

**THE DEVELOPMENT OF SPECIES DISTRIBUTION
MODELS AND THEIR APPLICATION TO BIRDS IN
RIVER HABITATS**

Thesis submitted for the degree of Doctor of Philosophy

IAN P. VAUGHAN

**Cardiff School of Biosciences
Cardiff University, July 2004**

UMI Number: U584653

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U584653

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

1. Distribution models are used as management tools to predict species' distributions and quantify their habitat preferences. Numerous methodological issues require further development, which are explored using the distribution of birds along rivers as a model system in which there is a need to i) develop the quantification and analysis of variation in river habitat features and ii) better quantify species' habitat preferences for conservation and prediction. These themes were linked using a combination of the British Trust for Ornithology's Waterways Breeding Bird Survey (WBBS) and the Environment Agency's River Habitat Survey (RHS), along with similar data from upland Wales and the Himalayan mountains.
2. Training data are the starting point for distribution models and their properties can have profound effects. These issues were investigated via a literature review, which identified key factors including the type of predictors chosen, the approach to environmental sampling and spatial autocorrelation in distribution patterns. Recommendations are made that should optimise model quality, whilst making the most efficient use of available resources.
3. Testing the performance of distribution models is vital. Using a review of the available literature, highlighting weaknesses in current practice, and a case study using a Himalayan river bird, recommendations are made for improved practice. Bootstrapping and independent data should be used to assess overfitting and transportability, respectively. Accuracy statistics should facilitate inter-model comparisons, examining both discrimination and calibration. Nominal presence/absence predictions are problematic: information-theoretic methods may be the most useful approach.

4. Complex habitat data, such as RHS, may create a range of problems during statistical analyses unless the sample size is large. Data reduction, using methods such as principal components analysis (PCA), is an effective solution, but the resulting axes may be difficult to interpret. Using models built with Welsh river bird-RHS data, I compared the interpretability and predictive efficacy of PCA used in its conventional form against PCA preceded by the clustering of RHS variables that referred to the same ecological factors. The two approaches produced similar predictive performance but habitat indices produced by the latter were easier to interpret. A variant of PCA devised for qualitative data was also examined, and benefited RHS analyses involving ordinal variables.
5. Predictive models for 28 river birds, built with the WBBS, represent the first quantitative study linking detailed river habitat data with river bird distributions across the UK. Accuracy varied widely, with better performance for species associated with the river channel, rather than floodplain habitats, reflecting the relative coverage of these river features in RHS. By using variable clustering, the likelihood of species occurrences could be easily related to the observed habitat.
6. By utilising important methodological developments, this project provides important evidence that RHS forms an effective basis for relating many river birds to their habitats, and that when used in conjunction with the WBBS, could bring valuable benefits to river bird conservation. More generally, the work illustrates how RHS can describe variations in river structure and anthropogenic modification in a manner that is relevant to riverine organisms, along with transferable methods for describing and modelling the resulting relationships.

Acknowledgements

Numerous people have contributed in some way to the production of this thesis and I should like to express my gratitude to them at the outset. First, my supervisor Prof. Steve Ormerod for guidance and encouragement throughout the project, particularly helping me to see my work in a broader context and whenever I put pen to paper.

The project was in collaboration with the Environment Agency (EA) and British Trust for Ornithology (BTO), both of whom provided important resources. The EA provided funding for the project, my accreditation in River Habitat Surveying and made the national River Habitat Survey database freely available. The RHS team in Warrington provided useful feedback about my ideas, especially from an applied viewpoint, and I would particularly like to thank Marc Naura, and Drs Helena Parsons and Kevin Hall in this respect. BTO volunteers collected the Waterways Breeding Bird Survey (WBBS) data. The BTO also collated the WBBS data set, overcoming major problems in so doing, and provided useful ideas for the analysis: I would particularly like to thank Drs David Noble, Dave Leech and John Marchant. Several other data sets were used in the thesis and acknowledgements are made at the end of the appropriate chapters.

I would also like to acknowledge the contribution of a range of other people, not directly connected to the project. First, my various friends and colleagues at Cardiff University over the past three plus years, for making the PhD process such a rewarding one. Secondly, I wish to thank Prof. Tony Walsby of Bristol University for vital preparations for postgraduate study during my time there and invaluable guidance in Excel®. I would also like to 'thank' Dr Stefan Bodnar, for suggesting the PhD idea in the first place...

Finally, I would like to thank my family for their continual support – both moral and logistical – throughout the last 45 months.

Ian Vaughan

July 2004

Contents

NB: each chapter within the thesis is presented as an individual paper. Consequently, each chapter is presented with its own self-contained reference list

List of tables xi

List of figures xiii

Chapter 1: Introduction	1
1.1 Species distribution modelling	2
1.1.1 A dichotomy in modelling methodology.....	6
1.1.2 The importance of further research in empirical distribution modelling.....	8
1.2 The importance of modelling river bird distributions	9
1.3 Background to the current project	15
1.3.1 The Waterways Breeding Bird Survey.....	16
1.3.1.1 Development.....	16
1.3.1.2 WBBS method	16
1.3.2 The River Habitat Survey	17
1.3.2.1 RHS method	19
1.4 The RHS-WBBS data set	24
1.4.1 Supplementary data sets used in the thesis	25
1.5 Project aims	25
1.6 References	28

Chapter 2: Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data.....	50
2.1 Summary.....	51
2.2 Introduction	53
2.3 Clarification of model aims	56
2.4 Accurately capturing organism-environment relationships.....	61
2.5 Sampling scale for analysing organism-environment relationships ...	65
2.6 Improving analytical success.....	67
2.7 Tackling spatial autocorrelation	70
2.8 Collecting suitable data with which to evaluate the model	73
2.9 Conclusions	74
2.10 Acknowledgements	77
2.11 References	78

Chapter 3: Testing the predictive performance of species distribution models: unresolved issues and omissions from current practice.....	90
3.1 Summary.....	91
3.2 Introduction	93
3.3 Testing generalisability.....	95
3.3.1 The relationship between resample testing and independent test data.....	96
3.3.2 Collection of test data	99
3.3.3 Describing predictive performance in independent data	103
3.4 Selection of accuracy statistics 1: probabilistic/continuous-scale predictions	104
3.4.1 Measures of discriminatory ability for probabilistic models.....	104
3.4.2 Measures of calibration for probabilistic models	107
3.5 Selection of accuracy statistics 2: nominal presence/absence predictions	111
3.6 Case study.....	117
3.6.1 Resubstitution and resampling performance	117

3.6.2	Performance in the test set.....	118
3.6.3	Conclusions from testing.....	121
3.7	Conclusions	122
3.8	Acknowledgements	124
3.9	References	125

Chapter 4: Simplifying complex data in ecology: improving the interpretability and versatility of principal components analysis..... 138

4.1	Summary.....	139
4.2	Introduction	141
4.3	Methods	145
4.3.1	River Habitat Survey	145
4.3.2	Principal components analysis.....	147
4.3.3	Variable clustering.....	147
4.3.3.1	Cluster formation.....	147
4.3.3.2	Scoring clusters.....	148
4.3.4	River bird models	149
4.3.4.1	Waterways Breeding Bird Survey	151
4.4	Results	152
4.4.1	Overall PCA	152
4.4.2	Variable clusters	152
4.4.2.1	Land use.....	156
4.4.2.2	Anthropogenic modification.....	156
4.4.2.3	Overall channel characteristics.....	162
4.4.2.4	Flow types and substrata.....	163
4.4.3	Scoring clusters with PCA.....	163
4.4.4	River bird models	165
4.4.4.1	Waterways Breeding Bird Survey models.....	170
4.5	Discussion.....	170
4.5.1	Scoring variable clusters.....	177
4.6	Conclusions	179
4.7	Acknowledgements	180
4.8	References	181

Chapter 5: River birds and their habitats: linking the Waterways Breeding Bird Survey and the UK River Habitat Survey.....	190
5.1 Summary.....	191
5.2 Introduction	193
5.3 Methods	194
5.3.1 Waterways Breeding Bird Survey	194
5.3.2 River Habitat Survey	196
5.3.3 Indices of river habitat	197
5.3.4 River bird-habitat models	198
5.4 Results	204
5.5 Discussion.....	209
5.5.1 Pragmatic properties of RHS.....	210
5.5.2 Linking river birds to RHS	212
5.5.3 Applications of RHS models	217
5.6 Conclusions	219
5.7 Acknowledgements	219
5.8 References	220
Chapter 6: Discussion.....	231
6.1 Introduction	232
6.2 Species distribution modelling	232
6.2.1 Improving the value and uptake of distribution models.....	234
6.3 An appraisal of RHS for species distribution modelling.....	236
6.4 Waterways Breeding Bird Survey and RHS.....	240
6.5 References	242
Appendix 1: River Habitat Survey.....	246
1.1 Reproduction of the River Habitat Survey forms for the 1997 revision of the survey	246
1.2 Complete listing of variables in the 1997 revision of the River Habitat Survey.....	253

1.3 Preparation of River Habitat Survey data for the analyses in Chapters 4
& 5 260

**Appendix 2: The 161 sites selected for the Waterways Breeding Bird Survey
analysis..... 265**

**Appendix 3: Published version of Chapter 2 (*Conservation Biology*, 17, 1601–
1611)..... 271**

List of tables

Chapter 1

Table 1	A range of applications for species distribution models	4
Table 2	Some research directions with empirical distribution models	10
Table 3	Summary of the River Habitat Survey (1997 revision)	21

Chapter 2

Table 1	Major problems experienced in the use of distribution models	54
Table 2	The choice between indirect and direct predictor variables.....	57
Table 3	A summary of the major recommendations for field study design	75

Chapter 3

Table 1	A simple bootstrapping method to estimate overfitting.....	98
Table 2	Possible differences between training data and potential applications for distribution models that should be addressed in the selection of test data.....	100

Chapter 4

Table 1	RHS variable clusters and their constituent variables.....	157
Table 2	Predictive abilities of the distribution models for four river bird species using the three different data reduction methods	166
Table 3	Structures and interpretation of the variable cluster models.....	168
Table 4	A comparison of conventional PCA and variable clustering.....	172
Table 5	Potential benefits and disadvantages of optimally-scaled PCA.....	178

Chapter 5

Table 1	Species considered in the analysis of the WBBS.....	195
Table 2	Description of the river habitat indices used in the river bird models	199
Table 3	The predictive performance of WBBS river bird models.....	205

Chapter 6

Table 1	Factors that may influence the perceived value of species distribution models in practice	233
----------------	--	------------

List of figures

Chapter 2

- Figure 1 The potential impact of incomplete sampling of an environmental gradient upon subsequent interpretation 62

Chapter 3

- Figure 1 Sample calibration plot, comparing predicted and observed prevalence at decile probability intervals 109
- Figure 2 Non-parametric calibration curves for the plumbeous redstart model 119
- Figure 3 The accuracy of predictions for individual sites in the test set 120

Chapter 4

- Figure 1 The scree plot from principal components analysis of RHS data 153
- Figure 2 The distribution of Pearson correlation coefficients for the variables that contributed significantly to PCs 1–3 154
- Figure 3 Average-linkage clustering tree for the full RHS data set 155
- Figure 4 Typical form of the optimised transformation for ordinal sweep-up variables 164

Chapter 1

General introduction

1.1 Species distribution modelling

Distribution models are ubiquitous in ecology and conservation biology. They comprise a diverse set of tools that quantify the relationships between the environment and some aspect of species' actual or potential distribution patterns: absolute or relative abundance, presence or absence, the likelihood of occurrence or a measure of the suitability of a site for the species. Such models may also be generalised to other ecological attributes, such as whole communities (e.g. Zimmermann & Kienast 1999), diversity (e.g. Gioia & Pigott 2000), vegetation physiognomy (e.g. Mackey 1994), breeding performance (e.g. Paradis *et al.* 2000), genetic diversity (e.g. Scribner *et al.* 2001) or carnivore predation of livestock (e.g. Treves *et al.* 2004). Distribution models have been applied to many taxa, especially birds (e.g. Chamberlain *et al.* 1999; Franco, Brito & Almeida 2000; Suárez-Seoane, Osborne & Alonso 2002), mammals (e.g. Lindenmayer *et al.* 1991; Mladenoff *et al.* 1995; Bonesi, Rushton & Macdonald 2002) and vascular plants (e.g. Franklin 1998; Wisser, Peet & White 1998; Thuiller, Araújo & Lavorel 2003), but also to bryophytes (e.g. Suren & Ormerod 1998), algae (e.g. Ter Braak & Van Dam 1989), cyanobacteria (e.g. Maier *et al.* 1998), fungi (e.g. Backhouse & Burgess 1995), invertebrates (e.g. Wright 1995; Naura & Robinson 1998), fish (e.g. D'Angelo *et al.* 1995; Lek *et al.* 1996) and herpetofauna (e.g. Sarre *et al.* 1995; Knapp *et al.* 2003). The methods used for distribution modelling are similarly diverse (see Guisan & Zimmermann 2000 for a recent review).

Species distribution models first appeared in the early 1960s, but it is since the 1970s that the field has burgeoned, driven by the rapid growth in computing power and availability, and the application of multivariate methods (see Stauffer 2002 for a

history of distribution modelling). This growth has continued to the present, where distribution modelling studies are abundant in the ecological and biogeographical literature. Between 1999 and 2003 inclusive, for example, distribution modelling was incorporated into 7.0% of the research papers in *Biological Conservation*, 5.1% in *Conservation Biology*, 7.2% in *Ecological Applications*, 10.4% in the *Journal of Applied Ecology* and 8.1% in the *Journal of Biogeography*.

The current prevalence of distribution modelling can be attributed to the multitude of potential applications for such models in both theoretical and applied ecology (Table 1). In all cases, however, there is an essential dichotomy between explanatory and predictive applications (Table 1; Ripley 1994; Mac Nally 2000; O'Connor 2002). Explanatory modelling aims to describe the environmental requirements of species, quantifying their relationships with particular aspects of their habitats. Such information readily feeds into applied ecology, guiding species management strategies (Table 1). In models of this type, transparency and clear ecological rationale are paramount: prediction accuracy may be sacrificed for examining the relation of certain variables to species distribution, perhaps testing specific hypotheses.

In many recent modelling studies, especially in applied ecology, the primary focus has been upon prediction (Rushton, Ormerod & Kirby 2004). This reflects the range of pressing conservation issues to which predictive distribution models can be applied (Table 1). Predictive modelling differs from explanatory in so far as the structure of a model may be considered to be unimportant: the primary concern is predictive ability. Indeed, Copas (1983) argues that the inclusion of variables with no known relationship to a species and even systematic biases in a model may not be important

Table 1. A range of applications for species distribution models.

Explanatory applications

- Exploratory modelling: distribution modelling may be amongst the tools used in data-driven hypothesis generation
- Quantifying the effects that different factors have upon species' distributions:
 - identify habitat requirements (e.g. Guerry & Hunter 2002)
 - inform management advice to increase populations of desirable species (e.g. Chamberlain *et al.* 1999) or reduce populations of pest or alien species (e.g. Gregory, Carter & Baillie 1997)
 - diagnose causes of species' declines (e.g. Bustamante 1997)
- Testing ecological or biogeographical hypotheses (e.g. Manel, Buckton & Ormerod 2000)

Predictive applications

- Estimate basic distribution data and improve the quality of biodiversity databases or atlases (e.g. Osborne & Tigar 1992; Tushabe, Reynolds & Pomeroy 2001)
 - Estimate the extent of biodiversity loss (e.g. Lek-Ang *et al.* 1999)
 - Identify sites expected to hold important species – guide reserve selection (e.g. Loiselle *et al.* 2003), target conservation efforts (e.g. Ferrier *et al.* 2002) or identify sites to survey for rare species (Wiser, Peet & White 1998).
 - Select possible sites for re-introductions of species (e.g. Donázar *et al.* 1993).
 - Predict re-colonisation patterns (e.g. Corsi, Dupré & Boitani 1999)
-

-
- Identify unexpected gaps in species' distributions, from which causes can be diagnosed, either for the conservation of a species (e.g. Lawton & Woodroffe 1991) or using it as a bioindicator (e.g. Wright 1995)
 - Support estimates of population persistence/extinction where detailed demographic studies cannot be carried out (e.g. Araújo & Williams 2000)
 - Predict pest outbreaks (e.g. Lindblad 2001) or identify sites susceptible to invasion by alien species (e.g. Collingham *et al.* 2000)
 - Predict species' responses to environmental change, for example:
 - predict the possible impacts of invading/alien species upon existing flora/fauna (e.g. Rushton *et al.* 2000)
 - predict the effects of large scale environmental changes, such as climate change (e.g. Erasmus *et al.* 2002), changes to the Common Agricultural Policy (e.g. Bustamante 1997) or habitat fragmentation (e.g. Bolger 1997)
 - predict the effects of local habitat change, such as species' responses to alternative management strategies (e.g. McCune *et al.* 2003)
-

provided that the predictive performance is proven to be sufficient. To successfully carry out both predictive and explanatory modelling applications for a species, distinct analytical approaches may be required (Mac Nally 2000).

The current popularity of distribution modelling may also, to some degree, reflect the relative ease of carrying out such analyses, rather than having clear *a priori* goals.

This phenomenon has been identified and fiercely criticised in medical research (Altman 1994). Whilst such harsh criticism does not seem to be warranted in ecology, O'Connor (2002) raises some similar concerns.

1.1.1 A DICHOTOMY IN MODELLING METHODOLOGY

The diversity of applications for distribution models is reflected in the range of modelling methodologies that have been used (see Franklin 1995; Guisan & Zimmermann 2000; Scott *et al.* 2002 for overviews of this diversity). Broadly, two types of model are recognised, representing the ends of a continuum, with process-based mechanistic models at one extreme and static, empirical models at the other (Guisan & Zimmermann 2000).

Mechanistic models are potentially more powerful, aiming to simulate the actual biological processes that affect distribution (e.g. Sykes *et al.* 2001). By directly modelling species' responses to specific elements of their environment, mechanistic models should be well suited to predicting their reaction to environmental changes, such as those expected for climate change (Sutherland 1996). However, these responses may not be stable over time, due to factors such as rapid evolutionary change (Thomas *et al.* 2001). Modelling of a species' physiological or behavioural

responses to environmental factors, without equal detail to describe its interactions with other species, means that dynamic models often model the fundamental niche: this is unlikely to be realised in field conditions (Pearson & Dawson 2003). Probably the biggest weakness of mechanistic models, however, is the extensive *a priori* information required to specify and parameterise them. This prevents their use in many instances.

Static models, by contrast, may be formulated when there is virtually no prior knowledge of the species being modelled, although any ecological knowledge that can be applied is liable to benefit modelling (Austin 2002; Burnham & Anderson 2002; Vaughan & Ormerod 2003). They are based on the concept of the realised niche occupied by a species, following Hutchinson's (1957) definition, approximated from the correlations between the observed species' distribution and a series of potential predictor variables: the n -dimensions in which the niche is defined (Green 1971; Austin, Nicholls & Margules 1990; Guisan, Edwards & Hastie 2002). As a minimum, all that is therefore required to develop an empirical model is a simple statistical technique, such as multiple linear regression, and a set of matched species and environmental data with which to calibrate it.

