes into account the random effects, while the predicted values were calculated using the parameter estimates of the output thus ignoring the random effects and predicting for the average species and average site.
5.4 RESULTS

5.4.1 Presence/Absence

Analysis using the two different datasets for the two common sites resulted into the same minimal adequate model therefore only the first model (based on Dataset 2) is presented and discussed. The results of the second model (based on Dataset 1) are reported in Appendix IV (Tables A.3, A.4). The $\Omega^2$ value of the model was 0.10, showing that the fixed effects included explain only a small percentage of the individual species variation of presence/absence. Most of the variation in the model is explained by the random effects and particularly the “Species” effect (Table 5.2).

Table 5.2 Variance explained by the random effects in the presence/absence model

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species (n=108)</td>
<td>4.1625</td>
<td>2.0402</td>
</tr>
<tr>
<td>Sites (n=22)</td>
<td>9.7953e-10</td>
<td>3.1298e-05</td>
</tr>
</tbody>
</table>

Residual Variance: 1.0, AIC: 2124, BIC: 2251, Deviance: 2080

As in SEM analysis, canopy openness does not significantly influence the bird communities of the forest and there is no significant effect of body length on the probability of presence of the species. The rest of the factors included influence the birds to a certain degree although the p-values reported are not statistically significant for all the bird guilds (Table 5.3). For example the distance to the forest edge does not influence significantly the presence of the specialists despite the fact that structural equation modelling shows that this variable influences both the richness and the abundance of the particular guild.
Table 5.3 Results showing the importance of the depended variables and their interactions on presence/absence of the species

|                          | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------------|----------|------------|---------|---------|
| Intercept                | 0.37     | 2.72       | 0.136   | 0.891567 |
| Plant Diversity          | 0.05     | 0.02       | 2.638   | 0.008345 ** |
| Foliage Height Diversities| -4.49   | 1.67       | -2.688  | 0.007192 ** |
| Number of Stumps         | 0.94     | 0.38       | 2.436   | 0.014833 * |
| Distance to Forest Edge  | 1.15     | 0.33       | 3.482   | 0.000498 *** |
| Clutch Size              | -3.49    | 2.39       | -1.460  | 0.144221 |
| Forest Guild (FS)        | 2.48     | 1.69       | 1.466   | 0.142772 |
| Forest Guild (FV)        | 2.50     | 1.89       | 1.332   | 0.182890 |
| Forest Guild (NF)        | 3.21     | 2.99       | 1.075   | 0.282479 |
| Plant Diversity:Forest Guild FS | 0.03 | 0.02       | 1.065   | 0.286827 |
| Plant Diversity:Forest Guild FV | -0.10 | 0.03       | -3.396  | 0.000684 *** |
| Plant Diversity:Forest Guild NF | -0.10 | 0.05       | -2.112  | 0.034644 * |
| FHD:Clutch Size          | 4.45     | 1.73       | 2.568   | 0.010216 * |
| Stumps:Forest Guild FS   | -1.10    | 0.50       | -2.181  | 0.029172 * |
| Stumps:Forest Guild FV   | -0.33    | 0.55       | -0.592  | 0.554071 |
| Stumps:Forest Guild NF   | -1.46    | 0.90       | -1.615  | 0.106265 |
| Distance FE:Clutch Size  | -1.08    | 0.32       | -3.422  | 0.000621 *** |
| Distance FE:Forest Guild FS | 0.09   | 0.25       | 0.366   | 0.714622 |
| Distance FE:Forest Guild FV | -0.81 | 0.27       | -3.019  | 0.002536 ** |
| Distance FE:Forest Guild NF | -0.45 | 0.45       | -1.017  | 0.309011 |

Significance levels: *** 0.001, ** 0.01, * 0.05

Figure 5.1 a,b shows that the direction of the impact of the habitat variables on birds is in accordance with the results of the SEM analysis. For example, plant diversity and distance to the forest edge influence positively the presence of the forest specialists and generalists and negatively the presence of the forest visitors. Non-forest birds are also influenced negatively by these two variables. The direct impact of the number of stumps also agrees with that reported in structural equation modelling with forest specialists influenced negatively and the rest of the guilds influenced positively (Figure 5.1c). It should be noted though that the corresponding path coefficients were only significant for forest visitors’ abundance while in this case of the GLMM analysis p-values are only significant for forest generalists and specialists.
Based on the results of this analysis, there is a significant interaction effect between clutch size and two of the habitat variables: a) foliage height diversity and b) distance to the forest edge (Figure 5.2 a,b). The probability of presence is generally higher for species with larger clutch sizes and species with small clutch sizes are usually found in areas with low foliage height diversity and areas deeper in the forest (Figure 5.2 a,b).
The influence of species traits on the presence and abundances of birds in Kakamega

Figure 5.2 Graph depicting the interaction effect of clutch size with foliage height diversity and distance to the forest edge on the probability of presence of the species

5.4.2. Abundance model

The $\Omega^2$ value for the abundance model is also extremely low (0.08) confirming that interspecies variation is very high. As before, the species random effect explains a large amount of the variance compared to sites which is very close to zero (Table 5.4). When the observed abundances were plotted against the fitted values an $R^2$ value of 0.73 was obtained. However, when the parameter estimates were used to predict abundances in the forest (by taking into account only the fixed effects) the $R^2$ was only 0.04.