Unfortunately, the simple structures of empirical distribution models tend to be relatively inflexible. Their over-riding assumption is that the species-environment relationships, as modelled, will remain constant over space and time: an assumption regarded to be highly unrealistic (Sutherland 1996; Guisan & Theurillat 2000; O'Connor 2002). Nevertheless, the ability to make predictions in 'knowledge poor' situations means that empirical models are invaluable to conservation biology,

because of the lack of information concerning many species. The models produced may also be very accurate (e.g. Osborne, Alonso & Bryant 2001; Luoto, Kuussaari & Toivonen 2002; Miller & Franklin 2002). As a consequence, empirical models are used for virtually all distribution modelling and have directly informed conservation planning in different parts of the world (e.g. Ferrier *et al.* 2002). They are the focus of the current thesis for two reasons. First, because of the paucity of data to parameterise mechanistic models for river birds, and secondly to ensure that findings concerning modelling methodology will have the broadest generic relevance. Subsequently, distribution modelling refers exclusively to static, empirical models.

1.1.2 THE IMPORTANCE OF FURTHER RESEARCH IN EMPIRICAL DISTRIBUTION MODELLING

Despite the extensive development of distribution modelling in the last 20 years and the number of modelling studies carried out, much work is still required to improve the methods used. Such development is made a priority by the number of applications that could benefit from successful modelling (Table 1), coupled with the potentially serious 'costs' of making inaccurate predictions. From a conservation standpoint, the stakes involved in modelling can be very high if models are relied upon to inform planning and management decisions. This assumes particular importance when non-specialists are involved in the use of models and interpretation of the outputs, as they may apply them uncritically. Prediction errors, and the concomitant misdirection of conservation resources – financial and/or logistical – may be very damaging. At best, the prioritisation of conservation targets may be compromised. At worst, sites containing rare species may be overlooked, and therefore not be conserved, or

resources may be wasted on sites of no importance to a species. There is real incentive, then, to continue to improve distribution modelling methodology.

In a field as extensive and diverse as distribution modelling, numerous unresolved issues and areas for improvement can be identified (Table 2). Some of these points are relevant to all modelling studies, such as the measures used to quantify prediction accuracy. A common theme is a need for greater knowledge of the system being modelled, often with the aim of adding more mechanistic elements (e.g. using predictors with direct biological relationships with the organism or more realistic networks of predictive factors; Table 2). Whilst highly desirable, such developments have more limited applicability to distribution modelling as a whole, being as they require much greater knowledge of the systems being modelled from the outset.

With so many issues requiring further work, the value of any distribution modelling study will be increased if it can address some of them. The intention is that the current project will do just this, addressing a few of the methodological issues highlighted.

1.2 The importance of modelling river bird distributions

Globally, rivers form an important habitat for birds, with around 60 specialist riverine species recognised and up to 23% of all bird species utilising freshwaters, including rivers, for part or all of their life histories (Tyler & Ormerod 1993; Buckton 1998; Buckton & Ormerod 2002). Less specialised species are often associated either with riverside vegetation (e.g. *Acrocephalus* species) or undrained wetlands found in the riparian environment, such as several species of wading bird (Marchant & Hyde 1980b). For the purposes of this project, river birds are defined in this looser

Table 2. Some research directions with empirical distribution models and selected sources in which the issues have been raised

Theoretical framework

- use of constraints/limiting factors to distribution as predictors rather than correlates of distribution (Huston 2002; O'Connor 2002)
- incorporation of predictors from multiple scales, interacting species and other increases in biological realism (Vaughan & Ormerod 2003)

Data collection

- size of data sets required to train and test models (Karl *et al.* 2002; Stockwell & Peterson 2002)
- optimal sampling methodology (Margules & Austin 1994; Hirzel & Guisan 2002)
- how to target relevant spatial and temporal scales
- are certain classes of predictors consistently more effective/useful?

Modelling methods

- do alternative methods offer consistent advantages? e.g. machine learning algorithms versus traditional statistical approaches (Fielding 1999; Manel *et al.* 1999a; Olden & Jackson 2002)
 - how to increase the generality of models and reduce overfitting of training data
 - how to incorporate multi-scale information (extent & grain; Buckland & Elston 1993; Mackey & Lindenmayer 2001)
 - how to generalise models to non-equilibrium situations e.g. modelling spread of species over time (Buckland & Elston 1993; Augustin, Muggleston & Buckland 1998)
 - integration of expert knowledge into models (Pearce *et al.* 2001)
 - value and development of specialist spatial modelling methods e.g. autoregressive models (Lichstein *et al.* 2002), spatial interpolation (Miller & Franklin 2002)
 - how to cope with non-stationarity in the modelled species-environment relationships (Osborne & Suárez-Seoane 2002)
-

Model evaluation

- best methods to test accuracy (Fielding & Bell 1997; Boyce *et al.* 2002)
- most efficient testing methods where data are sparse
- when can models be considered to be validated for a certain application? (Altman & Royston 2000)
- diagnosing the causes of prediction errors (Miller, Hui & Tierney 1991; Pearce & Ferrier 2000)

Application of distribution models

- how to increase the connectivity between modellers and conservation planners/practitioners (Wiens 2002)
- ways to increase the conservation value of models and their uptake by practitioners
- can agreed standards be devised for how models are developed, employed and validated?
- are models demonstrably better than expert opinion?

Ecological concerns

- what aspects of distribution modelling are consistently the most important limiting factors? e.g. properties of training data versus choice of modelling algorithm?
 - to what degree can modelling success be predicted *a priori* based upon the known ecology of the species involved? (Boone & Krohn 1999)
 - how much can be learnt from published distribution modelling studies? Until recently performance was usually assessed using overall prediction success, giving no idea of improvement over chance, and typically only assessed using the training data (Manel *et al.* 2001; McPherson *et al.* in press). Neither provide reliable information about the success of modelling.
 - the potential impacts of evolutionary change on model generality (Thomas *et al.* 2001; Pearson & Dawson 2003)
-

context: as birds that utilise the river and/or neighbouring riparian corridor during the breeding season (the period of data collection). This is consistent with other research into UK river birds (e.g. Carter 1989; Marchant & Hyde 1980a, 1980b; Rushton, Hill & Carter 1994; Buckton & Ormerod 1997).

River birds are an important focus for distribution modelling for three reasons. First, they are a valuable model system for studying a range of ecological issues. Secondly, the habitat requirements of many river birds remain poorly known. Finally, river bird models could be valuable in detecting, and predicting the effects of, anthropogenic impacts upon rivers. Each of these three points will be discussed briefly.

Collectively, birds are often valuable for addressing fundamental ecological questions. Their current and historical popularity amongst natural historians has allowed a wealth of background information to be collected about them, removing a fundamental barrier to studying generic issues, by which the majority of (less well-known) taxa are currently excluded (Lawton 1996). Bird species that regularly breed along rivers display a wide variety of life history strategies in terms of such factors as diet, movement patterns and habitat associations (Snow & Perrins 1998; Wernham *et al.* 2003). The study of river birds also has a series of practical benefits, especially with regard to the ease and reliability with which they can be monitored. Confirming the presence of dippers *Cinclus cinclus* on a river during the breeding season, for example, is very easy and reliable: a study in the Pyrénées indicated that three survey visits ensured that all breeding pairs were detected and that the timing within the breeding season was unimportant (D'Amico & Hemery 2003). Many river birds are similarly easy to census, to trap and to monitor breeding in (e.g. Bibby *et al.* 2000;

D'Amico 2002). In a modelling context, such survey reliability is vital, considering the severity of biases that may be introduced to models where surveys falsely record a species as absent from a site (Tyre *et al.* 2003).

Whilst birds are the most common focus for distribution modelling, accounting for 33% of studies identified by Manel, Williams & Ormerod (2001), river birds are poorly represented. Modelling studies focusing upon river birds or floodplain habitats have been carried out in several parts of the world, including Britain (e.g. Rushton *et al.* 1994; Gregory *et al.* 1997), North America (e.g. Strong & Bock 1990; Groom & Grubb 2002; Hatten & Paradzick 2003), Australia (e.g. Woinarski *et al.* 2000) and the Himalaya of India and Nepal (Manel *et al.* 1999a, 1999b, 2000). Nevertheless, this represents a small proportion of the total set of avian models. In the UK, this lack of quantitative study of habitat requirements contrasts with relatively detailed qualitative knowledge (e.g. RSPB, NRA & RSNC 1994; Snow & Perrins 1998). Where quantitative study has been carried out, it has often been of a univariate nature and restricted to relatively small numbers of catchments in distinct geographical areas, such as the Peak District (e.g. Yalden 1986) or mid-Wales (e.g. Round & Moss 1984). Only one quantitative, multivariate study of river bird-habitat relationships has been carried out at the national level (Rushton *et al.* 1994).

The UK is unusual in having relatively comprehensive networks for surveying the distributions of certain taxa, such as butterflies (Asher *et al.* 2001) and birds (e.g. Gibbons *et al.* 1993). This may negate some of the uses frequently cited for predictive models (Table 2; Fielding & Haworth 1995). However, even if national distributions are relatively well known, predictive models could still represent valuable support

tools for local conservation management. In the case of British river birds, two applications in particular could be valuable: monitoring water quality and predicting species' responses to proposed river management. Both are likely to become increasingly important over the next 10–15 years as the EU Water Framework Directive (2000/60/EC) requires the surveillance of river conditions and the development of 'Programmes of Measures' to help rivers across Europe to attain or maintain 'good ecological status' by 2015 (Logan & Furse 2002).

River birds may act as effective indicators of water quality (Ormerod & Tyler 1993). Dippers in particular have great potential in this respect, being exclusively riverine and largely sedentary, and a series of studies has shown their potential as indicators of acidification, organic pollution and toxins such as polychlorinated biphenyls (PCBs) and organochlorines (e.g. Ormerod *et al.* 1986; Ormerod & Tyler 1990; Ormerod, Tyler & Jüttner 2000; Sorace *et al.* 2002; O'Halloran *et al.* 2003). To use simple presence/absence of a species to indicate water quality problems requires that its occurrence on a site – in the absence of pollution – can be predicted with a high degree of confidence. Predictive models trained at unpolluted reference sites may perform this role (e.g. Wright 1995). Clearly, the basis for such a system would have to be an accurate river bird distribution model, applicable to all sites where the indicator system might be used.

The physical environment of rivers is often heavily modified, particularly for flood defence works, and such activities may have negative impacts upon river bird communities (e.g. Taylor 1984; Raven 1986). To minimise the impact of such works in future, and to help devise the Programmes of Measures for the Water Framework

Directive, the ability to assess how the flora and fauna of the river habitat would be altered by different management options would be extremely valuable. Ultimately, successful prediction of this type may require detailed mechanistic models rather than empirical ones. Nevertheless, empirical models may be able to inform decision making – even if they lack very high predictive power – or at least suggest the components of the river habitat that need to be incorporated into a mechanistic model.

The ability to predict river bird distributions and describe habitat preferences in a meaningful manner is reliant upon an effective approach to capturing and characterising the variation that exists in the structure of river habitats. A critical component of such a method is to integrate anthropogenic modifications alongside natural features of the physical environment. This reflects the heavily modified state of some rivers and allows the effects of past modifications to be investigated, as well as predicting the impacts of future modification. Producing a comprehensive habitat survey of this type represents a major challenge (Raven *et al.* 1997).

1.3 Background to the current project

Data to construct river bird models were obtained from two matched data sets, collected across the UK by teams of surveyors. Bird data were derived from the pilot years of the British Trust for Ornithology's (BTO) Waterways Breeding Bird Survey (WBBS). Habitat data for the sites were in the format of the Environment Agency's (EA) River Habitat Survey (RHS).

1.3.1 THE WATERWAYS BREEDING BIRD SURVEY

1.3.1.1 *Development*

The WBBS is one of two surveys currently used by the BTO to monitor the population trends of birds along streams, rivers and canals in the UK – the other is the longer running Waterways Bird Survey (WBS; Marchant & Coombes 2003). They are essential to provide accurate monitoring of the population trends of several river bird species that are poorly monitored by the more general Breeding Bird Survey (Marchant *et al.* 2002). The WBS was introduced in 1974, but has several disadvantages as a national survey. Sampling sites are not randomly located, being selected by observers; the survey is inefficient, requiring nine visits to a site each year and extensive processing of the results to calculate individual territories; the sample size is relatively small (up to around 100 sites per year) due to the effort required at each site; and it is not integrated with other riverine monitoring methods (Langston, Marchant & Gregory 1997). In an attempt to address these limitations, WBBS was initiated as a pilot study in 1998 (Marchant *et al.* 2002). It aims to use a smaller amount of effort per site (two visits), so that a larger number of randomly selected sites can be visited in an attempt to increase the precision of monitoring. It is also designed to better integrate with the RHS and other riverine monitoring schemes (Marchant *et al.* 2002).

1.3.1.2 *WBBS method*

WBBS methodology is derived from the transect counting method used in the Breeding Bird Survey. Two visits, at least four weeks apart, are made to each site during the breeding season (early April until late June). All birds seen or heard are recorded by distance category: 0–25m, 25–100m or >100m from the channel, or as a

fourth 'in-flight' category (Marchant *et al.* 2002). Transects run alongside the waterways and are divided into 500m sections, with birds recorded separately for each section: the overall survey length may be 1–10 sections (500m to 5-km).

Three phases of WBBS have occurred so far. In phase one (1998), 107 randomly selected sites were surveyed, as well as further sites to aid comparison with the WBS and sites for investigating the impact of removing the closed season to coarse fishing on canals (Marchant *et al.* 2002). During phase two (1999–2000) similar numbers of surveys were carried out on randomly located sites each year, such that over the three years, 144 sites were surveyed at least once (Marchant *et al.* 2002). Phase three (2001–2004) involves increasing the number of sites to improve the spatial coverage for population monitoring (Marchant *et al.* 2002).

1.3.2 THE RIVER HABITAT SURVEY

RHS was developed in the early to mid-1990s by the National Rivers Authority: later the EA (Fox, Naura & Scarlett 1998). The objective was to develop a robust, quantitative and objective method for describing river habitats throughout the UK, with particular focus on the ability to describe conservation value (Raven *et al.* 1997; Fox *et al.* 1998). RHS records detailed information about the channel flow and morphology, and vegetation and bank structure, as well as the vegetation and land use in a 50m wide corridor along both banks. The use of a single, standard methodology should facilitate communication between different users and interest groups.

Analogous schemes have been developed elsewhere in Europe, such as the Système d'Evaluation de la Qualité du Milieu Physique in France and the Länderarbeitsgemeinschaft Wasser in Germany (Raven *et al.* 2002).

The first application of RHS in 1994–1996 was blanket coverage of the UK, with three randomly-selected river sections sampled from every 10-km grid square across England and Wales (one per year) and at least one sample from every 10-km square in Scotland and Northern Ireland (Raven *et al.* 1997). This network of reference sites allows the geographical variability of rivers to be described nationally and individual sites to be placed in a broader context. Subsequently, RHS surveying has been carried out for a wide variety of projects, often concentrating upon particular catchments or fisheries. Most of these have been collated alongside the reference sites in the national RHS database, which contains in excess of 15 000 sites from across the UK (Version 3.3; Environment Agency 2002).

RHS has been employed in several ways. The first is essentially a descriptive role, using the standard survey format to provide site descriptions that can be compared to others nation-wide (Raven *et al.* 1998). The second, extending from its descriptive role, is the potential to develop river classifications. Previous classifications have generally used water chemistry or biota such as macroinvertebrates (e.g. Wright 1995) or macrophytes (e.g. Rodwell 1995; Holmes, Boon & Rowell 1998). Such approaches fail to reflect the particular importance that the physical environment has in defining the river habitat, however (Edwards 1997). Preliminary RHS-based classifications appear to successfully reflect intuitive river types (Raven *et al.* 1997) or geophysical characters (Newson *et al.* 1998).

A third application of RHS is in habitat assessment. Subjective scoring patterns for some of the variables within RHS – especially those describing anthropogenic modification – have been developed to allow overall habitat quality assessment and

habitat modification scores to be calculated (Raven *et al.* 1998). RHS data, with the addition of a suite of extra variables and a scoring system, provide the basis for the System for Evaluating Rivers for Conservation (SERCON; Wilkinson *et al.* 1998). A natural progression from its roles in river description, classification and assessment is to use RHS to define management objectives (Walker, Diamond & Naura 2002).

The ability to link RHS data directly to species' distributions and to use them to describe species' habitat requirements has been relatively little studied. At a fundamental level, it is important to demonstrate that species respond to the habitat patterns described by RHS because a basic tenet of the survey is to describe habitats in conservation terms (Raven *et al.* 2002). To be able to describe species' habitat requirements – and ultimately management recommendations – in the same terms used throughout river planning and management would be extremely valuable for conservation, as it may help to reduce the division often apparent between modellers and conservation practitioners (see Wiens 2002). Previous RHS-based modelling studies have examined river birds (Buckton & Ormerod 1997; Buckton *et al.* 1998; Manel *et al.* 1999a, 1999b) and benthic macroinvertebrates (Brewin, Buckton & Ormerod 1998; Manel *et al.* 2001), including crayfish (Naura & Robinson 1998) and molluscs (Hastie *et al.* 2003).

1.3.2.1 RHS method

In common with other riverine surveys (e.g. River Corridor Surveys, macrophyte surveys), RHS is based upon 500m stretches of waterway (Raven *et al.* 1997; Langston *et al.* 1997). Raven *et al.* (1997) and Fox *et al.* (1998) provide a detailed outline of the RHS method, with full instructions given to surveyors in the form of a

comprehensive manual (Environment Agency 1997): only a brief description is given here. The survey uses a four-page form, which is reproduced in Appendix 1.1.

The first section of the RHS records a series of map-derived variables, to provide background information about a site (Table 3). Field surveys comprise the majority of the RHS and take around an hour to complete, although this varies markedly between sites (Raven *et al.* 1997). Surveys are typically carried out in May and June (Environment Agency 1997). A two-stage method is used in the field: a series of so-called 'spot-checks' and an overall 'sweep-up'. In the field, these correspond to walking along the 500m reach and then back again.

Ten spot-checks are made, spread equidistantly along the 500m reach. At each one, an assessment of the river channel, banks and immediate land use is made for a 1m wide strip across the channel (10m for channel and bank vegetation; Table 3). Separate entries are used for the two banks of the river. Spot-check variables are generally recorded either as the predominant feature type from a predefined list (e.g. the predominant flow type or channel substratum within the 1m strip) or up to two features from a pre-defined list (e.g. bank and channel features). Channel vegetation types are given a simple ordinal cover score (absent, present <33% cover, present ≥33% cover) within a 10m stretch of river at each spot check (Table 3).

Sweep-up variables provide an holistic assessment of the characteristics of the river and neighbouring corridor, also ensuring that features falling between the 50m spot-

Table 3. Summary of the River Habitat Survey (1997 revision). The 19 sections on the RHS form are given, along with brief descriptions. ‘Spot-check’ and ‘sweep-up’ parts of the survey are marked where appropriate. For a full description, see Appendix 1.2.