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species (n=108)</td>
<td>1.078331</td>
<td>1.03843</td>
</tr>
<tr>
<td>Sites (n=22)</td>
<td>0.014284</td>
<td>0.11951</td>
</tr>
<tr>
<td>Residual</td>
<td>0.591355</td>
<td>0.76900</td>
</tr>
</tbody>
</table>

As in the presence/absence model the canopy openness and the body length have no impact on the species abundances. Plant diversity has a positive impact on forest specialists and generalists and a negative impact on the visitors (Table 5.5, Figure 5.3a). The same is true for foliage height diversity (Figure 5.3b), a result also shown in SEM analysis although the path coefficient was only significant for the abundances of the visitors. Surprisingly, distance to the forest edge does not seem to significantly influence the abundances of the birds in the forest so it was removed from the model.
The influence of species traits on the presence and abundances of birds in Kakamega

despite the fact that the direction was in accordance with the SEM results. The strong interaction between clutch size and the two habitat variables identified in the presence/absence model was not significant in this case but clutch size seems to interact with the number of stumps. Species with smaller clutch sizes are generally more abundant as before and they are usually found in areas with lower levels of habitat disturbance (Figure 5.3 d)

Figure 5.3 Graphs depicting the interaction effects between the species traits and the habitat variables on bird abundances
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-4.93</td>
<td>1.57</td>
<td>-3.143</td>
<td>0.0017**</td>
</tr>
<tr>
<td>Plant Diversity</td>
<td>0.002</td>
<td>0.01</td>
<td>0.260</td>
<td>0.7946</td>
</tr>
<tr>
<td>Foliage Height Diversities</td>
<td>0.61</td>
<td>0.67</td>
<td>0.905</td>
<td>0.3655</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>1.01</td>
<td>0.37</td>
<td>2.766</td>
<td>0.0058**</td>
</tr>
<tr>
<td>Clutch Size</td>
<td>2.36</td>
<td>0.85</td>
<td>2.759</td>
<td>0.0059**</td>
</tr>
<tr>
<td>Forest Guild (FS)</td>
<td>-1.34</td>
<td>1.60</td>
<td>-0.834</td>
<td>0.4048</td>
</tr>
<tr>
<td>Forest Guild (FV)</td>
<td>-0.81</td>
<td>2.11</td>
<td>-0.382</td>
<td>0.7029</td>
</tr>
<tr>
<td>Forest Guild (NF)</td>
<td>-8.77</td>
<td>4.18</td>
<td>-2.098</td>
<td>0.0362*</td>
</tr>
<tr>
<td>Plant Diversity:Forest Guild FS</td>
<td>0.02</td>
<td>0.01</td>
<td>1.427</td>
<td>0.1538</td>
</tr>
<tr>
<td>Plant Diversity:Forest Guild FV</td>
<td>-0.05</td>
<td>0.01</td>
<td>-2.639</td>
<td>0.0085**</td>
</tr>
<tr>
<td>Plant Diversity:Forest Guild NF</td>
<td>0.02</td>
<td>0.03</td>
<td>0.731</td>
<td>0.4650</td>
</tr>
<tr>
<td>FHD:Forest Guild FS</td>
<td>1.67</td>
<td>0.78</td>
<td>2.129</td>
<td>0.0335*</td>
</tr>
<tr>
<td>FHD:Forest Guild FV</td>
<td>-1.21</td>
<td>1.05</td>
<td>-1.155</td>
<td>0.2484</td>
</tr>
<tr>
<td>FHD:Forest Guild NF</td>
<td>4.61</td>
<td>2.11</td>
<td>2.181</td>
<td>0.0294*</td>
</tr>
<tr>
<td>Stumps:Clutch Size</td>
<td>-0.83</td>
<td>0.32</td>
<td>-2.585</td>
<td>0.0099**</td>
</tr>
<tr>
<td>Stumps:Forest Guild FS</td>
<td>-0.41</td>
<td>0.26</td>
<td>-1.565</td>
<td>0.1178</td>
</tr>
<tr>
<td>Stumps:Forest Guild FV</td>
<td>1.036</td>
<td>0.33</td>
<td>3.150</td>
<td>0.0017**</td>
</tr>
<tr>
<td>Stumps:Forest Guild NF</td>
<td>0.63</td>
<td>0.68</td>
<td>0.919</td>
<td>0.3582</td>
</tr>
</tbody>
</table>