A. BACKGROUND MAP-BASED INFORMATION

- 11 variables e.g. altitude, distance from source and water quality class

B. FIELD SURVEY DETAILS

- location and quality control variables, such as the surveyor’s details and conditions that may compromise the survey (e.g. very high or low flow levels)

C. PREDOMINANT VALLEY FORM – *sweep-up*

- selected from six basic shapes

D. NUMBER OF RIFFLES, POOLS AND POINT BARS – *sweep-up*

- counts over the 500m reach

E. PHYSICAL ATTRIBUTES – *spot-check*

- bank material, modifications and features (left and right bank separately)
- channel substratum, predominant flow type, channel modifications and features (e.g. bars, boulders)

F. BANKTOP LAND USE AND VEGETATION STRUCTURE – *spot-check*

- land use within 5m of channel (left and right bank separately)
- vegetation structure on the bank face and top of the bank (left and right bank separately)

G. CHANNEL VEGETATION TYPES – *spot-check*

- the cover of each of 10 vegetation types, defined by physiognomy (c.f. taxonomy)

H. LAND USE WITHIN 50M OF BANKTOP – *sweep-up*

- land use within 50m of channel (left and right bank separately)
-

I. BANK PROFILES – *sweep-up*

- extent and type of 13 different cross-sections (both natural and artificial)

J. EXTENT OF TREES AND ASSOCIATED FEATURES – *sweep-up*

- tree cover (left and right bank separately)
- six variables describing trees' influences upon the channel (e.g. extent of shading)

K. EXTENT OF CHANNEL FEATURES – *sweep-up*

- 19 variables describing different flow types and fluvial features (e.g. bars, islands, silt deposits)

L. CHANNEL DIMENSIONS – *sweep-up*

M. ARTIFICIAL FEATURES – *sweep-up*

- counts of different types of features (e.g. weirs, outfalls, revetments)

N. EVIDENCE OF RECENT MANAGEMENT – *sweep-up*

- occurrence of activities such as dredging and weed cutting

O. FEATURES OF SPECIAL INTEREST – *sweep-up*

- occurrence of 14 features of conservation interest (e.g. leafy debris in channel, water meadows, flushes)

P. CHOKED CHANNEL – *sweep-up*

- presence of vegetation restricting channel flow

Q. NOTABLE NUISANCE SPECIES – *sweep-up*

- occurrence of giant hogweed *Heracleum mantegazzianum*, Himalayan balsam *Impatiens glandulifera*, Japanese knotweed *Fallopia japonica* or other invasive species

R. OVERALL CHARACTERISTICS – *sweep-up*

- largely subjective site descriptions, using key-words to identify features not covered elsewhere (e.g. housing, afforestation, parkland)

S. ALDERS – *sweep-up*

- extent of alders *Alnus glutinosa* and proportion showing symptoms of *Phytophthora* disease
-

check intervals are recorded (Raven *et al.* 1997). Most of the sweep-up variables record the extent of features over the entire 500m reach, describing them either as absent, present, but accounting for less than 33% of the survey reach, or extensive ($\geq 33\%$). This includes the extent of different land uses within 50m of the channel, bank profiles (natural and artificial), trees and features of the channel (Table 3). The sweep-up also incorporates a range of information about anthropogenic modifications of the channel, a measurement of the channel cross section and a series of other features of interest, such as specific wildlife habitats and the presence of alien riparian weeds (Table 3; Appendix 1.2).

1.4 The RHS-WBBS data set

RHS surveys were carried out at 76% of WBBS sites between 1998 and 2000.

Consequently, a matched RHS-WBBS data set should have been available at the start of the current project (October 2000). Unfortunately, problems were encountered in linking the two data sets. Different numbers of 500m sections were recorded for RHS and WBBS on many rivers and the six-figure grid references used to match the sites were frequently different (Marchant *et al.* 2002). Part of the problem may have arisen from the field-based measurements of 500m sections used by RHS surveyors, compared with map-based allocation by BTO volunteers (Marchant *et al.* 2002). Map-derived sections are likely to be longer than field-based measures, especially on sinuous stretches of waterway.

Matching of the data sets took a prolonged period as a result of these problems and so the data set was received from the BTO in summer 2002. RHS and WBBS sections had been matched as closely as possible – giving the greatest possible overlap – based

upon the grid references supplied by both RHS and WBBS surveyors (Marchant *et al.* 2002). The final data set supplied by the BTO contained information on 28 river bird species separated into each of the three years and by three distance classes (0–25m, 25–100m and >100m). In total, 595 transects had both RHS and WBBS data, sampled from 109 waterways.

1.4.1 SUPPLEMENTARY DATA SETS USED IN THE THESIS

Three data sets were used in addition to WBBS-RHS data. The complete RHS database was used for some analyses (Chapters 4 & 5). The other data sets were derived from river bird surveys with matched RHS data, carried out in upland Wales and the Himalaya of India and Nepal between 1995 and 2002, by the Catchment Research Group, Cardiff University and collaborators. Details of all of these data sets are given in the relevant chapters (3, 4 & 5).

1.5 Project aims

The current project had two broad aims, *viz.* to contribute to the ongoing body of research concerning the development of distribution models and to relate the WBBS to RHS. Methodological issues were selected to benefit analysis of the RHS-WBBS data set. Specifically, five aims were identified:

1. *To identify how properties of the data use to create species distribution models may affect the performance and utility of the resulting models* (Chapter 2). Until recently, data collection specifically for distribution modelling was a neglected issue. In the last few years, however, several papers have appeared addressing such issues as sampling methods and the effects of sample size (e.g. Hirzel &

Guisan 2002; Stockwell & Peterson 2002). Even though the RHS-WBBS data were already collected, an understanding of the interaction between the performance of a model and the data used to create it provides important insight into the modelling process and helped to highlight likely problems with the RHS-WBBS.

2. *Examine how best to evaluate species distribution models – quantify their accuracy, diagnose problems and identify when and where they may be used* (Chapter 3). Like data collection issues, this was a largely neglected issue in distribution modelling for many years. More recently, a series of papers has addressed this issue (e.g. Fielding & Bell 1997; Boyce *et al.* 2002), but most have only examined a relatively limited set of issues. Aspects such as model generalisability have received little attention.

These two modelling issues (data collection and evaluation) were selected from the range of possibilities (Table 2) because of their generic importance and their relevance to the current project. Many of the other issues were ruled out by the nature of the RHS-WBBS data set. For example, the spatial and temporal scales were already fixed and the small sample size (109 river stretches) did not lend itself to a detailed comparison of different modelling methods.

3. *Develop a strategy for making RHS data more amenable to subsequent statistical analysis and modelling* (Chapter 4). The potential number of variables available in RHS data (>>100) coupled with the combination of quantitative and semi-quantitative (ordinal) variables, can make it difficult to analyse. An effective,

standard strategy for pre-processing RHS data could prove very valuable not only for RHS-based modelling, but also for a range of the other RHS applications.

4. *Analyse the WBBS-RHS data set (Chapter 5). Look at the potential to link river birds to RHS, propose preliminary models for predicting river bird distributions using RHS and explore how river bird distribution relates to RHS features.*

5. *Draw some preliminary conclusions about the potential value of the RHS to describe species' habitats and as a predictor of distribution patterns (Chapters 5 & 6).*

1.6 References

- Altman, D.G. (1994) The scandal of poor medical research. *British Medical Journal*, **308**, 283–284.
- Anderson, D.R., Burnham, K.P. & Thompson, W.L. (2000) Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management*, **64**, 912–923.
- Araújo, M.B. and Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.
- Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G. & Jeffcoate, S. (2001) *The Millennium Atlas of Butterflies in Britain and Ireland*. Oxford University Press, Oxford.
- Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (1998) The role of simulation in modelling spatially correlated data. *Environmetrics*, **9**, 175–196.
- Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Austin, M.P., Nicholls, A.O. & Margules, C.R. (1990) Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs*, **60**, 161–177.

Backhouse, D. & Burgess, L.W. (1995) Mycogeography of *Fusarium* – climatic analysis of the distribution within Australia of *Fusarium* species in section Gibbosum. *Mycological Research*, **99**, 1218–1224.

Bibby, C.J., Burgess, N.D., Hill, D.A. & Mustoe, S. (2000) *Bird Census Techniques* (2nd edn.). Academic Press, London.

Bolger, D.T. (1997) Response of rodents to habitat fragmentation in Coastal South California. *Ecological Applications*, **7**, 552–563.

Bonesi, L., Rushton, S. & Macdonald, D. (2002) The combined effect of environmental factors and neighbouring populations on the distribution and abundance of *Arvicola terrestris*. An approach using rule-based models. *OIKOS*, **99**, 220–230.

Boone, R.B. & Krohn, W.B. (1999) Modeling the occurrence of bird species: are the errors predictable? *Ecological Applications*, **9**, 835–848.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300.

Brewin, P.A., Buckton, S.T & Ormerod, S.J. (1998) River habitat surveys and biodiversity in acid-sensitive rivers. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **8**, 501–514.

Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology*, **30**, 478–495.

Buckton, S.T. (1998) *Spatio-temporal patterns in the distribution and ecology of river birds*. Unpublished PhD Thesis, University of Wales, Wales.

Buckton, S.T., Brewin, P.A., Lewis, A., Stevens, P. & Ormerod, S.J. (1998) The distribution of dippers, *Cinclus cinclus* (L.), in the acid-sensitive region of Wales, 1984–95. *Freshwater Biology*, **39**, 387–396.

Buckton, S.T. & Ormerod, S.J. (1997) Use of a new standardized habitat survey for assessing the habitat preferences and distribution of upland river birds. *Bird Study*, **44**, 327–337.

Buckton, S.T. & Ormerod, S.J. (2002) Global patterns of diversity among the specialist birds of riverine landscapes. *Freshwater Biology*, **47**, 695–709.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, New York.

Bustamante, J. (1997) Predictive models for Lesser Kestrel *Falco naumanni* distribution, abundance and extinction in southern Spain. *Biological Conservation*, **80**, 153–160.

Carter, S.P. (1989) The Waterways Bird Survey of the British Trust for Ornithology: an overview. *Regulated Rivers: Research & Management*, **4**, 191–197.

Chamberlain, D.E., Wilson, A.M., Browne, S.J. & Vickery, J.A. (1999) Effects of habitat type and management on the abundance of skylarks in the breeding season. *Journal of Applied Ecology*, **36**, 856–870.

Collingham, Y.C., Wadsworth, R.A., Huntley, B. & Hulme, P.E. (2000) Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology*, **37** (Suppl. 1), 13–27.

Copas, J.B. (1983) Regression, prediction and shrinkage. *Journal of the Royal Statistical Society, Series B*, **45**, 311–354.

Corsi, F., Dupré, E. & Boitani, L. (1999) A large-scale model of wolf distribution in Italy for conservation planning. *Conservation Biology*, **13**, 150–159.

D'Amico, F. (2002) High reliability of linear censusing for Common Sandpiper (*Acitits hypoleucos*) breeding along upland streams in the Pyrenees, France. *Bird Study*, **49**, 307–309.

D'Amico, F. & Hemery, G. (2003) Calculating census efficiency for river birds: a case study with the white-throated Dipper *Cinclus cinclus* in the Pyrénées. *Ibis*, **145**, 83–86.

D'Angelo, D.J., Howard, L.M., Meyer, J.L., Gregory, S.V. & Ashkenas, L.R. (1995) Ecological uses for genetic algorithms: predicting fish distributions in complex physical habitats. *Canadian Journal of Fisheries and Aquatic Science*, **52**, 1893–1908.

Dawson, F.H. & Holland, D. (1999) The distribution in bankside habitats of three alien invasive plants in the UK in relation to the development of control strategies. *Hydrobiologia*, **415**, 193–201.

Donázar, J.A., Hiraldo, F. & Bustamante, J. (1993) Factors influencing nest site selection, breeding density and breeding success in the bearded vulture (*Gypaetus barbatus*). *Journal of Applied Ecology*, **30**, 504–514.

Edwards, R.W. (1997) Introduction. *Freshwater Quality: Defining the Indefinable?* (eds P. J. Boon & D. L. Howell), pp. 1–6. The Stationery Office, Edinburgh.

Environment Agency (1997) *River Habitat Survey: Field Survey and Guidance Manual*. Environment Agency, Bristol.

Erasmus, B.F.N., Van Jaarsveld, A.S., Chown, S.L., Kshatriya, M. & Wessels, K.J. (2002) Vulnerability of South African animal taxa to climate change. *Global Change Biology*, **8**, 679–693.

Ferrier, S., Watson, G., Pearce, J. & Drielsma, M. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species level modelling. *Biodiversity and Conservation*, **11**, 2275–2307.

Fielding, A.H. (1999) An introduction to machine learning methods. *Machine Learning Methods for Ecological Applications* (ed A. H. Fielding) pp 1–35. Kluwer Academic Publishers, Boston.

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

Fielding, A.H. & Haworth, P.F. (1995) Testing the generality of bird-habitat models. *Conservation Biology*, **9**, 1466–1481.

Fox, P.J.A., Naura, M. & Scarlett, P. (1998) An account of the derivation and testing of a standard field method, River Habitat Survey. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 455–475.

Franco, A.M.A., Brito, J.C. & Almeida, J. (2000) Modelling habitat selection of common cranes *Grus grus* wintering in Portugal using multiple logistic regression. *Ibis*, **142**, 351–358.

Franklin, J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474–499.

Franklin, J. (1998) Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science*, **9**, 733–748.

Gibbons, D.W., Reid, J.B. & Chapman, R.A. (1993) *The New Atlas of Breeding Birds in Britain and Ireland: 1988–1991*. T. & A. D. Poyser, London.

Gioia, P. & Pigott, J.P. (2000) Biodiversity assessment: a case study in predicting richness from the potential distributions of plant species in the forests of south-western Australia. *Journal of Biogeography*, **27**, 1065–1078.

Green, R.H. (1971) A multivariate statistical approach to the Hutchinsonian niche: bivalve molluscs of central Canada. *Ecology*, **52**, 543–556.

Gregory, R.D., Carter, S.P. and Baillie, S.R. (1997) Abundance, distribution and habitat use of breeding Goosanders *Mergus merganser* and Red-breasted Mergansers *Mergus serrator* on British rivers. *Bird Study*, **44**, 1–12.

Groom, J.D. & Grubb, T.C. (2002) Bird species associated with riparian woodland in fragmented, temperate-deciduous forest. *Conservation Biology*, **16**, 832–836.

Guerry, A.D. & Hunter, M.L. (2002) Amphibian distributions in a landscape of forests and agriculture: an examination of landscape composition and configuration. *Conservation Biology*, **16**, 745–754.

Guisan, A., Edwards, T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89–100.

Guisan, A. & Theurillat, J.-P. (2000) Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia*, **30**, 353–384.

Hastie, L.C., Cooksley, S.L., Scougall, F., Young, M.R., Boon, P.J. & Gaywood, M.J. (2003) Characterization of freshwater pearl mussel (*Margaritifera margaritifera*) riverine habitat using River Habitat Survey data. *Aquatic Conservation: marine and freshwater ecosystems*, **13**, 213–224.

Hatten, J.R. & Paradzick, C.E. (2003) A multiscaled model of southwestern willow flycatcher breeding habitat. *Journal of Wildlife Management*, **67**, 774–788.

Hirzel, A. & Guisan, A. (2002) Which is the optimal strategy for habitat suitability modelling? *Ecological Modelling*, **157**, 331–341.

Holmes, N.T.H., Boon, P.J. & Rowell, T.A. (1998) A revised classification system for British rivers based on their aquatic plant communities. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 555–578.

Huston, M.A. (2002) Critical issues for improving predictions. *Predicting Species*

Occurrences: issues of accuracy and scale (eds J.M. Scott, P.J. Heglund, M.L.

Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 7–21. Island Press, Washington.

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposium on*

Quantitative Biology, **22**, 415–427.

Karl, J.W., Svancara, L.K., Heglund, P.J., Wright, N.M. & Scott, M.J. (2002) Species

commonness and the accuracy of habitat relationship models. *Predicting Species*

Occurrences: issues of accuracy and scale (eds J.M. Scott, P.J. Heglund, M.L.

Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 573–580.

Island Press, Washington.

Knapp, R.A., Matthews, K.R., Preisler, H.K. & Jellison, R. (2003) Developing

probabilistic models to predict amphibian site occupancy in a patchy landscape.

Ecological Applications, **13**, 1069–1082.

Langston, R.H.W., Marchant, J.H. & Gregory, R.D. (1997) Waterways Bird Survey:

evaluation of population monitoring and appraisal of future requirements. *Freshwater*

Quality: Defining the Indefinable? (eds P.J. Boon & D.L. Howell), pp. 282–289. The

Stationery Office, Edinburgh.

Lawton, J.H. (1996) Population abundances, geographic ranges and conservation: 1994 Witherby lecture. *Bird Study*, **43**, 3–19.

Lawton, J.H. and Woodroffe, G.L. (1991) Habitat and the distribution of water voles: why are there gaps in a species' range? *Journal of Applied Ecology*, **60**, 79–91.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., & Aulanier, S. (1996) Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, **90**, 39–52.

Lek-Ang, S., Deharveng, L., & Lek, S. (1999) Predictive models of collembolan diversity and abundance in a riparian habitat. *Ecological Modelling*, **120**, 247–260.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, **72**, 445–463.

Lindblad, M. (2001) Development and evaluation of a logistic risk model: predicting frit fly infestation in oats. *Ecological Applications*, **11**, 1563–1572.

Lindenmayer, D.B, Cunningham, R.B., Tanton, M.T., Nix, H.A. & Smith, A.P. (1991) The conservation of arboreal marsupials in the montane ash forests of the Central Highlands of Victoria, South-East Australia: III. The habitat requirements of Leadbeater's Possum *Gymnobelideus leadbeateri* and models of the diversity and abundance of arboreal marsupials. *Biological Conservation* **56**, 295–315.

Logan, P. & Furse, M. (2002) Preparing for the European Water Framework Directive – making the links between habitat and aquatic biota. *Aquatic Conservation: marine and freshwater ecosystems*, **12**, 425–437.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G. & Williams, P.H. (2003) Avoiding pitfalls in using species distribution models in conservation planning. *Conservation Biology*, **17**, 1591–1600.

Luoto, M., Kuussaari, M., & Toivonen, T. (2002) Modelling butterfly distribution based on remote sensing data. *Journal of Biogeography*, **29**, 1027–1037.

McCune, B., Berryman, S.D., Cissel, J.H. & Gitelman, A.I. (2003) Use of a smoother to forecast occurrence of epiphytic lichens under alternative forest management plans. *Ecological Applications*, **13**, 1110–1123.

McPherson, J.M., Jetz, W. & Rogers, D.J. (in press) The effects of species' range size on the accuracy of environmental distribution models – ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, in press.

Mackey, B.G. (1994) A spatial analysis of the environmental relations of rainforest structural types. *Journal of Biogeography*, **20**, 303–336.

Mackey, B.G. & Lindenmayer, D.B. (2001) Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography*, **28**, 1147–1166.

Mac Nally, R. (2000) Regression and model-building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – ‘predictive and ‘explanatory’ models. *Biodiversity and Conservation*, **9**, 655–671.

Maier, H.R., Dandy, G.C. & Burch, M.D. (1998) Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. *Ecological Modelling*, **105**, 257–272.

Manel, S., Buckton, S.T. & Ormerod, S.J. (2000) Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan rivers. *Journal of Applied Ecology*, **37**, 756–770.

Manel, S., Dias, J.M., Buckton, S.T. & Ormerod, S.J. (1999a) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734–747.

Manel, S., Dias, J.M. & Ormerod, S.J. (1999b) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, **120**, 337–347.

- Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921–931.
- Marchant, J.H. & Coombes, R. (2003) Breeding birds on waterways. *BTO News*, **249**, 22–23.
- Marchant, J.H. & Hyde, P.A. (1980a) Population changes for waterways birds, 1978–79. *Bird Study*, **27**, 179–182.
- Marchant, J.H. & Hyde, P.A. (1980b) Aspects of the distribution of riparian birds on the waterways in Britain and Ireland. *Bird Study*, **27**, 183–202.
- Marchant, J.H., Noble, D.G., Leech, D.I & Freeman, S.N. (2002) *River Habitat Survey and Waterways Breeding Bird Survey 1999–2000: final report*. R&D technical report W1-043/TR, BTO Research Report 291.
- Margules, C.R. & Austin, M.P. (1994) Biological models for monitoring species decline: the construction and use of data bases. *Philosophical Transactions of the Royal Society of London, Series B*, **344**, 69–75.
- Miller, J. & Franklin, J. (2002) Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*, **157**, 227–247.

Miller, M.E., Hui, S.L. & Tierney, W.M. (1991) Validation techniques for logistic regression models. *Statistics in Medicine*, **10**, 1213–1226.

Mladenoff, D.J., Sickley, T.A., Haight, R.G. & Wydeven, A.P. (1995) A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great Lakes region. *Conservation Biology*, **9**, 279–294.

Naura, M. and Robinson, M. (1998) Principles of using River Habitat Survey to predict the distribution of aquatic species: an example applied to the native white-clawed crayfish *Austropotamobius pallipes*. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 515–527.

Newson, M.D., Clark, M.J., Sear, D.A. & Brookes, A. (1998) The geomorphological basis for classifying rivers. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 415–430.

O'Connor, R.J. (2002) The conceptual basis of species distribution modeling: time for a paradigm shift? *Predicting Species Occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 25–33. Island Press, Washington.

O'Halloran, J., Irwin, S., Harrison, S., Smiddy, P. & O'Mahony, B. (2003) Mercury and organochlorine content of Dipper *Cinclus cinclus* eggs in south-west Ireland: trends during 1990-1999. *Environmental Pollution*, **123**, 85–93.

Olden, J.D. & Jackson, D.A. (2002) A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, **47**, 1976–1995.

Ormerod, S.J., Allinson, N., Hudson, D. & Tyler, S.J. (1986) The distribution of breeding dippers (*Cinclus cinclus* (L), Aves) in relation to stream biology in upland Wales. *Freshwater Biology*, **16**, 501–507.

Ormerod, S.J. & Tyler, S.J. (1990) Environmental pollutants in the eggs of Welsh dippers *Cinclus cinclus* – a potential monitor of organochlorine and mercury contamination in upland rivers. *Bird Study*, **37**, 171–176.

Ormerod, S.J. & Tyler, S.J. (1993) Birds as indicators of changes in water quality. *Birds as Monitors of Environmental Change* (eds R.W. Furness & J.J.D. Greenwood) pp 179–216. Chapman & Hall, London.

Ormerod, S.J., Tyler, S.J. & Jüttner, I. (2000) Effects of point-source PCB contamination on breeding performance and post-fledging survival in the dipper *Cinclus cinclus*. *Environmental Pollution*, **109**, 505–513.

Osborne, P.E., Alonso, J.C. & Bryant, R.G. (2001) Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology*, **38**, 458–471.

Osborne, P.E. & Suárez-Seoane, S. (2002) Should data be partitioned spatially before building large-scale distribution models? *Ecological Modelling*, **157**, 249–259.

Paradis, E., Baillie, S.R., Sutherland, W.J., Dudley, C., Crick, H.Q.P. & Gregory, R.D. (2000) Large-scale spatial variation in the breeding performance of song thrushes *Turdus philomelos* and blackbirds *T. merula* in Britain. *Journal of Applied Ecology*, **37** (Suppl. 1), 73–87.

Pearce, J.L., Cherry, K., Drielsma, M., Ferrier, S. & Wish, G. (2001) Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. *Journal of Applied Ecology*, **38**, 412–423.

Pearce, J. and Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.

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

Raven, P.J., Fox, P., Everard, M., Holmes, N.T.H & Dawson, F.H. (1997) River Habitat Survey: a new system for classifying rivers according to their habitat quality. *Freshwater Quality: Defining the Indefinable?* (eds P.J. Boon & D.L. Howell), pp. 215–234. The Stationery Office, Edinburgh.

Raven, P.J., Holmes, N.T.H., Charrier, P., Dawson, F.H., Naura, M. & Boon, P.J. (2002) Towards a harmonized approach for hydromorphological assessment of rivers

in Europe: a qualitative comparison of three survey methods. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **12**, 405–424.

Raven, P.J., Holmes, N.T.H., Dawson, F.H. & Everard, M. (1998) Quality assessment using River Habitat Survey data. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 477–499.

Ripley, B.D. (1994) Neural networks and related methods for classification. *Journal of the Royal Statistical Society, Series B*, **56**, 409–456.

Rodwell, J.S. (1995) *British Plant Communities Volume 4: Aquatic communities, swamps and tall-herb fens*. Cambridge University Press, Cambridge.

RSPB, NRA & RSNC (1994) *The New Rivers and Wildlife Handbook*. RSPB, Sandy.

Rushton, S.P., Barreto, G.W., Cormack, R.M., Macdonald, D.W. & Fuller, R. (2000) Modelling the effects of mink and habitat fragmentation on the water vole. *Journal of Applied Ecology*, **37**, 475–490.

Rushton, S.P., Hill, D. & Carter, S.P. (1994) The abundance of river corridor birds in relation to their habitat: a modelling approach. *Journal of Applied Ecology*, **31**, 313–328.

Rushton, S.P., Ormerod, S.J. & Kerby, G. (2004) New paradigms for modelling species' distributions? *Journal of Applied Ecology*, **41**, 193–200.

Sarre, S., Smith, G.T. & Meyers, J.A. (1995) Persistence of 2 species of Gecko (*Oedura reticulata* and *Gehyra variegata*) in remnant habitat. *Biological Conservation*, **71**, 25–33.

Scribner, K.T., Arntzen, J.W., Cruddance, N., Oldham, R.S. & Burke, T. (2001) Environmental correlates of toad abundance and population genetic diversity. *Biological Conservation*, **98**, 201–210.

Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. & Samson, F.B. (2002) *Predicting Species Occurrences: issues of accuracy and scale*. Island Press, Washington.

Snow, D.W. & Perrins, C.M. (eds) (1998) *The Birds of the Western Palearctic. Concise Edition*. Oxford University Press, Oxford.

Stauffer, D.F. (2002) Linking populations and habitats: where have we been? Where are we going? *Predicting Species Occurrences: issues of accuracy and scale* (eds J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall & F. B. Samson), pp. 53–61. Island Press, Washington.

Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1–13.

Strong, T.R. & Bock, C.E. (1999) Bird species distribution patterns in riparian habitats in south-eastern Arizona. *Condor*, **92**, 866–885.

Suárez-Seoane, S., Osborne, P.E. & Alonso, J.C. (2002) Large-scale habitat selection by agricultural steppe birds in Spain: identifying species-habitat responses using generalized additive models. *Journal of Applied Ecology*, **39**, 755–771.

Suren, A.M. & Ormerod, S.J. (1998) Aquatic bryophytes in Himalayan streams: testing a distribution model in a highly heterogeneous environment. *Freshwater Biology*, **40**, 697–716.

Sutherland, W.J. (1996) *From Individual Behaviour to Population Ecology*. Oxford University Press, Oxford.

Sykes, M.T., Prentice, I.C, Smith, B., Cramer, W. & Venevsky, S. (2001) An introduction to the European Terrestrial Ecosystem Modelling Activity. *Global Ecology & Biogeography*, **10**, 581–593.

Taylor, K. (1984) The influence of watercourse management on Moorhen breeding biology. *British Birds*, **77**, 144–148.

Ter Braak, C.J.F. & Van Dam, H. (1989) Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia*, **178**, 209–223.

Thomas, C.D., Bodsworth, E.J., Wilson, R.J., Simmons, A.D., Davies, Z.G., Musche, M. & Conradt, L. (2001) Ecological and evolutionary processes at expanding range margins. *Nature*, **411**, 577–581.

Thuiller, W., Araújo, M.B. & Lavorel, S. (2003) Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, **14**, 669–680.

Treves, A., Naughton-Treves, L., Harper, E.K., Mladenoff, D.J., Rose, R.A., Sickley, T.A., & Wydeven, A.P. (2004) Predicting human-carnivore conflict: a spatial model derived from 25 years of data on wolf predation on livestock. *Conservation Biology*, **18**, 114–125.

Tushabe, H., Reynolds, J. & Pomeroy, D. (2001) Innovative aspects of the bird atlas of Uganda. *Ostrich* (Suppl. 15), 183–188.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003) Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, **13**, 1790–1801.

Vickery, J. (1991) Breeding density of Dippers *Cinclus cinclus*, Grey Wagtails *Motacilla cinerea* and Common Sandpipers *Actitis hypoleucos* in relation to the acidity of streams in south-west Scotland. *Ibis*, **133**, 178–185.

Walker, J., Diamond, M. & Naura, M. (2002) The development of Physical Quality Objectives for rivers in England and Wales. *Aquatic Conservation: marine and freshwater ecosystems*, **12**, 381–390.

Wernham, C.V., Toms, M.P., Marchant, J.H., Clark, J.A., Siriwardena, G.M. & Baillie, S.R. (2002) *The Migration Atlas: movements of the Birds of Britain and Ireland*. T. & A. D. Poyser, London.

Wiens, J.A. (2002) Predicting species occurrences: progress, problems and prospects. *Predicting Species Occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 739–749. Island Press, Washington.

Wilkinson, J., Martin, J., Boon, P.J. & Holmes, N.T.H. (1998) Convergence of field survey protocols for SERCON (System for Evaluating Rivers for Conservation) and RHS (River Habitat Survey). *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 579–596.

Wiser, S.K., Peet, R.K. & White, P.S. (1998) Prediction of rare-plant occurrence: a southern Appalachian example. *Ecological Applications*, **8**, 909–920.

Woinarski, J.C.Z., Brock, C., Armstrong, M., Hempel, C., Cheal, D. & Brennan, K. (2000) Bird distribution in riparian vegetation in the extensive natural landscape of Australia's tropical savanna: a broad-scale survey and analysis of a distributional data base. *Journal of Biogeography*, **27**, 843–868.

Wright, J.F. (1995) Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology*, **20**, 181–197.

Yalden, D.W. (1986) The habitat and activity of Common Sandpipers *Actitis hypoleucos* breeding by upland rivers. *Bird Study*, **33**, 214–222.

Zimmermann, N.E. & Kienast, F. (1999) Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science*, **10**, 469–482.

Chapter 2

**Improving the quality of distribution models for
conservation by addressing shortcomings in the field
collection of training data**

2.1 Summary

1. Conservation biology can benefit greatly from models that relate species' distributions to their environments. The foundation of successful modelling is a high-quality set of field data, and distribution models have specialised data requirements.
2. The role of a distribution model may be primarily predictive or, alternatively, may emphasise relationships between an organism and its habitat. For the latter application, the environmental variables recorded should have direct, biological relationships with the organism. Interacting species may be valuable predictors and can improve understanding of distribution patterns.
3. Sampling should cover the full range of environmental conditions within the study region, with samples stratified across major environmental gradients to ensure thorough coverage. Failure to sample correctly can lead to erroneous organism-environment relationships, affecting predictive ability and interpretation. Sampling ideally should examine a series of spatial scales, increasing the understanding of organism-environment relationships, identifying the most effective scales for predictive modelling and complementing the spatial hierarchies often used in conservation planning.
4. Consideration of statistical issues could benefit most studies. The ratio of sample sites to environmental variables considered should ideally exceed 10:1, to improve the analytical power and reliability of subsequent modelling. Presence/absence models may be biased if training data detect the study organism at an atypical proportion of sites.

5. Different strategies for spatial autocorrelation were considered and it is recommended that it be included wherever possible for the benefits in biological realism, predictive accuracy and model versatility.
6. Finally, I stress the importance of collecting independent evaluation data and suggest that, as with training data, a systematic approach be used to ensure broad environmental coverage, rather than relying on a random selection of test sites.

Key words: analytical power, distribution modelling, environmental space, model evaluation, sampling scale, species prediction

2.2 Introduction

Relationships between organisms and their environment are fundamental to ecology (Begon, Harper & Townsend 1996). In the past, detailed autecology was the foundation for studying such relationships, generally over limited spatial extents. As the scope of conservation biology grows to cover huge spatial scales, capture global environmental change and make itself relevant to remote regions of the world, so models to relate organisms to their environments are needed to fulfil this role (Fielding & Haworth 1995; Gaston & Blackburn 1999; Manel *et al.* 1999, 2000). Once carefully validated, distribution models can provide useful insights into species/environment relationships and represent valuable tools in conservation planning (e.g. Mladenoff, Sickley & Wydeven 1999; Franco, Brito & Almeida 2000; Guisan & Theurillat 2000; Clevenger *et al.* 2002; Woolf *et al.* 2002). Such models are now being applied in conservation, from studies concerned with single species to those predicting whole communities and ecosystem attributes such as diversity and vegetation physiognomy (Franklin 1995). Two recent reviews examine the diversity of these modelling techniques and their applications (Franklin 1995; Guisan & Zimmermann 2000).

With so much at stake biologically and the need to target conservation resources as efficiently as possible, it is imperative that the methodology for distribution modelling be optimised. A good set of training data is the basic foundation for successful distribution modelling, yet there is a surprising paucity of guidance for such data collection, especially compared with the wealth of general fieldwork texts. Here I address this shortcoming, discussing a series of field-survey ideas that could benefit distribution modelling. Problems with training data could be contributing to a range of

Table 1. Major problems experienced in the use of distribution models and ways in which shortcomings in the training data may contribute.

Problem	Possible data-related causes
Poor predictive ability (either reclassifying training data or application to new data)	<ul style="list-style-type: none">• Failure to record important predictor variables• Failure to take account of interacting species• Predictor variables recorded at spatial scales where they correlate weakly with distribution• Poor coverage of environmental space during sampling, leading to inaccurate characterisation of species-environment relationships• Training data unrepresentative of the wider environment• Statistical weaknesses in the training data (e.g. too few sampling sites for the number of predictor variables analysed)• Spatial autocorrelation overlooked, ignoring a potentially important predictor and reducing the reliability of modelling (e.g. increased risk of including spurious variables)
Consistent over- or under-optimism in predictions with new data (presence/absence models)	<ul style="list-style-type: none">• Study organism recorded at an atypically high or low proportion of sites in the training data compared with wider environment

Apparent responses to environmental conditions
inconsistent with biological understanding

- Relationships between organism and predictor variables poorly sampled and so inaccurately modelled
- Failure to record critical environmental variables
- Failure to take account of interacting species
- Training data poorly suited to model development (e.g. too few sampling sites for the number of variables analysed)
- Spatial autocorrelation overlooked; local processes generating autocorrelation (e.g. dispersal), may modify correlations between organisms and environmental variables

Organism-environment relationships difficult to
interpret

- Use of environmental variables that have indirect relationships with the distribution of the organism, rather than directly affecting it, necessitating an additional level of interpretation of organism-environment relationships
-

weaknesses observed in distribution models (Table 1). Further support for this idea comes from studies showing similar predictive performance for radically different model types trained with the same data set (e.g. discriminant analysis and logistic regression compared with artificial neural networks, Manel *et al.* 1999). My hope is that by illustrating some of the problems and potential solutions, the quality of the training data can be improved, thereby underpinning analytical success.

2.3 Clarification of model aims

All distribution models attempt to quantify the relationships between organisms and their environments. A dichotomy exists, however, between emphasis placed primarily on absolute predictive ability of the model and one on interpretation of the structure of the model – exploring the organism-environment relationship (Ripley 1994; Mac Nally 2000). The nature of the study dictates whether prediction, explanation/interpretation, or as is more commonly the case, both, approaches are of interest. During subsequent model development, the analysis may benefit from different methods for these two roles (Mac Nally 2000). At the point of fieldwork design, the subsequent aim of the study should be considered when the choice is made of which environmental variables to record.

Many environmental variables correlate with species' distribution patterns. All such variables could theoretically be employed for predictive modelling, but to improve the understanding of an organism-environment relationship, variables whose correlations represent actual biological relationships are required – so-called direct variables (Table 2). Working on Himalayan river birds, Manel *et al.* (1999) showed that distribution models based on altitude and slope had, in many instances, predictive

Table 2. The choice between indirect and direct predictor variables in the design of field studies^a.

	Direct predictors ^b	Indirect predictors
Definition	Variables with biological relationships with the study species	Variables whose correlations with the study species represent correlations with a series of intermediate, direct factors, rather than a direct relationship
Examples	Climate, nesting sites, soil nutrients (plants), interacting species, site isolation	Altitude, topography, geographical position, geology, soil nutrients (animals)
Strengths	<ul style="list-style-type: none"> • Structure of the model more easily interpreted in biologically meaningful terms, potentially greater ecological value • Direct biological relationships should generalise better to new areas and be more effective for use in climate/habitat change scenarios than indirect ones • Provides more valuable information for conservation management 	<ul style="list-style-type: none"> • Data sets often widely available <i>a priori</i>, in maps or GIS • Low cost, ease of collection • Can be effective predictors (e.g. altitude/topography in mountainous areas) • Encompass a range of correlated variables so should: <ul style="list-style-type: none"> - reduce multicollinearity problems - increase stability of model development as fewer candidate variables are considered - result in more parsimonious models - save resources by recording fewer variables

Weaknesses	<ul style="list-style-type: none">• Variables often require greater effort to record, especially for biotic variables• Several variables may need to be recorded where a single indirect predictor (e.g. altitude) would suffice• Data sets may need to be estimated for large spatial extents, often using indirect variables, reducing their overall accuracy (e.g. estimation of climatic data based on altitude)	<ul style="list-style-type: none">• Correlation with direct variables tends to be location-specific, making models particularly weak in extrapolation (e.g. latitudinal change in altitude/climate correlations)• Limited value for interpretation – biological meaning has to be read into the relationships, adding a further level of interpretation, thereby increasing uncertainty
------------	--	--

^aSources: Austin (1980); Austin *et al.* (1984); Austin & Meyers (1996); Franklin (1995); Guisan & Zimmermann (2000)

^bThis is a combination of Austin's (1980) direct and resource gradients, to simplify the generalisation of this framework across all distribution modelling.

abilities similar to those of models based on detailed habitat data. For explanatory applications, however, such models are of little value, necessitating further investigation into the correlations between altitude and slope and river habitat. Variables such as altitude and slope are therefore considered indirect (Table 2).

Indirect variables can be useful in predictive modelling, most notably for the increases in overall survey efficiency they provide (Table 2). The ability to use map-derived variables to predict river bird distribution in the Himalaya, rather than having to visit each site on the ground in difficult terrain, would be highly beneficial (Manel *et al.* 1999). Nevertheless, direct predictors are favoured in most cases for their reliability and generality (Austin & Meyers 1996). This stems from their biological relationships with the organism, which should remain relatively constant compared with indirect variables, whose correlations with direct variables by which their relationships to the organism actually occur, are location specific (Table 2; Austin *et al.* 1984). Studies encompassing large spatial extents, perhaps looking at the complete biogeographical range of an organism, should place particular importance on the use of direct predictors, where feasible. Similarly, correlations between indirect and direct variables may alter over time following major environmental change, in which case direct predictors are again preferred.