Significance levels: *** 0.001, ** 0.01, * 0.05

5.5 DISCUSSION

The results should be interpreted with caution as the interspecific variation is too high to be able to extract meaningful conclusions from the two models. Although overall the patterns resulting from the analysis seem to be in accordance with the results of the SEM analysis (i.e. confirming the positive impact of the vegetation variables on forest bird communities and the negative impact of shorter distances to the forest edge in the case of species richness) the extremely low variance explained by the fixed effects and consequently the low predictability of the models shows that results cannot be used to explain the individual species presence or abundance of birds in Kakamega Forest.

In other words, although, as also identified in Chapter 2, the habitat variables identified through SEM analysis explain a satisfactory percentage of species richness and abundances at the community and guild level, individual species are likely to be influenced by more diverse mechanisms not captured in this analysis. This is indicated by the fact that most of the variance in the two models is explained by the species random effect. On the other hand, the inter-site variation is relatively low showing that results of SEM are generalizable at the site level but not the species level.
This very high variation between species, although it was unexpected, is understandable as we know from results of Chapter 2 and also from previous research that tropical birds respond rather differently to various habitat variables. It’s not only the fact that birds belonging to different guilds may respond in opposite ways to the variables examined, as is the case for forest specialists and forest visitors’ response to distance to the forest edge, but also in other cases there seems to be considerable within guild variation on resource and habitat use (Schleuning et al. 2011). For example, large variation in the sensitivity to forest edge has been observed even within insectivores in three forest types in southeastern Madagascar with the canopy-insectivores being “edge-sensitive” and the sallying-insectivores “edge-preferring” (Watson et al. 2004).

These patterns are likely to be the reason for the observed levels of variation. Yet, it is surprising that the abundance model did not show any significant effects for the distance to the forest edge (neither through an interaction with the forest guild variable nor as a main effect) since in all SEM analyses and also in the presence/absence model this variable was consistently significant at least for some species. A plausible explanation is that the significance has been masked by the large variation in species abundances. This alludes to the fact that the lack of significance of the effect of body length on species presence and abundance may be a result of this large variation, something that needs to be taken into consideration when interpreting the results. Similarly the observed interaction effects of clutch sizes must be interpreted with caution for the same reasons.

In conclusion, the insights gained from this analysis show that although the forest management authorities in Kakamega Forest need to address the levels of disturbance and the quality of the vegetation of the forest in order to conserve the overall community of the birds recorded, more research is needed to understand the factors that are influencing individual species in the forest that might be of conservation interest.
CHAPTER 6

DISCUSSION
6.1 THESIS OVERVIEW

The aim of this thesis was to test whether we could effectively utilize large amount of available data and information from disparate published and unpublished sources to answer new conservation questions and more specifically to study and quantify the factors that influence the distributions of birds in Kakamega Forest, in western Kenya and in the Natura 2000 sites in Cyprus. In Chapter 1 I presented the main methodological approach used in the thesis to link the data and explained the advantages of this method. In Chapter 2 I reviewed the large amount of published studies for Kakamega Forest and developed a path model that illustrated the most important factors influencing the species richness of the birds in the forest. The attempt to parameterize the model was not successful as the necessary information was not available. An updated version of the model, based on experts’ opinion, was parameterized using existing raw data and data I collected myself. I identified habitat variables that determine species richness and abundance of various bird guilds in the area. A similar methodological approach was followed in Chapter 3, and using unpublished data from disparate sources, I parameterized a path model that illustrates the important habitat variables that influence various bird guilds in Natura 2000 sites in Cyprus. In Chapter 4 I conducted a species-trait analysis to study the species characteristics that influence the presence of birds in those sites. Results revealed significant interaction effects between the habitat variables identified in Chapter 3, and also that species with certain life history characteristics are differently influenced by the variables identified. Lastly, in Chapter 5, I studied the influence of certain species traits on the presence and abundance of birds in Kakamega Forest, showing that the inter-species variation of the birds recorded was too high to extract any meaningful conclusions, indicating that individual species are probably influenced by different factors.