The distributions of interacting species should also be considered as possible variables in distribution modelling. Competing species, generalist predators, parasites, and pathogens can exclude a species from a region, either locally or at broad spatial scales (Begon *et al.* 1996). As such, they may account for a proportion of the unexplained variation in the distribution pattern that remains once a model based purely on the

physical environment is fitted (Table 1). The extra proportion such interacting species explains represents a difference between the fundamental and realised niches. The former considers simply the range of environmental conditions the organism can tolerate, whilst the latter also accounts for the regions from which the organism is excluded by other species (Hutchinson 1957). Field data always pertain to the realised niche (Austin, Nicholls & Margules 1990; Westman 1991), and failure to account for this may weaken the analytical process, confound interpretation and generate false positive (commission, type I) prediction errors.

In many cases, data describing the distributions of interacting species could be recorded with little extra effort. Similar organisms may be recorded in the same sampling technique, such as potential avian competitors in a standard point count or transect. Where competitors or predators have been included in distribution models, valuable increases in predictive ability and especially explanatory value have been observed. Özesmi and Özesmi (1999) showed major improvements in modelling success if competing wetland breeding birds were considered in their model.

Similarly, distance to the nearest golden eagle (*Aquila chrysaetos*) nesting cliff proved to be a useful predictor for peregrine (*Falco peregrinus*) nesting cliff selection (Gainzarain, Arambarria & Rodríguez 2000). Lowe and Bolger (2002) showed how piscivorous predators were valuable predictors of salamander abundance. Rushton *et al.* (2000) showed a similar effect with a nonindigenous predator, the American mink (*Mustela vison*), on the distribution of the water voles (*Arvicola terrestris*) in the United Kingdom. There is also good potential to develop these simplistic approaches for plant distribution modelling, in instances where distribution patterns are strongly affected by small numbers of easily recorded species. For example, the populations of

certain mistletoe species in New Zealand are being affected by reduced populations of their avian pollinators (Robertson *et al.* 1999), while introduced herbivores and livestock can affect floristic composition in a range of habitats (e.g. Veblen *et al.* 1992; Nugent, Fraser & Sweetapple 2001).

2.4 Accurately capturing organism-environment relationships

Wildlife-habitat models define the distribution of a species within environmental space – a multivariate concept covering all possible combinations of environmental conditions, rather than individual environmental gradients (Margules & Stein 1989). To accurately capture this distribution, samples need to be taken across the complete region of environmental space within which the organism occurs, including sites that define the boundaries of the organism's niche. Failure to do so undermines the analytical process and could lead to serious errors in both explanatory work and predictive applications (Table 1).

The effects of incomplete sampling are clearly illustrated when organism-environment relationships are visualised in terms of species response curves to individual gradients, as is widely done in plant distribution modelling. If, for example, a species shows a simple Gaussian response to an environmental gradient, incomplete sampling that overlooked the upper end of the gradient may result in a sigmoid curve being fitted (Fig. 1a). Erroneous predictions of occurrence (Type I or commission errors) would then be made in applying the model to sites high up on the gradient. This type of sampling-generated problem can be seen in practice where models are built from a limited geographic region, especially the edge of a species' range, and then applied more widely. Similarly, missing both ends of a gradient may provide the erroneous

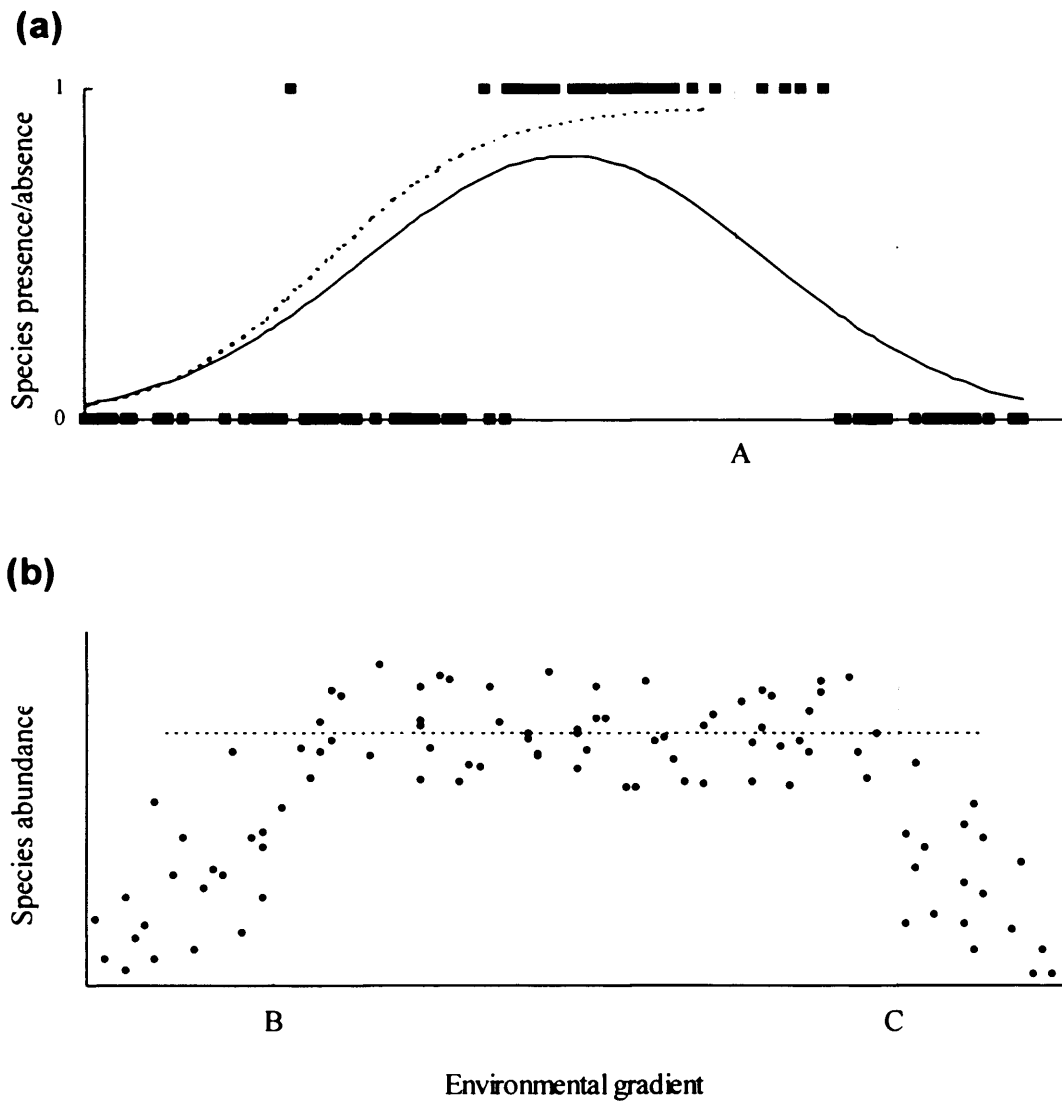


Figure 1. The potential impact of incomplete sampling of an environmental gradient upon subsequent interpretation. In these hypothetical examples, the occurrence or abundance of a species is recorded along a single environmental gradient. In (a), sampling only as far as point A, missing the top end of the gradient, could lead to a sigmoid response curve being fitted (dashed line) in place of the Gaussian response curve (solid line). In (b), sampling only between points B and C would lead to the erroneous conclusion that the environmental factor has no relationship with the organism (dashed line).

impression that the organism shows no response to that environmental factor, when, again, complete sampling would reveal a Gaussian response (Fig. 1b). Austin, Cunningham & Fleming (1984) provide an empirical example of sampling effects, obtaining different response curves for a *Eucalyptus* species depending upon the completeness with which a rainfall gradient was sampled. Further discussion is provided by Mohler (1983) and Westman (1991). The implications for sampling design are profound: through different combinations of oversampling and undersampling regions of environmental space, any response curve could theoretically be obtained (Yee & Mitchell 1991).

Environmental stratification is the only systematic approach to sampling environmental space (Margules & Austin 1994). Major environmental gradients are identified, such as altitude and geology, and each is divided into a series of strata at regular intervals. Stratification then attempts to place samples within every possible combination of the strata from these environmental gradients. For example, with five altitude and four geological categories, there are 20 potential strata to sample. This number may become very large for stratification by several gradients, but fortunately only a fraction usually occur in the environment (e.g. 215 of 504 possible strata in a forested region of New South Wales [Austin & Heyligers 1989] or 15 of 20 broad climatic strata in southern Australian forest [Neave, Norton & Nix 1996]). Sampling sites should be allocated equally between all strata present, with several replicates in each, to provide robust, wide coverage of environmental space for later analysis (Guisan & Zimmermann 2000). Hirzel & Guisan (2002) demonstrate how, for a fixed sample size, stratifying sites in this way improves model performance compared with randomised sampling.

Successful environmental stratification is conditional on identifying the complete expanse of environmental space found within the study region to ensure comprehensive sampling. Every region in which a predictive model may be applied must be identified, and the geographical limit of explanatory conclusions clarified. Collingham *et al.* (2000) demonstrated the consequences of failing to account for the region of model application. They constructed a predictive model for alien riparian weeds at a regional scale (263 km²), then applied it nationally to England and Wales. Performance was poor, especially when compared with a model built at the national scale and applied regionally (Collingham *et al.* 2000).

The basic sampling protocol may be further customised to the individual needs of the study, conserving precious financial resources, especially when large spatial scales are involved. An increased level of subjectivity may be introduced, but Austin and Heyligers (1989) argue that this should not be a problem provided that sampling rules are stated explicitly so that any biases are evident. Two ideas in particular are gaining recognition. The first is gradsect sampling, in which sampling is limited to discrete areas selected to contain maximum environmental heterogeneity to minimise travelling between sites. Empirical work collecting biological inventories demonstrated similar performance between gradsect sampling and full environmental stratification across the area (Wessels *et al.* 1998). The second is the use of rules to limit the resources expended in sampling each site. An example is specifying a maximum distance from the road network at which a site will still be sampled (e.g. Austin & Heyligers 1989). Examples of environmental stratification methods and

gradsect sampling can be found in Gillison and Brewer (1985), Austin and Heyligers (1989), Neave *et al.* (1996) and Wessels *et al.* (1998).

2.5 Sampling scale for analysing organism-environment relationships

The correlations observed between organisms and their environments are frequently dependent on the spatial scale of observation and may vary between scales in a complex manner (e.g. Wiens, Rotenberry & van Horne 1987; Orians & Wittenberger 1991; Carroll, Zielinski & Noss 1999; Collingham *et al.* 2000; Orrock *et al.* 2000; Johnson, Parker & Heard 2001). This is a manifestation of the way in which the apparent importance of such factors as different environmental variables, behavioural decisions, interspecific interactions and demographic factors change as the scale of observation does (Wiens 1989). Consequently, models employing a single study resolution provide only a limited 'snapshot' of organism-environment relationships. The generality of these relationships to other scales is unknown, and it may not be possible to find strong organism-environment relationships at the spatial scale of the sampling, creating weak predictive models (Table 1). The solution is to employ a multi-scale sampling approach wherever possible, unless there are good reasons for restricting modelling to a single scale (e.g. testing a standardised habitat survey for birds, Buckton & Ormerod 1997). In such cases caveats to the model's use regarding spatial scale should always be made clear.

Multi-scale distribution modelling is still in its relative infancy, often requiring greater quantities of data and new, more complex analytical approaches. Many studies now incorporate simple information to describe the landscape context of the sampling sites, in addition to simple habitat quality at those sites. This may include the distance

to neighbouring patches of similar habitat or site isolation (e.g. Hinsley *et al.* 1995; Lowe & Bolger 2002) or the density of similar habitat types in the area (e.g. Vander Haegen, Dobler & Pierce 2000). An alternative approach is to build models at two distinct scales, such as for quadrats of 1 m² and 100 m² for rare plant species in the Appalachians (Wiser, Peet & White 1998).

More systematic approaches to multi-scale investigations utilise scale hierarchies, incorporating four or more spatial scales (e.g. Carroll *et al.* 1999; Söderström & Pärt 2000; Apps *et al.* 2001; Mackey & Lindenmayer 2001). This provides a convenient conceptual framework for multi-scale modelling and is readily compatible with multi-scale management frameworks often employed at regional or national levels (e.g. Saunders, Hobbs & Arnold 1993; Wikramanayake *et al.* 1998; Zacharias & Roff 2000). This should maximise the relevance of models in decision-making, enhancing their overall conservation value. Sampling at several spatial scales increases the chance of identifying the scale at which distribution is most effectively predicted, increasing the potential efficacy of the model (Wiens 1989).

The choice of scales to examine in a hierarchy requires careful thought, unless it is predetermined by management objectives. Convenient scales for research, such as standard grid squares on a map, are entirely arbitrary. Assuming little prior knowledge of the species, study design may benefit from identifying an objective, logical classification of the environment into a series of distinct spatial scales, at each of which the organism-environment relationships could be analysed (Poizat & Pont 1996). Tools for spatial analysis provide objective ways of identifying the scales at which environmental patterns change, from which it is inferred that the major factors

shaping the environment change (Wiens 1989). Krummel *et al.* (1987) provided an example of this, applying fractal analysis to a fragmented woodland landscape. Regions of self-similarity across spatial scales were observed, suggesting similar controlling factors, but with a marked change when woodlands reached around 60–73 ha in area. This appeared to be the scale at which the fragmentation process was occurring and would be an obvious choice for a spatial scale in a modelling study examining the distribution of woodland birds in fragmented landscapes.

2.6 Improving analytical success

Well-planned experiments consider the subsequent data analysis during their conception, and this applies equally to distribution modelling, even though it does not conform to a rigid experimental structure. Adequate analytical power at the model-development stage is crucial to successful model development, yet has scarcely been mentioned in the ecological literature, in contrast to the medical literature (e.g. Harrell *et al.* 1984, 1996; Peduzzi *et al.* 1996). The crucial stage in model development is the selection of which environmental variables are to be retained in the final model and, to a lesser extent, the estimation of their weightings/coefficients. Although most work has concerned regression models, the recommendations should still provide useful starting points for other distribution modelling techniques.

The analytical power and reliability of regression models is largely a function of the number of sites sampled relative to the number of candidate variables considered (Peduzzi *et al.* 1996). For presence/absence models, the number of sampling sites is restricted to that fraction representing the rarer outcome event (presence or absence). As the number of observations per variable increases, so does the analytical power

and the reliability of the analysis (Steyerberg, Eijkemans & Habbema 1999). As the ratio decreases, so the risk of spurious variables being included grows, and the chance of overlooking 'important' variables increases. A range of studies has concentrated on establishing a minimum number of sampling sites per candidate variable to provide adequate statistical power. Most suggest a minimum advisable ratio of around 10:1 (e.g. Harrell *et al.* 1984; Peduzzi *et al.* 1996), although some suggest as little as half that (e.g. Freedman, Pee & Midthune 1992).

Conservation biologists face an acute challenge regarding analytical power. Sampling resources limit the overall sample size, and environmental complexity demands the use of extensive numbers of variables. A large proportion of studies has used ratios far short of 10:1, requiring further caution in their interpretation. The use of biological knowledge to select candidate predictors for recording, and during the model selection process, should help to some degree, as could the use of indirect variables, because they act as proxies for a number of other variables (Table 2). Data reduction techniques such as principal components analysis may also be valuable, especially for prediction, because they can dramatically reduce the dimensionality of the data set, and overcome collinearity between variables (Guisan *et al.* 1998; Wiser *et al.* 1998; Manel *et al.* 1999; Cumming 2000; Lowe & Bolger 2002). Their major disadvantage is in complicating interpretation.

For distribution models utilising presence/absence data, the prevalence of the organism within the data set – the proportion of sites at which it is observed – has two important effects upon the subsequent data analysis. The first concerns analytical power, which, as stated above, is limited by the frequency of the rarer outcome event

(presence or absence). Conservation biology's primary concern with rare species, whose prevalence in training data may be well below 20%, leaves it exposed to serious analytical weaknesses. The second implication of prevalence is the potential for bias problems if the prevalence observed in the training data is atypically high or low (Pearce & Ferrier 2000). Many presence/absence modelling algorithms are calibrated such that their average prediction matches the observed prevalence. If the prevalence is markedly different in subsequent applications of the model, perhaps in a different geographic region, predictions could consistently over- or underestimate the probability of occurrence (Pearce & Ferrier 2000).

There is no panacea for the problems of species prevalence, and rare organisms will continue to cause problems by necessitating the collection of large amounts of data in the hope of obtaining a reasonable level of analytical power. This may be aided by biasing sampling towards rare strata expected to support the organism or including sites already known to hold the species (e.g. traditional nesting sites for raptors, Fielding & Haworth 1995). This must be done sparingly if the prevalence is not to be inflated to the point where bias is introduced to the model. Environmental stratification divided equally between strata will itself bias sampling toward rarer strata and so may increase the prevalence. The consequences of these artificial increases in prevalence could include overoptimistic assessments of the species' status and resources wasted targeting conservation efforts towards unsuitable sites. When working with very rare or endangered species, however, these false positive (Type I, commission) errors may be considered less serious than overlooking any of the limited number of occupied sites.

Consistent biases in the model may only come to light after widespread application, across new areas and between years (due to population fluctuations). Even then, they will not affect a model that is merely required to rank sites according to their relative probability of occupancy or suitability for the species (Pearce & Ferrier 2000).

However, if absolute occupancy or suitability predictions are required, the model could be easily re-calibrated at a later date to improve its accuracy. This is perhaps to be expected if a distribution model is to be widely used as a conservation tool.

2.7 Tackling spatial autocorrelation

Awareness of spatial autocorrelation, and the need to address it in fieldwork and data analysis, has increased dramatically in recent years (Legendre 1993; Koenig 1999; Carroll & Pearson 2000). From the standpoint of field study design, before distribution data are available to characterise autocorrelation, a decision is required to either attempt to exclude autocorrelation from the training data or to use it as a candidate predictor variable. Attempting to model a distribution showing spatial autocorrelation, without taking account of it, can affect both predictive and explanatory analyses (Table 1).

Two main ideas underpin the desire to exclude autocorrelation. The first is to retain simple descriptions of organism-environment relationships at discrete sites for management purposes. Practitioners may find limited value in a model that draws much of its information from landscape context and surrounding sites – factors likely to be beyond their management. The second idea is to increase the reliability of the modelling process, because site independence is a basic requirement for valid statistical testing in model selection (Legendre 1993; Guisan & Theurillat 2000).

Positive spatial autocorrelation, for example, tends to increase the probability of spurious predictor variables being included in the model (see Legendre 1993).

Attempts to remove autocorrelation generally involve the imposition of a minimum distance between samples, in response to an assumed patchy distribution. Guisan and Theurillat (2000), for example, employed a 250-m spacing between samples in distribution modelling for alpine vegetation. Such a distance is necessarily speculative, unless distribution data are available in advance.