6.2 CONSERVATION CONCLUSIONS FOR EACH STUDY REGION

6.2.1 Natura 2000 sites, Cyprus

Cyprus is considered an important migratory corridor for birds supporting a large number of species (Iezekiel et al. 2004). Several protected areas have been established
in the country, under the EU’s Habitats Directive, to protect its biodiversity including its rich avian diversity. Yet, the factors that influence species richness and the presence of birds in those areas are largely unknown since the region is understudied. The results of this project have shown that the birds in the Natura 2000 sites are mostly positively affected by the size of the protected area and its habitat diversity, while they are negatively influenced by the density of the roads within each area; interestingly the negative effects of the road density are higher in smaller areas. Species with certain traits, such as larger incubation periods, smaller clutch size and larger body lengths are influenced differently by the three factors identified and seem to benefit from larger areas with higher habitat diversity and lower road densities. The documented negative effects of the road network and the positive effects of area size and habitat diversity on the species richness and presence of the birds recorded in the protected areas need to be consider by the management authorities in Cyprus to ensure the conservation of the avifauna. Also, the results of the analysis should be confirmed using the rest of the Natura 2000 sites not included in this analysis (the Special Protection Areas) once the data on their habitat diversity become available.

6.2.2 Kakamega Forest, Kenya

Despite the ongoing attempts to protect the biodiversity of the Kakamega Forest habitat destruction is still occurring in the area. The high density of human populations and the large growth rate, which results in a need for more agricultural land around and within the forest (Lung and Schaab 2004) are likely to increase human encroachment and habitat fragmentation and consequently the edge effects, a factor that has been shown in this study to negatively impact birds. In addition the large need for trees as firewood and construction poles (Bleher et al. 2006) along with the other related subsistence activities are likely to increase habitat disturbance and exert additional pressure on the bird communities in the forest since this analysis has shown that disturbance negatively influences the forest’s vegetation with a subsequent negative effect on birds.

The overall impact of continuing habitat degradation on the birds, if not controlled, will impoverish the rich avifauna of the forest, and it is likely to affect the rest of the biodiversity in the region as it is widely accepted that birds perform significant
ecosystem functions as pollinators and dispersers for example (Sekercioglu et al. 2004, Berens et al. 2006). In fact, a large number of tree species in Kakamega Forest rely on birds for their dispersal (Kirika et al. 2008b) such as the Common Bulbul (*Pycnonotus barbatus*), the Black and White casqued Hornbill (*Bycanistes subcylindricus*) (Farwig et al. 2006b) and the Blackcap (*Sylvia atricapilla*) (Schleuning et al. 2011). The documented negative impact may also have an economic consequence on the ecotourism in the region which is comprised mostly of birdwatchers who visit the forest because of its known value as important bird area and its high species richness.

To reduce the impact of the human activities the responsible management authorities need to address the current subsistence needs of the locals by offering viable alternative options to their daily requirements. There have been efforts to promote more efficient cooking devices and paper pulp as fuel to reduce the need for firewood (personal observation) but the success of these interventions are yet to be evaluated. The increasing need for more agricultural land that leads to fragmentation must be also addressed, possibly by improving and promoting more efficient agricultural methods.

Although these measures will undoubtedly benefit the conservation status of the forest’s biodiversity in general, results of Chapter 5 indicate that more research is needed to understand the specific habitat requirements and the factors that influence specific bird species especially those that might be of conservation importance.

### 6.2 PROBLEMS/CAVEATS WITH THE APPROACH

The results of the analyses of this thesis show that with the right tools scientists can take advantage of the large amount of data already available to answer new questions possibly of more complex nature that have not been addressed before. However, it is unlikely that the proposed method of parameterizing a model using only the information in published literature would be practicable for any study area. As shown in the example of Kakamega Forest, important information is missing and this is to be expected as significant relationships used for the development of the model are not always quantified or even examined by the individual studies.
For example in the Kakamega Forest literature there were several studies in which the authors assessed the effects of fragmentation and habitat disturbance on various species by first categorizing qualitative different study sites and then comparing the status of the species between those sites. In such cases, although the relationships reported will be part of the path model, the necessary statistical estimates will not be available. Even more, to construct the necessary correlation matrix that would allow the parameterization of the model, one would also need the correlations values between factors that have not explicitly studied by any of the available individual studies reviewed. In rare cases in which all the information might be available there may be still issues with using different correlation values calculated based on heterogeneous sample sizes making model testing problematical.

Despite the fact that parameterization based on the published literature seems impracticable, the results of the analyses in this thesis show that the method is nevertheless useful in a constructing a path model that successfully illustrates the possible underlying ecosystem mechanisms of interest, based on the conclusions of the individual studies. The model can then be parameterized using raw data, and additional data if needed as in the case of Kakamega Forest. Because the particular forest has been studied so extensively there were large amounts of data that researchers were willing to share and which allowed us conduct the analysis. Collecting all these data in the field would have required large and possibly preventive amounts of money and time.