Attempting to filter out autocorrelation carries several disadvantages, however. It involves deliberately overlooking a potentially valuable variable for both prediction and explanation. Autocorrelation between neighbouring sites, at distances less than the rigid inter-site spacing, may still affect species occurrence at the sampling sites, perhaps through local territorial behaviour, in a way that cannot be explained with autocorrelation removed. This could create 'noise' in the data set, weakening the analysis. From a sampling viewpoint, fixing a rigid inter-site spacing may make it difficult to sample environmental space fully, especially where extensive heterogeneity exists over short distances. Finally, in terms of subsequent model application, there may be problems if there is a desire to apply the model to sets of contiguous sites.

Including spatial autocorrelation within distribution modelling provides a series of potential benefits. A greater understanding of the processes controlling distribution may be achieved, and sites are no longer considered in isolation, but in the landscape context. At local scales in particular, the accuracy of predictions may be improved

(Augustin, Muggleston & Buckland 1996; Araújo & Williams 2000). Strong autocorrelation effects may also allow more parsimonious models to be built, as the quantity of environmental information required may fall (Bivand 1984). All of these could have real benefits in conservation biology.

Field study design may be little affected by a decision to model autocorrelation should it be detected in the data set; the major differences are reserved for the model development stage. Nevertheless, it is worth considering how best to locate sampling sites with respect to future autocorrelation considerations. Primarily, the concern should be to obtain a large sample of sites over a wide range of inter-site distances so that the analytical process is powerful and can accurately quantify the distance over which autocorrelation is observed. There are no absolute rules for sample sizes, but the reliability of the analysis increases as the sample size does (Sokal & Oden 1978). The irregular location of sites within environmental stratification is advantageous in this respect, because regularly spaced sites may fall in or out of phase with environmental patterns, rendering their analytical power weak (Fortin, Drapeau & Legendre 1989). It is a straightforward piece of geometry to calculate the distances between all possible pairings of the sampling sites once the coordinates are known, and this allows a simple confirmation of good coverage across a range of inter-site distances. Where there are obvious weaknesses in the coverage, attempts should be made to adjust site positions to remedy the situation.

Once the data are collected, simple tests will indicate the presence of autocorrelation. Correlograms, showing how autocorrelation statistics such as Moran's I vary with inter-site distance, are the most valuable tools. Details of such tests are given by Sokal

and Oden (1978), Cliff and Ord (1981), and Legendre and Fortin (1989) and examples of simple autocorrelation testing on training data can be found in Beard, Hengartner & Skelly (1999), Villard, Trzcinski & Merriam (1999), Higgins *et al.* (1999), Fisher, Suarez & Case (2002) and Lobo & Martin-Piera (2002). If autocorrelation patterns are evident, the environmental variables should then be compared to see whether the patterns could be accounted for solely by autocorrelation in the environment. If so, or if no autocorrelation is evident, spatial autocorrelation need not be considered further (Smith 1994).

2.8 Collecting suitable data with which to evaluate the model

Thorough evaluation is an integral part of developing wildlife-habitat models, and consequently resources should always be allocated during fieldwork to collect data for this purpose. The correlative nature of distribution modelling demands that such evaluation with independent data be carried out to increase the scientific rigour of the work and the confidence in the conclusions (Manel *et al.* 1999; Gaston & Blackburn 1999). Only through proper evaluation can the modelling problems in Table 1 be correctly identified. In the past, the standard of model evaluation was poor, with little evaluation based on new data (Manel, Williams & Ormerod 2001), although the situation is improving with independent test data being used in an increasing number of studies. Such data, collected from a geographically discrete region, are the only valid test of a distribution model (Chatfield 1995). Prediction success with the training data has little value, while resampling techniques, such as bootstrapping and jack-knifing, may be affected by spatial autocorrelation between sites (Manel *et al.* 1999). Longer-range autocorrelation could also affect the training and test regions themselves, rendering evaluation data only pseudo-independent. This risk is slight,

however, compared with resampling approaches, where sites in close proximity can be used for training and testing. A simple autocorrelation test can confirm the independence of an evaluation set from the training data.

Aside from geographical independence, evaluation data should be collected so as to provide the model with a thorough test. Although a random selection of sites allows basic evaluation, a much more effective approach is to systematically collect sites representing the full range of environmental space, mirroring collection of the training data. This gains particular importance when it is considered that the evaluation set will generally contain fewer sites than the training set. Samples collected both inside and outside the niche identified by the distribution model are important, testing both the overall predictive accuracy and the reliability and logic of the model in widespread application. The coverage of the evaluation data can be roughly assessed relative to the training data by ordinating the training and test data using principal components analysis, and examining the ordination biplots produced. Regions of environmental space sampled in the training data, but missing in the evaluation set, become clear on simple biplots because the evaluation sites fail to completely overlap the training sites. Where gaps in coverage exist there should be caveats to model application, until such time as the model is tested in those conditions too.

2.9 Conclusions

Distribution modelling is a complex undertaking with specialised data requirements. Problems with the training data may be pervasive, contributing to a series of observed modelling problems (Table 1). By addressing these problems directly (Table 3), it is

Table 3. A summary of the major recommendations for improved field-study design and the potential benefits.

Recommendation	Potential advantages
Use of variables showing a direct relationship with the organism wherever possible	<ul style="list-style-type: none">• Improved predictive ability, especially over large geographical extents or in predicting the responses to environmental change• Greater ease of interpretation
Consideration of interacting species	<ul style="list-style-type: none">• Improved predictive ability• Greater biological validity (modelling of realised niche)• Greater explanatory power and ease of interpretation
Identification of complete geographical region of interest, prior to sampling	<ul style="list-style-type: none">• Improved predictive ability with new data – model does not need to extrapolate to new conditions• Explanatory conclusions more widely applicable
Environmental stratification, with an equal division of samples between strata	<ul style="list-style-type: none">• Improved predictive ability• More accurate explanatory analysis
Multi-scale approach to sampling	<ul style="list-style-type: none">• Improved predictive ability• Greater explanatory understanding• Greater relevance to conservation planning

Aim to sample at least 10 sites for every environmental variable considered

- More reliable model development and explanatory analysis
- Improved predictive ability

Greater awareness of the effects of prevalence in presence/absence modelling

Aim to model spatial autocorrelation, where present; test to ensure adequate statistical power for autocorrelation analyses in the design of sampling scheme

- Facilitate detection, characterisation and subsequent modelling of autocorrelation
- Improved understanding of mechanisms generating distribution pattern
- Greater predictive accuracy

Partition resources to collect independent evaluation data;
environmental stratification used in the collection processes

- Essential to test distribution models, thereby increasing the scientific rigour of observational analyses
 - Essential to assess model generality and predictive ability
 - Environmental stratification ensures evaluation data will provide a thorough test
-

hoped that valuable improvements in the quality of training data can be achieved. Ultimately, improved predictive ability and better explanatory analysis should result, and greater confidence could be placed in both. Given the burgeoning use of distribution modelling in conservation biology, any such improvements stand to be very valuable.

2.10 Acknowledgements

I thank three anonymous reviewers and the editors of *Conservation Biology* for their helpful and insightful comments on earlier versions of the manuscript.

2.11 References

- Apps, C.D., McLellan, B.N., Kinley, T.A. & Flaa, J.P. (2001) Scale-dependent habitat selection by mountain caribou, Columbia Mountains, British Columbia. *Journal of Wildlife Management*, **65**, 65–77.
- Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.
- Augustin, N.H., Muggleston, M.A. & Buckland, S.T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339–347.
- Austin, M.P. (1980) Searching for a model for use in vegetation analysis. *Vegetatio*, **42**, 11–21.
- Austin, M.P., Cunningham, R.B. & Fleming, P.M. (1984) New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio*, **55**, 11–27.
- Austin, M.P. & Heyligers, P.C. (1989) Vegetation survey design for conservation: gradsect sampling of forests in north-eastern New South Wales. *Biological Conservation*, **50**, 13–32.
- Austin, M.P. & Meyers, J.A. (1996) Current approaches to modelling the environmental niche of eucalypts: implications for management of forest biodiversity. *Forest Ecology and Management*, **85**, 95–106.

Austin, M.P., Nicholls, A.O. & Margules, C.R. (1990) Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs*, **60**, 161–177.

Beard, K.H, Hengartner, N. & Skelly, D.K. (1999) Effectiveness of predicting breeding bird distributions using probabilistic models. *Conservation Biology*, **13**, 1108–1116.

Begon, M., Harper, J.L. & Townsend, C.R. (1996) *Ecology: individuals, populations and communities* (3rd edn.). Blackwell Science, Oxford.

Bivand, R.S. (1984) Regression modelling with spatial dependence: an application of some class selection and estimation methods. *Geographical Analysis*, **16**, 25–37.

Buckton, S.T. & Ormerod, S.J. (1997) Use of a new standardized habitat survey for assessing the habitat preferences and distribution of upland river birds. *Bird Study*, **44**, 327–337.

Carroll, S.S. & Pearson, D.L. (2000) Detecting and modelling spatial and temporal dependence in conservation biology. *Conservation Biology*, **14**, 1893–1897.

Carroll, C., Zielinski, W.J. & Noss, R.F. (1999) Using presence/absence data to build and test spatial habitat models for the Fisher in the Klamath Region, U.S.A. *Conservation Biology*, **13**, 1344–1359.

Chatfield, C. (1995) Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A.*, **158**, 419–466.

Clevenger, A.P., Wierzchowski, J., Chruszcz, B. & Gunson, K. (2002) GIS-generated, expert-based models for identifying habitat linkages and planning mitigation passages. *Conservation Biology*, **16**, 503–514.

Cliff, A.D. & Ord, J.K. (1981) *Spatial processes: models and applications*. Pion, London.

Collingham, Y.C., Wadsworth, R.A., Huntley, B. & Hulme, P.E. (2000) Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology*, **37**(suppl. 1), 13–27.

Cumming, G.S. (2000) Using between-model comparisons to fine-tune linear models of species range. *Journal of Biogeography*, **27**, 441–455.

Fielding, A.H. & Haworth, P.F. (1995) Testing the generality of bird-habitat models. *Conservation Biology*, **9**, 1466–1481.

Fisher, R.N., Suarez, A.V. & Case, T.J. (2002) Spatial patterns in the abundance of the coastal horned lizard. *Conservation Biology*, **16**, 205–215.

Fortin, M.-J., Drapeau, P. & Legendre, P. (1989) Spatial autocorrelation and sampling design in plant ecology. *Vegetatio*, **83**, 209–222.

Franco, A.M.A., Brito, J.C. & Almeida, J. (2000) Modelling habitat selection of Common Cranes *Grus grus* wintering in Portugal using multiple logistic regression. *Ibis*, **142**, 351–358.

Franklin, J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474–499.

Freedman, L.S., Pee, D. & Midthune, D.N. (1992) The problem of underestimating the residual error variance in forward stepwise regression. *The Statistician*, **41**, 405–412.

Gainzarain, J.A., Arambarria, R. & Rodríguez, A.F. (2000) Breeding density, habitat selection and reproductive rates of the Peregrine Falcon *Falco peregrinus* in Álava (northern Spain). *Bird Study*, **47**, 225–231.

Gaston, K.J. & Blackburn, T.M. (1999) A critique for macroecology. *OIKOS*, **84**, 353–368.

Gillison, A.N. & Brewer, K.R.W. (1985) The use of gradient directed transects or gradsects in natural resource surveys. *Journal of Environmental Management*, **20**, 103–127.

Guisan, A. & Theurillat, J.-P. (2000) Assessing alpine plant vulnerability to climate change: a modelling perspective. *Integrated Assessment*, **1**, 307–320.

Guisan, A., Theurillat, J.-P. & Kienast, F. (1998) Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science*, **9**, 65–74.

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

Harrell, F.E., Lee, K.L., Califf, R.M., Pryor, D.B. & Rosati, R.A. (1984) Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, **3**, 143–152.

Harrell, F.E., Lee, K.L. & Mark, D.B. (1996) Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, **15**, 361–387.

Higgins, S.I., Richardson, D.M., Cowling, R.M. & Trinder-Smith, T.H. (1999) Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. *Conservation Biology*, **13**, 303–313.

Hinsley, S.A., Bellamy, P.E., Newton, I. & Sparks, T.H. (1995) Habitat and landscape factors influencing the presence of individual breeding bird species in woodland fragments. *Journal of Avian Biology*, **26**, 94–104.

Hirzel, A. & Guisan, A. (2002) Which is the optimal strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331–341.

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, **22**, 415–427.

Johnson, C.J., Parker, K.L. & Heard, D.C. (2001) Foraging across a variable landscape: behavioural decisions made by woodland caribou at multiple spatial scales. *Oecologia*, **127**, 590–602.

Koenig, W.D. (1999) Spatial autocorrelation of ecological phenomena. *Trends in Ecology and Evolution*, **14**, 22–26.

Krummel, J.R., Gardner, R.H., Sugihara, G., O'Neill, R.V. & Coleman, P.R. (1987) Landscape patterns in a disturbed environment. *OIKOS*, **48**, 321–324.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659–1673.

Legendre, P. & Fortin, M.-J. (1989) Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107–138.

Lobo, J.M. & Martin-Piera, F. (2002) Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables.

Conservation Biology, **16**, 158–173.

Lowe, W.H. & Bolger, D.T. (2002) Local and landscape-scale predictors of salamander abundance in New Hampshire headwater streams. *Conservation Biology*,

16, 183–193.

Mackey, B.G. & Lindenmayer, D.B. (2001) Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography*, **28**, 1147–

1166.

Mac Nally, R. (2000) Regression and model-building in conservation biology,

biogeography and ecology: the distinction between – and reconciliation of –

‘predictive’ and ‘explanatory’ models. *Biodiversity and Conservation*, **9**, 655–671.

Manel, S., Buckton, S.T. & Ormerod, S.J. (2000) Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan

ivers. *Journal of Applied Ecology*, **37**, 756–770.

Manel, S., Dias, J.M., Buckton, S.T & Ormerod, S.J. (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of*

Applied Ecology, **36**, 734–747.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence/absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921–931.

Margules, C.R. & Austin, M.P. (1994) Biological models for monitoring species decline: the construction and use of data bases. *Philosophical Transactions of the Royal Society of London B*, **344**, 69–75.

Margules, C.R. & Stein, J.L. (1989) Patterns in the distributions of species and the selection of nature reserves: an example from the *Eucalyptus* forests in south-eastern New South Wales. *Biological Conservation*, **50**, 219–238.

Mladenoff, D.J., Sickley, T.A. & Wydeven, A.P. (1999) Predicting gray wolf landscape recolonization: logistic regression models vs. new field data. *Ecological Applications*, **9**, 37–44.

Mohler, C.L. (1983) Effect of sampling pattern on estimation of species distributions along gradients. *Vegetatio*, **54**, 97–102.

Neave, H.M., Norton, T.W. & Nix, H.A. (1996) Biological inventory for conservation evaluation I. Design of a field-survey for diurnal, terrestrial birds in southern Australia. *Forest Ecology and Management*, **85**, 107–122.

Nugent, G., Fraser, W. & Sweetapple, P. (1992) Top down or bottom up? Comparing the impacts of introduced arboreal possums and 'terrestrial' ruminants on native forests in New Zealand. *Biological Conservation*, **99**, 65–79.

Orians, G.H. & Wittenberger, J.F. (1991) Spatial and temporal scales in habitat selection. *The American Naturalist*, **137**(suppl.), 29–49.

Orrock, J.L., Pagels, J.F., McShea, W.J. & Harper, E.K. (2000) Predicting presence and abundance of a small mammal species: the effect of scale and resolution. *Ecological Applications*, **10**, 1356–1366.

Özesmi, S.L. & Özesmi, U. (1999) An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling*, **116**, 15–31.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. & Feinstein, A.R. (1996) A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, **49**, 1373–1379.

Poizat, G. & Pont, D. (1996) Multi-scale approach to species-habitat relationships: juvenile fish in a large river. *Freshwater Biology*, **36**, 611–622.

Ripley, B.D. (1994) Neural networks and related methods for classification. *Journal of the Royal Statistical Society, Series B.*, **56**, 409–456.

Robertson, A.W., Kelly, W., Ladley, J.J. & Sparrow, A.D. (1999) Effects of pollinator loss on endemic New Zealand mistletoes (Loranthaceae). *Conservation Biology*, **13**, 499–508.

Rushton, S.P., Barreto, G.W., Cormack, R.M., Macdonald, D.W. & Fuller, R. (2000) Modelling the effects of mink and habitat fragmentation on the water vole. *Journal of Applied Ecology*, **37**, 475–490.

Saunders, D.A., Hobbs, R.J. & Arnold, G.W. (1993) The Kellerberrin project on fragmented landscapes: a review of current information. *Biological Conservation*, **64**, 185–192.

Smith, P.A. (1994) Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters*, **4**, 47–61.

Söderström, B. & Pärt, T. (2000) Influence of landscape scale on farmland birds breeding in semi-natural pastures. *Conservation Biology*, **14**, 522–533.

Sokal, R.R. & Oden, N.L. (1978) Spatial autocorrelation in biology 1. Methodology. *Biological Journal of the Linnean Society*, **10**, 199–228.

Steyerberg, E.W., Eijkemans, M.J.C. & Habbema, J.D.F. (1999) Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology*, **52**, 935–942.

Vander Haegen, W.M., Dobler, F.C. & Pierce, D.J. (2000) Shrubsteppe bird response to habitat and landscape variables in eastern Washington, USA. *Conservation Biology*, **14**, 1145–1160.

Veblen, T.T., Mermoz, M., Martin, C. & Kitzberger, T. (1992) Ecological impacts of introduced animals in Nahual-Huapi-National-Park, Argentina. *Conservation Biology*, **6**, 71–83.

Villard, M.A., Trzcinski, M.K. & Merriam, G. (1999) Fragmentation effects on forest birds: Relative influence of woodland cover and configuration on landscape occupancy. *Conservation Biology*, **13**, 774–783.

Wessels, K.J., Van Jaarsveld, A.S., Grimbeek, J.D. & Van Der Linde, M.J. (1998) An evaluation of the gradsect biological survey method. *Biodiversity and Conservation*, **7**, 1093–1121.

Westman, W.E. (1991) Measuring realized niche spaces: climatic response of chaparral and coastal sage scrub. *Ecology*, **72**, 1678–1684.

Wiens, J.A. (1989) Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.

Wiens, J.A., Rotenberry, J.T. & Van Horne, B. (1987) Habitat occupancy patterns of North American shrubsteppe birds: the effects of spatial scale. *OIKOS*, **48**, 132–147.

Wikramanayake E.D., Dinerstein, E., Robinson, J.G., Karanth, U., Rabinowitz, A., Olson, D., Mathew, T., Hedao, P., Conner, M., Hemley, G. & Bolze, D. (1998) An ecology-based method for defining priorities for large mammal conservation: The tiger as a case study. *Conservation Biology*, **12**, 865–878.

Wiser, S.K., Peet, R.K. & White, P.S. (1998) Prediction of rare-plant occurrence: a southern Appalachian example. *Ecological Application*, **8**, 909–920.

Woolf, A., Nielson, C.K., Weber, T. & Gibbs-Kieninger, T.J. (2002) Statewide modelling of bobcat, *Lynx rufus*, habitat in Illinois, USA. *Biological Conservation*, **104**, 191–198.

Yee, T.W. & Mitchell, N.D. (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science*, **2**, 587–602.

Zacharias, M.A. & Roff, J.C. (2000) A hierarchical approach to conserving marine biodiversity. *Conservation Biology*, **14**, 1327–1334.

Chapter 3

Testing the predictive performance of species distribution models: unresolved issues and omissions from current practice

3.1 Summary

1. Species distribution models could bring manifold benefits to ecology and conservation biology. However, every model should be carefully tested before use, to ensure that valid conclusions are drawn and that conservation receives sound guidance. Here, I review the methods used to assess models developed from presence/absence data. Three concerns are highlighted i) poor characterisation of model generalisability, ii) the need for accuracy measures that allow inter-model comparisons, and iii) failure to examine the calibration of predictions – a fundamental part of prediction accuracy.
2. The assessment of generalisability requires a combination of resampling (e.g. bootstrapping) to examine overfitting and an independent test set – representing the range of conditions under which a model may be applied – to assess transportability. Within a test set, variation in accuracy with environmental conditions (e.g. the range of the predictors) or other factors (e.g. the effects of different users) should be described.
3. Predictions produced by most distribution models are continuous variables. Discrimination and calibration components of accuracy can be readily distinguished and tested separately, greatly benefiting model evaluation. Discrimination is best assessed using rank correlation coefficients, such as the concordance-index. Calibration is best described using smoothed calibration plots, although a logistic regression method provides a basis for characterising and correcting calibration problems in a linear context.
4. The problems of testing nominal presence/absence predictions are highlighted: especially when derived by dichotomising a continuous output. Chance-corrected measures of accuracy, such as Cohen's κ , may only be appropriate if prevalence is

held constant across modelling studies. Mutual information measures, based in information theory, may be more appropriate.

5. A case study, predicting the distribution of a Himalayan river bird, is used to demonstrate some recommended methods for assessing the discrimination, calibration and generalisability of models producing probabilistic predictions.
6. Model evaluation should be made more informative for model developers and users alike. It is imperative to: i) define generalisability in the greatest possible detail, ii) separate discrimination and calibration, testing both, iii) use flexible, general assessment techniques such as rank correlation coefficients and non-parametric calibration plots, iv) avoid nominal presence/absence evaluation where possible and consider information-theoretic measures, and v) utilise the full range of techniques to help diagnose causes of prediction problems. Currently, few modellers in conservation biology satisfy these needs, making it difficult for others to evaluate their models and leaving models open to potentially unrecognised misuses that could damage conservation efforts.

Key words: accuracy; bootstrapping; calibration; discrimination; generalisability; overfitting; transportability

3.2 Introduction

Species distribution models are being proposed with increasing frequency throughout ecology and conservation biology. Whilst some focus on increasing understanding of species' habitat requirements, the majority are being developed primarily as applied predictive tools (Rushton, Ormerod & Kerby 2004). The potential benefits are manifold (see Table 1 in Manel, Williams & Ormerod 2001), and as a consequence, distribution modelling studies are frequent in applied ecological journals. For example, over the five years 1999–2003, around 10% of the papers published in the *Journal of Applied Ecology* and 7% in *Ecological Applications* involved distribution modelling. Many papers focus upon methodological aspects of such modelling, and a recurrent theme in recent years has been how best to test models developed from presence/absence data (e.g. Fielding & Bell 1997; Fielding 1999, 2002; Pearce & Ferrier 2000; Manel *et al.* 2001; Boyce *et al.* 2002). Despite this attention, important shortcomings remain common in ecology.

Thorough testing is an important pre-requisite to all predictive applications of species distribution models. Validation reveals how models should perform in practice, establishes when and where they produce accurate predictions, and provides essential user-guidance. By analysing test results, problems in model specification can be diagnosed and corrected (Miller, Hui & Tierney 1991; Pearce & Ferrier 2000).

Typically, testing involves a direct comparison of a sample of a model's predictions against the observed species distribution. In this way, the efficacy of alternative models can be compared, irrespective of the modelling algorithm (e.g. regression versus machine learning methods: Fielding 1999; Manel *et al.* 1999a, 1999b; Olden & Jackson 2002). Finally, testing provides an opportunity to convince ecologists and end

users of a model's value by demonstrating its efficacy (Snee 1977; Wyatt & Altman 1995; Pearce, Ferrier & Scotts 2001).

The overall accuracy of a distribution model may be sub-divided into discriminatory ability and calibration, especially when predictions are made on a continuous scale (e.g. probabilities of occurrence). Discriminatory ability is a model's capacity to distinguish occupied from unoccupied sites (Harrell *et al.* 1984). Where continuous predictions are made this is manifested as the ability to place sites in rank order, assigning higher probabilities or habitat suitabilities to sites that are more likely to be occupied. Calibration describes the numerical accuracy of the predictions: whether the predicted prevalence of a species matches its observed prevalence, either across a complete data set or for subsets of sites (Harrell *et al.* 1984). For probabilistic predictions, calibration is concerned with whether, for example, a site given a predicted probability of 0.60 has a 60% chance of being occupied and whether this is twice as likely as for a site given a label of 0.30. If the predictions are not scaled between zero and one, an analogous property could be considered: whether the habitat suitability scores are proportional to prevalence, for example. The extent to which a model can make accurate predictions in new applications is its generalisability (Justice, Covinsky & Berlin 1999). To generalise to a new set of data, a model needs to minimise overfitting during training and in addition be transportable to the different conditions found within the new application, such as a different geographical region or range of environmental conditions (Justice *et al.* 1999). Overall predictive performance, the assessment of which is the aim of the testing process, is a combination of both accuracy and generalisability.

The assessment of overall predictive performance is often limited by weaknesses in three areas of model evaluation and these form the focus of this discussion. The first is the approach commonly adopted to assessing generalisability where, assuming a dedicated test set is available, only the average accuracy is reported. This provides minimal information to a model's users. The second weakness in model evaluation concerns the statistics used to describe the accuracy of predictions. Poor selection of accuracy measures makes it difficult to compare models in a meaningful way: a problem that is particularly acute for models making nominal presence/absence predictions. Finally, the third weakness in model evaluation is the virtual absence from ecology of tests of model calibration.

In the current paper, each of these three issues is discussed in turn and some potential solutions illustrated in a brief case study. I then outline a protocol for evaluating either probabilistic or nominal species distribution models.

3.3 Testing generalisability

All predictive species distribution models require some degree of generalisability since their basic purpose is to predict distribution in circumstances where direct field observation is difficult or impossible. Tests of generalisability are therefore fundamental to evaluation, and should have three basic aims. The first is to examine a model's generalisability to its intended applications. This requires consideration of when and where a model could be used and the subsequent collection of suitable test data. When such a test is carried out, accuracy will almost always be lower than when a model predicts the data used to train it (so-called resubstitution testing; Fielding & Bell 1997). It follows that the second aim of testing a model's generalisability is to



partition such losses in performance between its two possible sources: overfitting of the training data and an inability to transport to the conditions of the test set, and therefore to future applications.

The final aim is to describe the performance in independent test data in more detail, relating prediction accuracy to the conditions of application. The purpose of this is twofold. If a model has limited transportability, further analysis helps to diagnose the causes. For example, model accuracy may be poor over a particular range of the predictors or a model may not transport to sites within a certain geographical location. The second benefit of assessing transportability in this way is that it provides the detailed information about the limits to model applicability that subsequent users require.

3.3.1 THE RELATIONSHIP BETWEEN RESAMPLE TESTING AND INDEPENDENT TEST DATA

Independent data provide the only test of a model's overall generalisability, reflecting the performance expected in future applications (Chatfield 1995; Fielding 1999; Justice *et al.* 1999; Altman & Royston 2000). However, independent test data alone cannot identify the causes of reduced accuracy relative to resubstitution. This is because the effects of overfitting the training data – modelling its idiosyncrasies in addition to actual species-environment relationships – cannot be distinguished from the inability of a model to transport to new conditions (Justice *et al.* 1999). The solution is to use a resampling method (e.g. cross-validation) alongside independent test data. Being based solely upon the training data, resampling methods cannot estimate transportability to new conditions, but can produce near-unbiased estimates of performance under the model development conditions (Efron 1983; Verbyla &

Litvaitis 1989). The difference between such an estimate and that obtained from resubstitution is an estimate of the degree of overfitting present (Harrell *et al.* 1996). If a decrease in accuracy from resubstitution to independent test data exceeds that accounted for by overfitting, transportability problems are implied (Justice *et al.* 1999). Resampling and independent test data are therefore complementary to one another.

Resampling is a convenient first stage in model evaluation. Overfitting frequently underlies limits to generalisability and so resample testing can often identify problems prior to the expense of collecting specific test data (Charlson *et al.* 1987; Harrell *et al.* 1996). Equally, good overfitting-corrected performance may provide the necessary justification to further pursue model development, including the expense of test data collection (Verbyla & Litvaitis 1989).

Of the range of resampling methods available (Verbyla & Litvaitis 1989), jack-knifing and bootstrapping are generally the most useful. By averaging accuracy statistics across many iterations, they make more stable estimates of performance in small to moderately-sized data sets than those obtained by a single split of a data set into training and test fractions (Efron 1983; Steyerberg *et al.* 2001). They also allow all of the data to be used to fit the model, improving its quality (Hirsch 1991). At the same time, jack-knifing and bootstrapping can produce near-unbiased estimates of performance (Efron & Gong 1983; Verbyla & Litvaitis 1989; Steyerberg *et al.* 2001).

Bootstrapping is potentially more useful than jack-knifing for distribution modelling on pragmatic grounds. It produces large ‘training’ and ‘test’ sets in each iteration,

Table 1. A simple method for using bootstrapping to estimate, and correct for, overfitting in accuracy statistics. After: Efron (1983); Harrell *et al.* (1996); Steyerberg *et al.* (2001).

-
1. Obtain resubstitution estimates of accuracy statistic (accuracy in the training data).
 2. Generate a bootstrap of equal size to the training set by sampling training data with replacement.
 3. Fit the model in the bootstrap using the same methods as employed to fit it in the original training data: this includes the same variable selection strategy, where applicable.
 4. Estimate the accuracy statistic within the bootstrap resample. This simulates resubstitution testing.
 5. Using the same model as in (4), predict species distribution in the original training set and estimate the accuracy statistic. This simulates the use of independent test data.
 6. Overfitting = ('resubstitution estimate' in step 4) – ('test data estimate' in step 5).
 7. Repeat steps 2–6 for 100–200 bootstraps. Average the values calculated in step 6 to provide the overall estimate of overfitting.
 8. Subtract bias estimate from the resubstitution estimate in step 1 to provide an optimism-corrected value.
-

making it better suited to calculating many accuracy measures. Several realisations of the bootstrap are available, perhaps the most advanced being the ‘.632+’ method, which corrects for the slight bias that occurs due to the partial overlap between ‘training’ and ‘test’ data (Efron & Tibshirani 1997). Steyerberg *et al.* (2001), however, found no advantage in using this method over simpler bootstrapping approaches for estimating discrimination and calibration. They recommend an alternative approach which, rather than attempting to provide unbiased performance estimates in each bootstrap, estimates the degree of overfitting in model development, which can then be subtracted from the accuracy measured by resubstitution (Table 1).

3.3.2 COLLECTION OF TEST DATA

Independent data are employed in many distribution modelling studies and yet there is relatively little work on the important properties of test data (Altman & Royston 2000). The major concerns are obtaining full, representative coverage of future model applications and a sufficient sample size to estimate accuracy precisely.

Prior to the collection of test data, possible model applications should be defined, along with their differences from the training data (Table 2). This identifies the range of conditions over which generalisability is desired and provides a basis for the collection of test data. For conditions under which a model cannot be tested, the closest possible analogue should be the aim, such as some other form of temporal difference to validate a model developed to predict future distribution patterns. The spatial patterns in prediction errors may also be important, as they can influence the interpretation of model accuracy (Fielding & Bell 1997). With this in mind, test data

Table 2. Possible differences between training data and potential applications for distribution models that should be addressed in the selection of test data. Partly adapted from Justice *et al.* (1999).

Temporal separation

- Predictions of historical or future distributions.
- May be short term, such as the subsequent breeding season (e.g. Boyce *et al.* 2002), or longer term, such as species' responses under alternative climate change scenarios (e.g. Berry *et al.* 2002).

Geographical separation

- Models may be required to extrapolate to new regions, such as different islands (e.g. Fielding & Haworth 1995), or simply to interpolate between the sites included in the training data (e.g. Suárez-Seoane *et al.* 2002).

Different regions of environmental space

- Often related to differences in geographical space.
 - Model applications should ideally be within the range of environmental conditions sampled in the training data. Often, however, extrapolation is required, perhaps to positions further along environmental gradients.
-

'Deployment factors'

- Models intended for frequent use as conservation support tools may have to generalise to a range of new conditions. Many different people may use a model (e.g. the RIVPACS system, Wright 1995), for example, or data may be collected using different survey techniques (e.g. Pearce *et al.* 2001).
- The ability of a model to generalise over such variations in deployment needs to be tested.

Prevalence

- The prevalence of a species may differ in future applications.
 - The continued efficacy of a model needs to be confirmed or corrective action proposed: does the model need re-calibrating to the new prevalence or does the different prevalence represent a more fundamental change in a species' habitat use, weakening discriminatory ability (Poses *et al.* 1986)?
-

should attempt to match the expected spatial pattern of model applications, so that such errors may be described. If a model will be used to generate complete coverage of habitat suitability across an area, for example, tests made from a widely spaced set of sites may be relatively uninformative.

The only plausible alternative to independent test data is data partitioning stratified by such factors as distinct geographical regions. Whilst this represents a systematic approach to examining generalisability, such a test set is unlikely to address the full range of ways in which future applications will differ from the training data (Table 2). For example, if all of the data were collected in a single field season, they cannot measure transportability to subsequent years. The influence of such factors upon prediction accuracy is unlikely to be known *a priori* and so in most circumstances generalisability can only be reliably assessed using independent data that simulate real applications.

Test-set size presents problems in ecology. Many measures used to assess the accuracy of distribution models are non-parametric or, in the case of nominal presence/absence predictions, individual test sites convey relatively little information. Consequently, large sample sizes are required to obtain precise accuracy estimates and to test for differences between models. Differences in the area under the ROC (receiver operating characteristic) curve (AUC) statistic – now widely used in ecology – of 0.05 may reflect major differences in predictive ability, yet several hundred sites may be required to identify this statistically (Ash & Schwartz 1999; Cumming 2000; Steyerberg *et al.* 2003). Similar sample sizes may be required to separate kappa statistics (Donner 1998). For both, the precision of estimates increases with the

overall sample size, as species prevalence approaches 50% and as the discriminatory ability of the model increases (Hanley & McNeil 1982; Donner 1998). Hanley & McNeil's calculations for the standard error around estimates of AUC indicate that changes in any of these factors can make several-fold differences, especially at sample sizes below around 200 (McPherson *et al.* in press).

General sample size guidelines are difficult to formulate, as precision will vary between studies according to the basic properties of the species and the model. As a minimum, perhaps 200–300 sites should be the aim for a test set: substantially more if a heterogeneous set of applications is eventually intended (Table 2). Harrell (2001) suggests that a test set should contain at least 100 sites of the less common event (present/absent).

3.3.3 DESCRIBING PREDICTIVE PERFORMANCE IN INDEPENDENT DATA

Details of how model accuracy varies according to different conditions in the test data may provide invaluable information to model developers and users. Summarising the discrimination and calibration of the model under different operating conditions could be valuable, estimating the likely accuracy for distinct applications (Table 2).

Graphical analyses, plotting the accuracy of predictions for individual sites against factors hypothesised to affect them, may be used to identify conditions under which accuracy is degraded (Miller *et al.* 1991; Mackillop & Quirt 1997). Of particular interest are groups of sites for which the model appears inadequate that relate to specific ranges of the predictors, allowing transportability limits to be described in the terms most relevant to a model's users. This is demonstrated in the case study.

3.4 Selection of accuracy statistics 1: probabilistic/continuous-scale predictions

Several publications have considered the important characteristics of accuracy measures for distribution models and analogous systems (Swets 1988; Forbes 1995; Fielding & Bell 1997; Fielding 1999, 2002). Perhaps the single most important property – and the greatest challenge – is to provide meaningful comparisons of accuracy between models developed for different applications: where different species are of interest and models applied under different conditions (Manel *et al* 2001; McPherson *et al.*, in press). To achieve this, accuracy statistics must either be largely independent of, or attempt to correct for, potentially confounding properties of the particular data used to test a model, such as species' prevalence (Swets 1988; Miller *et al.* 1991; Fielding 1999, 2002; Manel *et al.* 2001). This rules out the use of goodness-of-fit measures such as R^2 for model evaluation (c.f. Ash & Schwartz 1999) – except for comparing alternative models for the same species in the same data set – because they are heavily reliant upon such factors as species' prevalence (Miller *et al.* 1991; Cox & Wermuth 1992).

3.4.1 MEASURES OF DISCRIMINATORY ABILITY FOR PROBABILISTIC MODELS

Many distribution modelling methods produce continuous predictions, at least as their initial output prior to dichotomisation either by software or the user (Fielding 2002). This includes logistic regression, Bayesian approaches, regression trees, discriminant analyses and artificial neural networks (Manel *et al.* 1999a; Guisan & Zimmermann 2000). The accuracy of models making such predictions can be readily separated into discrimination and calibration components, and this permits a much more general assessment of model performance than would otherwise be possible (c.f. nominal

presence/absence predictions, below). Discriminatory ability is the fundamental component of predictive ability – a model's 'ecological skill' – and consequently is the primary concern when testing and comparing models (Yates 1982; Harrell *et al.* 1984). It is independent of species prevalence as it simply addresses the capacity to rank sites according to likelihood of occurrence/habitat suitability and distinguish differences between sites, rather than considering the accuracy of the actual probability labels (Harrell *et al.* 1984; Ash & Schwartz 1999). By separating discrimination from calibration, therefore, and obtaining as pure a measure of it as possible, discrimination is readily comparable between studies.

Non-parametric correlation coefficients, relating predicted probabilities to observed presence/absence, are an effective approach to assessing discrimination. They are a pure measure of discrimination, as they only consider rank order of predictions, rather than incorporating information about the actual discrepancies between individual predictions and observed presence/absence (Harrell *et al.* 1996). This information, which is an aspect of calibration, is incorporated alongside discrimination into parametric correlation coefficients and goodness-of-fit measures, confounding comparative studies (Miller *et al.* 1991). For example, consider two possible test data sets for a model. In the first, the majority of sites are of low to moderate suitability for a species, and so a well-calibrated discriminating model will tend to concentrate predictions within a limited range of the unit probability scale. By contrast, in the second test set, sites are polarised between highly suitable and highly unsuitable. Accurate predictions from the model would then be concentrated at the extremes. In both cases the model may be just as discriminating, correctly placing sites in rank order of suitability, yet the use of a parametric correlation coefficient would suggest

much better discrimination in the latter due to smaller discrepancies between the probabilities and observed presence/absence. With such parametric measures, the choice of data set used to test a model may have major effects upon the subsequent results (Miller *et al.* 1991).

A range of rank correlation coefficients is available that could be used to assess discriminatory ability, including Somers' D_{XY} , Goodman and Kruskal's γ , and Kendall's τ (Harrell *et al.* 1996; SAS Institute 1999). The most useful for presence/absence modelling in ecology is probably the concordance index (c -index) developed by Harrell *et al.* (1982), because it is equivalent to the non-parametric AUC measure and the Wilcoxon statistic (Bamber 1975; Hanley & McNeil 1982). This allows direct comparisons to studies where continuous predictions may need to be dichotomised and so where ROC methodology is used to examine how classification performance varies over the range of possible classification thresholds (0–1).