In conclusion, if the right data from different sources are available it is likely that we can reuse them and link them together to draw new important insights with significant conservation implications, even if we have to complement them with new data, adding value to the old information and reaching new conclusions with conservation importance. Unfortunately though, the great potential of this approach is drastically compromised by the extremely small amount of data currently being made available by scientists.
6.3 IMPORTANCE OF DATA AVAILABILITY

Despite the benefits of data reuse to answer new questions and the potential to contribute to the increasingly data-intensive scientific endeavours the majority of the ecologists still have to rely on primary data for their research which are short-term and normally only cover the time period of the typical funding cycle (Baker and Millerand 2010).

It is unlikely that scientists will be able to utilize the full potential of the wealth of the information collected by researchers if the present approach towards data sharing does not change. Currently, only about 1% of the data are made available after publication (Reichman et al. 2011) and as a result the rest are eventually lost when the person who holds the dataset cannot be reached for any reason (e.g. because of career change or retirement) (Michener 2006, Whitlock 2011). This not only deprives the scientific community from data that would be costly to gather again but often from valuable information that is impossible to recollect since the existing conditions at the point of collection can not be replicated. Consider for example a dataset that includes the abundances of species in a forest that has been now logged. Such kind of data not only cannot be recollected but are also very useful for answering important conservation questions, that is if they are accessible.

The number of scientists arguing in favour of data sharing and open access to data are increasing and they are citing several reasons why that should be the case. Firstly, without the data the results and the conclusions of a study cannot be tested independently (Whitlock 2011), a central idea of the scientific process. Availability of data would allow scientists to confirm the results, correct errors, and minimize the rare cases of scientific misconduct thus making the field more transparent (Costello 2009, Whitlock 2011). Secondly, many scientists argue that since most of the data are collected using public funds researchers should be required to make them publicly available (Costello 2009). Thirdly and most importantly, ecology is a very data demanding science and the field will undoubtedly benefit from larger amounts of data made available (Borgman et al. 2006). Complex questions, such as on climate change for example, often need large and long-term datasets that can be easier obtained through data sharing.
Studies cite two major impediments in sharing data; the first one is cultural (Fegraus et al. 2005, Reichman et al. 2011, Whitlock 2011). Scientists are often reluctant to share their data even if the data have been analyzed and the results published because they feel that there maybe future potential to that information and they see sharing as a threat to that potential. Another reason explaining this reluctance is that at the moment the incentives for scientists to share their data are neither clear nor strong (Costello 2009, Reichman et al. 2011). Even though it has been shown that citation rates of manuscripts increase when data are shared (Costello 2009), because currently data are not cited in the same way as papers the incentives are not as strong as they could be.

The second impediment to data sharing is technological. Data in ecology are highly heterogeneous making data management and curation very challenging (Fegraus et al. 2005, Reichman et al. 2011, Whitlock 2011). In addition, the options for submitting data are not always obvious and in some cases the repositories or databases available cannot handle the large diversity in data (Madin et al. 2007). Scientists sometimes publish data on their website but the longevity of the information is relatively short and lasts only until that site is maintained. Some individual research groups set up and manage their own local repositories (Baker and Millerand 2010) which are either open to other people or access can be given upon request, but locating those repositories is not always an easy task and access is often restricted and conditional.

Nevertheless, steps are made in the right direction and ecological databases continue to increase both in numbers and volume (Michener 2006). There are several freely accessible repositories such as Dryad and Global Biodiversity Information Facility (GBIF) and recently more journals have started encouraging the submission of primary data, as supplementary information, which they make available on their websites (Parr 2007) and can be accessed along with the corresponding paper. Journals and funding organizations in particular can help change the process and promote data sharing by requiring scientists to make data available at the time of the publication. Already the National Science Foundation, a major funding organization in the United States, requires a data management plan to be included in the funding proposals and journals such as Science and Nature ask researchers to submit their data along with the manuscript accepted (Dittert et al. 2001, Reichman et al. 2011).
Yet, it will take some time before the cultural and technological obstacles are fully overcome and before we realize that collecting data but not sharing them with the rest of the scientific community inhibits the growth of the field and its potential to answer complex and data-demanding questions.
REFERENCES


and bird extinction in tropical forest fragments. Conservation Biology
**13**:1140-1150.

Brotons, L. and S. Herrando. 2001. Reduced bird occurrence in pine forest fragments
associated with road proximity in a Mediterranean agricultural area.
Landscape and Urban Planning **57**:77-89.

Volume 1: Ostriches to Falcons. Academic Press, San Diego, CA, USA.

Burger, J. and M. Gochfeld. 1991. Human Distance and Birds: Tolerance and
Response Distances of Resident and Migrant Species in India. Environmental
Conservation **18**:158-165.

A Two-Stage Approach. Psychological Methods **10**:40-64.

Clough, Y., D. Dwi Putra, R. Pitopang, and T. Tscharntke. 2009. Local and landscape
factors determine functional bird diversity in Indonesian cacao agroforestry.
Biological Conservation **142**:1032-1041.

Preparing Ecological Data Sets to Share and Archive. Bulletin of the
Ecological Society of America **82**:138-141.

Coreau, A. and J.-L. Martin. 2007. Multi-scale study of bird species distribution and
of their response to vegetation change: a Mediterranean example. Landscape
Ecology **22**:747-764.


of the Birds of Europe, the Middle East and North Africa, Oxford University
Press, Oxford, UK.


Understory Bird Community in a Logged Forest in Uganda. Conservation
Biology **14**:265-276.

/itc/cerc/danoff-burg/MBD_Links.html. NY, USA.


ESRI. 2009. ArcMap 9.2. ESRI (Environmental Systems Resource Institute), Redlands, California, USA.


http://CRAN.R-project.org/package=sem.


Disentangling the effects of area, energy and habitat heterogeneity on boreal 
forest bird species richness in protected areas. Global Ecology and 
Biogeography 19:61-71.

Iezekiel, S., C. Makris, and A. Antoniou. 2004. Important Bird Areas of European 
Union Importance in Cyprus. BirdLife Cyprus.

alternative for assessing causal relationships in threatened plant populations. 
Biological Conservation 113:367-377.


Jiguet, F., A. Gadot, R. Julliard, S. E. Newson, and D. Couvet. 2007. Climate 
envelope, life history traits and the resilience of birds facing global change. 

Julliard, R., F. Jiguet, and D. Couvet. 2004. Common birds facing global changes: 

Kambona, O. O. 2005. Environmental Protection, Ecotourism and Rural Livelihoods: 
The Case of Kakamega Forest Region, Kenya. PhD Thesis. University of 
Salzburg, Germany.

management of a Mediterranean mountainous reserve (Pindos National Park, 
Greece) using the bird community as an indicator. Journal for Nature 
Conservation 17:47-59.

Broadbills to Chats. Academic Press, San Diego, CA, USA.

Fragmentation and local disturbance of forests reduce frugivore diversity and 

Disturbance of Tropical Forests on Frugivores and Seed Removal of a Small-


References


APPENDIX I: Results of Multiple Regression Analysis

Table A.1 Results of the multiple regression analyses, beta coefficients, for FS, FG and FV using the five equations described in Chapter 3

<table>
<thead>
<tr>
<th></th>
<th>Specialists</th>
<th>Generalists</th>
<th>Visitors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>N</td>
<td>S</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>.091</td>
<td>.091</td>
<td>.091</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>.637***</td>
<td>.637***</td>
<td>.637***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.669***</td>
<td>-.669***</td>
<td>-.669***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.449*</td>
<td>-.449*</td>
<td>-.449*</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>.252</td>
<td>.304</td>
<td>.017</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.064</td>
<td>-.012</td>
<td>.019</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.042</td>
<td>-.013</td>
<td>.142</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>.467*</td>
<td>.546**</td>
<td>.411</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>.312</td>
<td>.405**</td>
<td>.572**</td>
</tr>
<tr>
<td>Canopy Openness</td>
<td>.327</td>
<td>.247</td>
<td>.046</td>
</tr>
<tr>
<td>Plant Diversity</td>
<td>.037</td>
<td>.283</td>
<td>.467</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>.492*</td>
<td>.733***</td>
<td>.276</td>
</tr>
</tbody>
</table>

Significance levels: *0.05, **0.01, ***0.001; S: Richness, N: Abundance

Table A.2 Results of the multiple regression analyses, beta coefficients, for frugivores and insectivores using the five equations described in Chapter 3.

<table>
<thead>
<tr>
<th></th>
<th>Frugivores</th>
<th>Insectivores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>.091</td>
<td>.091</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>.637***</td>
<td>.637***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.669***</td>
<td>-.669***</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.449*</td>
<td>-.449*</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>.056</td>
<td>.149</td>
</tr>
<tr>
<td>Number of Stumps</td>
<td>-.179</td>
<td>.229</td>
</tr>
<tr>
<td>Canopy Openness</td>
<td>.327</td>
<td>.247</td>
</tr>
<tr>
<td>Plant Diversity</td>
<td>.037</td>
<td>.283</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>.492*</td>
<td>.733***</td>
</tr>
</tbody>
</table>

Significance levels: *0.05, **0.01, ***0.001; S: Richness, N: Abundance
APPENDIX II: List of the experts consulted

1. Dr. Costas Kadis (Associate Professor at Frederick University and Head of the Nature Conservation Unit; Conservation Biologist with research experience on the wildlife of Cyprus)
2. Dr. Claire Papazoglou (Executive Director of BirdLife Cyprus; with research experience on the birds of Cyprus)
3. Martin Hellicar (Campaigns Manager of BirdLife Cyprus; with research experience on the birds of Cyprus)
4. Michael Miltiadou (Research Officer at BirdLife Cyprus; with research experience on the birds of Cyprus)
5. Despo Zavrou (Environment Department, Ministry of Agriculture, Natural Resources and Environment; involved with nature conservation issues in Cyprus)
6. Lefkios Sergides (Environment Department, Ministry of Agriculture, Natural Resources and Environment; involved with nature conservation issues in Cyprus)
7. Haris Nikolaou (Forests Departments, Ministry of Agriculture, Natural Resources and Environment; involved with forest management in Cyprus)
8. Minas Papadopoulos (Forests Departments, Ministry of Agriculture, Natural Resources and Environment; involved with forest management in Cyprus)
9. Nikos Kassinis (Game Fund, Ministry of Interior; involved with bird conservation issues in Cyprus)
10. Dr. Eleftherios Hadjisterkotis (Ministry of Interior; with research experience on the wildlife of Cyprus)
11. Dr. Iris Charalambidou (Research Fellow at University of Nicosia; ornithologist with research experience on the birds in Cyprus)
12. Dr. Alex Kirschel (Lecturer at University of Cyprus; ornithologist with research experience on birds)
## APPENDIX III: Natura 2000 sites (SACs) used in the analysis

<table>
<thead>
<tr>
<th>#</th>
<th>Code</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CY2000001</td>
<td>MAMMARI - DENEIA</td>
</tr>
<tr>
<td>2</td>
<td>CY2000002</td>
<td>ALYKOS POTAMOS - AGIOS SOZOMENOS</td>
</tr>
<tr>
<td>3</td>
<td>CY2000003</td>
<td>PERIOCHI MITSEROU</td>
</tr>
<tr>
<td>4</td>
<td>CY2000004</td>
<td>DASOS MACHAIRA</td>
</tr>
<tr>
<td>5</td>
<td>CY2000005</td>
<td>MADARI - PAPOUTSA</td>
</tr>
<tr>
<td>6</td>
<td>CY2000007</td>
<td>PERIOCHI PLATY</td>
</tr>
<tr>
<td>7</td>
<td>CY2000008</td>
<td>KOILADA KEDRON - KAMPOS</td>
</tr>
<tr>
<td>8</td>
<td>CY2000009</td>
<td>FOUNTOUKODASI PITSILIAS</td>
</tr>
<tr>
<td>9</td>
<td>CY2000010</td>
<td>KOILADA POTAMOU MAROULLENAS</td>
</tr>
<tr>
<td>10</td>
<td>CY2000011</td>
<td>POTAMOS PERISTERONAS</td>
</tr>
<tr>
<td>11</td>
<td>CY2000012</td>
<td>KOILADA KARGOTI</td>
</tr>
<tr>
<td>12</td>
<td>CY3000005</td>
<td>KAVO GKREKO</td>
</tr>
<tr>
<td>13</td>
<td>CY3000008</td>
<td>LIMNI PARALIMNIOU</td>
</tr>
<tr>
<td>14</td>
<td>CY4000001</td>
<td>PERIOCHI POLIS - GIALIA</td>
</tr>
<tr>
<td>15</td>
<td>CY4000002</td>
<td>CHA-POTAMI</td>
</tr>
<tr>
<td>16</td>
<td>CY4000003</td>
<td>KOILADA DIARIZOU</td>
</tr>
<tr>
<td>17</td>
<td>CY4000004</td>
<td>VOUNI PANAGIAS</td>
</tr>
<tr>
<td>18</td>
<td>CY4000005</td>
<td>EPISKOPI MOROU NEROU</td>
</tr>
<tr>
<td>19</td>
<td>CY4000007</td>
<td>XEROS POTAMOS</td>
</tr>
<tr>
<td>20</td>
<td>CY4000008</td>
<td>MAVROKOLYMPOS</td>
</tr>
<tr>
<td>21</td>
<td>CY4000009</td>
<td>PERIOCHI SKOULLI</td>
</tr>
<tr>
<td>22</td>
<td>CY4000010</td>
<td>CHERSONISOS AKAMA</td>
</tr>
<tr>
<td>23</td>
<td>CY4000011</td>
<td>PERIOCHI AGIATIS</td>
</tr>
<tr>
<td>24</td>
<td>CY4000012</td>
<td>PERIOCHI STAYROS TIS PSOKAS - KARKAVAS</td>
</tr>
<tr>
<td>25</td>
<td>CY4000013</td>
<td>FAROS KATO PAFOU</td>
</tr>
<tr>
<td>26</td>
<td>CY4000014</td>
<td>PERIOCHI DRYMOU</td>
</tr>
<tr>
<td>27</td>
<td>CY4000015</td>
<td>PERIOCHI KRITOU MAROTTOU</td>
</tr>
<tr>
<td>28</td>
<td>CY5000001</td>
<td>DASOS LEMESOU</td>
</tr>
<tr>
<td>29</td>
<td>CY5000004</td>
<td>ETHNIKO DASIKO PARKO TROODOUS</td>
</tr>
<tr>
<td>30</td>
<td>CY5000005</td>
<td>AKROTIRIO ASPRO - PETRA ROMIOU</td>
</tr>
<tr>
<td>31</td>
<td>CY5000006</td>
<td>KOILADA LIMNATI</td>
</tr>
<tr>
<td>32</td>
<td>CY5000007</td>
<td>PERIOCHI ASGATAS</td>
</tr>
<tr>
<td>33</td>
<td>CY6000002</td>
<td>ALYKES LARNAKAS</td>
</tr>
<tr>
<td>34</td>
<td>CY6000003</td>
<td>PERIOCHI LYMPION - AGIAS ANNAS</td>
</tr>
<tr>
<td>35</td>
<td>CY6000004</td>
<td>DASOS STAVROVOUNIOU</td>
</tr>
<tr>
<td>36</td>
<td>CY6000005</td>
<td>PERIOCHI LEFKARON</td>
</tr>
<tr>
<td>37</td>
<td>CY6000006</td>
<td>ETHNIKO DASIKO PARKO RIZOELIAS</td>
</tr>
<tr>
<td>38</td>
<td>CY6000011</td>
<td>LIMNI OROKLINIS</td>
</tr>
</tbody>
</table>
APPENDIX IV: GLMM results using Dataset 1 (Chapter 6)

Table A.3 Variance explained by the random effects in Presence/Absence Model, using Dataset

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species (n=108)</td>
<td>3.9070902</td>
<td>1.976636</td>
</tr>
<tr>
<td>Sites (n=22)</td>
<td>0.0042757</td>
<td>0.065389</td>
</tr>
</tbody>
</table>

Residual Variance: 1.0, Deviance: 2108
AIC: 2152, BIC: 2279

Table A.4 Parameter estimates for Presence/Absence Model, using Dataset 1

| Estimate       | Std. Error | z value | Pr(>|z|) |
|----------------|------------|---------|---------|
| Intercept      | 0.78       | 2.70    | 0.290   | 0.772090 |
| Plant diversity| 0.05       | 0.02    | 2.625   | 0.004721 ** |
| Foliage Height Diversities | -3.91     | 1.65    | -2.378  | 0.017389 * |
| Number of Stumps | 0.69     | 0.38    | 1.801   | 0.071655 • |
| Distance to Forest Edge | 0.95     | 0.32    | 2.934   | 0.003349 ** |
| Clutch Size     | -2.77      | 2.36    | -1.174  | 0.240227 |
| Forest Guild (FS) | 3.21     | 1.67    | 1.925   | 0.054264 • |
| Forest Guild (FV) | 1.58     | 1.86    | 0.850   | 0.395352 |
| Forest Guild (NF) | 3.47     | 2.91    | 1.193   | 0.232804 |
| Plant Diversity:Forest Guild FS | 0.03    | 0.02    | 1.067   | 0.286130 |
| Plant Diversity:Forest Guild FV | -0.10   | 0.03    | -3.498  | 0.000469 *** |
| Plant Diversity:Forest Guild NF | -0.10   | 0.04    | -2.161  | 0.030709 * |
| FHD:Clutch Size  | 3.47       | 1.70    | 2.036   | 0.041774 * |
| Stumps:Forest Guild FS | -1.31   | 0.50    | -2.630  | 0.008545 ** |
| Stumps:Forest Guild FV | -0.03   | 0.55    | -0.061  | 0.951643 |
| Stumps:Forest Guild NF | -1.45   | 0.88    | -1.654  | 0.098114 • |
| Distance FE:Clutch Size | -0.88  | 0.31    | -2.901  | 0.003715 ** |
| Distance FE:Forest Guild FS | 0.02    | 0.24    | 0.088   | 0.929490 |
| Distance FE:Forest Guild FV | -0.72  | 0.27    | -2.722  | 0.006492 ** |
| Distance FE:Forest Guild NF | -0.50   | 0.44    | -1.147  | 0.251561 |

Significance levels: *** 0.001, ** 0.01, * 0.05, • 0.1