Aside from its independence of prevalence and simplicity of computation, c is highly intuitive. It indicates the probability that if two sites (one occupied, the other unoccupied) were selected at random, the model would place them in the correct rank order of likelihood of occupancy (Harrell *et al.* 1982). This means that chance/random performance has a clear definition, 0.5, and perfect discrimination is unity.

The major weakness of non-parametric measures of discrimination is their inefficient use of data, necessitating large test sample sizes as discussed above. A particular problem arises when the prevalence is very low ($\ll 0.5$), as is often the case in conservation biology due to the frequent interest in rare species. Only paired

occupied/unoccupied sites are used to calculate c (AUC), with the result that large increases in the size of a test set may be required to usefully increase the effective sample size for calculating the statistic. The strength of this effect is demonstrated by the large impact that small changes in prevalence have near the extremes for a fixed sample size (McPherson *et al.* in press).

3.4.2 MEASURES OF CALIBRATION FOR PROBABILISTIC MODELS

Calibration, in contrast to discrimination, is directly concerned with species prevalence, describing how well predicted prevalence (mean probability) matches that observed in the test data (Harrell *et al.* 1984). It is a vital component of prediction accuracy, yet is rarely tested in ecology (e.g. Carroll, Zielinski & Noss 1999; Pearce & Ferrier 2000; Pearce *et al.* 2002; Meggs *et al.* 2004). Calibration at the level of a complete test set reflects the representativeness of the species' prevalence in the training data – to which the model is calibrated (Pearce & Ferrier 2000) – for subsequent applications (i.e. test data). Calibration accuracy is easily understood and compared between applications, and unlike discrimination may be easily corrected, so as to reflect a species' prevalence in a new application (Harrell *et al.* 1984; Poses *et al.* 1986).

The pattern of calibration across the unit probability scale should be examined in addition to the overall average. It largely reflects aspects of model development, such as overfitting (Harrell *et al.* 1996; Pearce & Ferrier 2000). In a model that is well-calibrated across the unit scale, predicted probabilities can be interpreted directly as likelihoods of occurrence, conveying much more information than nominal predictions. Decision making for conservation should benefit from the indications of

confidence in predicted species occurrences that probabilities provide and the ability to make *quantitative* comparisons of site suitability for, or use by, a species, compared to simple ranking/ordinal comparisons if a model only has good discriminatory ability. By providing well-calibrated probabilistic predictions to decision makers, the same data may be re-interpreted – as conditions dictate – using different thresholds to reflect the perceived costs of different prediction errors (Yates *et al.* 1996).

Unfortunately, it is common that only tests of discriminatory ability are indicated, before probabilities are displayed, often as maps of potential habitat use/suitability (e.g. Pereira & Itami 1991; Aspinall 1992; Franco, Brito & Almeida 2000; Suárez-Seoane *et al.* 2002; Johnson, Seip & Boyce 2004; Venier *et al.* 2004). Whilst it can be argued that for certain applications, such as prioritising conservation resources, only the ability to place sites in rank order of suitability is of interest (e.g. Pearce *et al.* 2001), it is dangerous to display such probabilities for interpretation with no knowledge of their calibration. Large apparent differences between sites based upon probabilities of occupancy may in fact be spurious or equally attempts to select relatively homogenous groups of sites for an application may be confounded by including sites that could in fact be markedly different.

The calibration plot is the basic tool for assessing the calibration of probabilistic predictions. At its simplest, a plot of average predicted probability (*x*-axis) versus species prevalence (*y*-axis) can be drawn for discrete groups of sites across the unit probability scale (Fig. 1). Groups may be formed at regular probability intervals (e.g. deciles) or for fixed numbers of sites across the probability range. Problems with

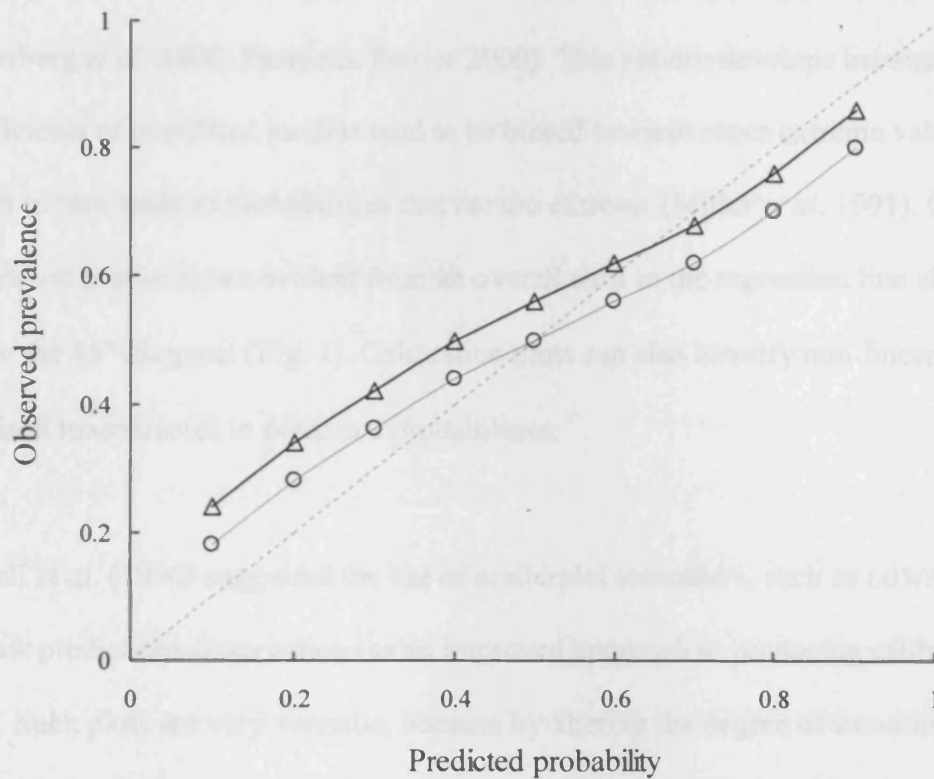


Figure 1. Sample calibration plot, comparing predicted and observed prevalence at decile probability intervals. The 45° diagonal (fine dashed line) indicates perfect calibration. The other two lines represent calibration problems: overfitting (circular symbols) and overfitting in a model with overall calibration problems, underestimating prevalence (triangular symbols).

calibration are evident as deviations in the agreement between predictions and observations from the 45° diagonal (Fig. 1). A pattern of calibration with a slope shallower than 45° is indicative of overfitting, at least in a regression context, with predictions at the lower end of the probability scale tending to under-estimate prevalence and those at the upper end of the scale over-estimating prevalence (Fig. 1; Steyerberg *et al.* 2000; Pearce & Ferrier 2000). This pattern develops because the coefficients of overfitted models tend to be biased towards more extreme values, which in turn leads to probabilities that are too extreme (Miller *et al.* 1991). Overall calibration problems are evident from an overall shift in the regression line above or below the 45° diagonal (Fig. 1). Calibration plots can also identify non-linear and localised inaccuracies in predicted probabilities.

Harrell *et al.* (1996) suggested the use of scatterplot smoothers, such as LOWESS, on the raw predictions-observations as an improved approach to producing calibration plots. Such plots are very versatile, because by altering the degree of smoothing used, they can reveal either the overall calibration pattern or localised aspects. Calibration plots are largely subjective, however, both in requiring an arbitrary choice of group size or smoothing neighbourhood, and in judging what constitutes an important calibration problem.

Cox (1958) proposed an objective method for assessing calibration, which includes a basis for testing the significance of calibration problems. Pearce & Ferrier (2000) described the method at length (e.g. Pearce *et al.* 2001; Pearce *et al.* 2002). A univariate logistic regression model is fitted in the test data, using the probabilistic predictions made by the model under test – logit-transformed – as the predictor for the

observed presence/absence. The intercept in the resulting regression equation addresses the overall calibration, whilst the slope coefficient considers the pattern of calibration across the unit probability scale. A perfectly calibrated model has an intercept of zero and a slope of one, and deviations from these can be tested using likelihood ratio tests (Miller *et al.* 1991; Pearce & Ferrier 2000). The slope is the more useful coefficient, as values between zero and one are indicative of overfitting (Steyerberg *et al.* 2000). Minor overfitting in regression models can be corrected by multiplying their coefficients by the calibration slope coefficient to 'shrink' them back towards their true values (Steyerberg *et al.* 2000).

The use of logistic regression to assess calibration has some weaknesses. With relatively small sets of test data, the estimation of regression coefficients may be unreliable, as with any small sample regression, affecting both the identification of calibration problems and the correction of any problems. This again argues for a large test set. Estimation of the slope and intercept coefficients is inter-related, making their interpretation difficult (Seillier-Moiseiwitsch 1996). For example, if the slope is not equal to unity, the intercept considers calibration at 0.50 rather than overall (Miller *et al.* 1991). Finally, the Cox approach cannot describe non-linear calibration problems (Harrell *et al.* 1998).

3.5 Selection of accuracy statistics 2: nominal presence/absence predictions

Model evaluation based upon nominal presence/absence predictions is required in a range of ecological situations: where a dedicated classifier is used to predict presence/absence (e.g. a classification tree), if a particular application demands presence/absence predictions (e.g. predicted species lists for sites) or where a simple

approach to the integration of the costs associated with different prediction errors is required (Guisan & Zimmermann 2000). In such situations, model evaluation considers classification accuracy, typically based upon the 2×2 confusion or classification matrix.

If continuous/probabilistic predictions are the initial output from a model, model evaluation is better performed prior to dichotomisation. Classification matrices only allow overall goodness-of-fit to be assessed and prevent clear separation of discrimination from calibration, making model evaluation highly dependent upon the particular test data, confounding inter-model comparisons (Miller *et al.* 1991; Hosmer & Lemeshow 2000). Calibration assessments are restricted to an overall comparison between predicted and observed prevalence. Once dichotomised, only limited information about a model can be extracted and the potential to diagnose the causes of prediction errors is reduced. A model is tested under one scenario out of the full range of possible thresholds between zero and one (Pearce & Ferrier 2000). Threshold-dependent evaluation can always be performed after a threshold independent one, to assess classification performance in the particular scenario. By presenting the results of both sets of testing, a model can be more easily compared to others.

Species' prevalence and the concomitant problem of 'chance' agreement complicate inter-model comparisons based upon confusion matrices (Manel *et al.* 2001). The overall prediction success (matching coefficient, correct classification rate) provides a useful summary of model performance but, as widely recognised, may assume high values even if a model has little or no predictive ability, if the prevalence is extreme (Fielding & Bell 1997; Olden, Jackson & Peres-Neto 2002). For example, a model

with no predictive ability that always predicted a species to be absent would be correct in a large proportion of cases if the species occurs at low prevalence, but would clearly have no conservation value.

Kappa statistics have been widely adopted in ecology because they attempt to correct for 'chance' agreement (Titus, Mosher & Williams 1984; Fielding & Bell 1997; Manel *et al.* 2001). Their use requires caution for two reasons: the interpretation of chance-corrected agreement and the way in which 'chance' is defined. Agreement beyond chance is a combined property of a model's accuracy and the prevalence in the test data, making it difficult to compare between studies where the prevalence differs. It does not describe how often a model's predictions are correct and, indeed, may mislead potential users of a model. At the extremes of prevalence the potential for chance agreement is large, making it virtually impossible for a model to demonstrate much agreement beyond chance, even if it is very accurate. This is evident in the unimodal relationship often observed between prevalence and Cohen's κ (e.g. Kraemer 1979; Thompson & Walter 1988; Feinstein & Cicchetti 1990; Manel *et al.* 2001; McPherson *et al.* in press). Highly accurate and valuable models, notably those for rare species, may be disregarded due to relatively low kappa values.

Similarly, the relationship between κ and prevalence renders scales for its interpretation (e.g. Landis & Koch 1977) meaningless. Standardised prevalence across studies would be required to make meaningful model comparisons based upon chance-corrected measures (McPherson *et al.* in press). Where kappa is used, a second statistic, such as overall prediction success, is required to indicate the overall level of accuracy.

The definition of 'chance' also causes problems. The model of chance selected can strongly influence the value of κ and therefore the assessment made of a model (Brennan & Prediger 1981). Different models of chance define the different formulations of kappa, and there has been debate in the social and medical sciences concerning which is most useful (e.g. Brennan & Prediger 1981; Zwick 1988; Feinstein & Cicchetti 1990). This debate has surfaced in ecology with a consideration of a form of κ sometimes called the τ coefficient in place of Cohen's κ (Fielding & Bell 1997; Couto 2003). A different model of chance may be required for models that are true classifiers, rather than dichotomising continuous predictions (Hoehler 2000). An alternative, and perhaps better, approach is to actually simulate 'chance' agreement directly using randomisation testing. Olden *et al.* (2002) estimated chance agreement by repeatedly randomising the observed presence/absence in the training data relative to the predictors and refitting the model in each case. The average chance classification ability, measured via resubstitution, resampling or independent test data, can then be estimated with its 95% confidence interval for virtually any statistic (Olden *et al.* 2002). The null models will have identical calibration to the real model, because they use training data with identical prevalence, but no discriminatory ability.

A widely used alternative to κ in medicine is the paired sensitivity-specificity measures of a test. They are independent of prevalence and can be calculated with and without correction for chance agreement (Brenner & Gefeller 1994). The use of two measures to describe performance may further complicate inter-model comparisons, however. If both sensitivity and specificity vary between models it can be difficult to

say which models are better (Glas *et al.* 2003), especially as the relative importance of sensitivity and specificity will vary between applications.

A range of other statistics is available to summarise confusion matrices within a single measure. The odds ratio is widely applied in medicine, but has seen little application in ecology, largely because of the problems associated with its calculation when the confusion matrix contains zero values (Manel *et al.* 2001). A simple continuity correction, adding 0.5 to each of the cells in the matrix, overcomes this problem, allowing the odds ratio and other statistics sensitive to zero entries to be estimated (Forbes 1995). The odds ratio provides a definition of chance performance (unity), rather than attempting to correct for it. Unfortunately, at extremes of prevalence – commonly encountered with rare species, for example – small changes in accuracy can have massive effects upon the odds ratio, exaggerating agreement and making it difficult to interpret (Kraemer 2004).

Information theory provides an alternative paradigm for model evaluation (c.f. model development paradigm; Rushton *et al.* 2004) that may allow more general model comparisons. Rather than attempting to directly describe classification accuracy, information-theoretic measures aim to quantify the amount of information that a set of predictions provides about its matched observations (Finn 1993; Forbes 1995). This information can also be considered as the information that the observations provide about the predictions, and so is denoted mutual information (Forbes 1995). If a model's predictions are totally unrelated to the observed presence/absence, mutual information is zero. It is maximal when knowledge of predictions allows perfect classification (Forbes 1995). This need not be perfect classification of predictions,

however, because a model that always predicts the reverse of what is observed can be used to infer perfect classification. This is the non-monotonic behaviour of mutual information measures upon which information-theoretic criteria have been criticised (e.g. Fielding & Bell 1997). It is analogous to an AUC < 0.5 , indicating that a model has predictive ability, but that it is in the opposite direction from that expected i.e. predicting absence in place of presence. A cursory examination of the confusion matrix immediately reveals whether this is happening, allowing the mutual information statistic to be qualified (Forbes 1995).

The Normalised Mutual Information (NMI) has seen some application in ecology (e.g. Manel *et al.* 2001; Wright & Fielding 2002). It is the difference between the overall information contained in the confusion matrix and that in the predictions, divided by the information contained in the observed presence/absence, all taken from one (Forbes 1995). The formula is given in terms of the confusion matrix in Table 4 of Manel *et al.* (2001), except that the quantity defined should be subtracted from one. The NMI is scaled so that it ranges from zero (no predictive ability) to one (complete information). Its derivation means that it does not correct for chance agreement or provide a clear definition of it (Forbes 1995). However, data sets with extreme prevalence contain little information and models that purely predict the most common class provide virtually no information about the classification problem, resulting in a low NMI even with a high level of prediction/observation agreement. Empirical demonstrations of the NMI indicate that it does not vary systematically with prevalence (Manel *et al.* 2001). Finn (1993) and Couto (2003) describe a measure closely related to the NMI: the Average Mutual Information (AMI).

Mutual information may prove to be very useful in ecology, summarising the most fundamental concern about predictions – how much they reveal about the actual distribution – whilst circumventing the problems associated with chance-correction. More work to assess its potential would be valuable.

3.6 Case study

A simple case study provides a useful opportunity to illustrate the issues of measuring discriminatory ability, alternative approaches to testing calibration and analysing generalisability for probabilistic models. As such it represents a common situation in current applications of species distribution models.

Logistic regression was used to predict the distribution of plumbeous redstarts *Rhyacornis fuliginosus* along Himalayan rivers using altitude and slope as predictors. Plumbeous redstart distribution was recorded as part of three surveys of riverine biodiversity in the Himalaya of India and Nepal, made in the winters of 1994–1996 (Manel *et al.* 1999a, 1999b and Manel, Buckton & Ormerod 2000 give details of the data collection). The 1994 survey collected data from three regions of Nepal ($n = 75$) and was used to train the model. The model was tested using data collected in 1995 and 1996 ($n = 105$), from different regions of the Himalaya of both India and Nepal. Due to the relatively small size of the test set, no attempt was made to stratify by region or year, so that more precise, reliable estimates of accuracy could be made.

3.6.1 RESUBSTITUTION AND RESAMPLING PERFORMANCE

The model produced highly discriminatory predictions of plumbeous redstart occurrence in the training data ($c = 0.91, \pm 0.03$). Two hundred bootstraps were then

generated and overfitting estimated by the bootstrapping method described above (Table 1). Overfitting was slight, estimated at 0.01 over-optimism in c and 0.11 in the Cox calibration slope, giving corrected values of 0.90 and 0.89 for c and slope respectively.

3.6.2 PERFORMANCE IN THE TEST SET

Accuracy in the test data was substantially lower than that estimated by bootstrapping, indicating transportability problems. Whilst the overall calibration was good, underestimating prevalence by 3%, the c -index decreased to 0.80 (± 0.04) and major non-linear calibration problems were evident (Fig. 2). Across low to intermediate probabilities (0.15–0.50) the prevalence of plumbeous redstarts was under-estimated by 10–20%. More dramatically, plumbeous redstarts were only detected at around 40–50% of sites with predicted probabilities greater than 0.7 (Fig. 2). The calibration slope, 0.53, similarly identified calibration problems, but was a poor descriptor of them because it could not convey their non-linear nature.

To clarify the apparent limits to transportability, the accuracy of individual predictions was plotted against the predictor variables (altitude and slope). Accuracy was described by the contribution that individual predictions made to the overall Pearson χ^2 goodness-of-fit statistic for the logistic model used to test calibration ($\Delta_i \chi^2$; Pregibon 1981; Miller *et al.* 1991). Brier's (1950) scoring rule, the squared difference between a probabilistic prediction and observed presence/absence, could also have been used. An initial plot of $\Delta_i \chi^2$ against the predicted probability (Fig. 3a) identified a cluster of sites causing the calibration problems at high probability values. These were found to reflect poor performance at altitudes below 1000-m (Fig. 3b)

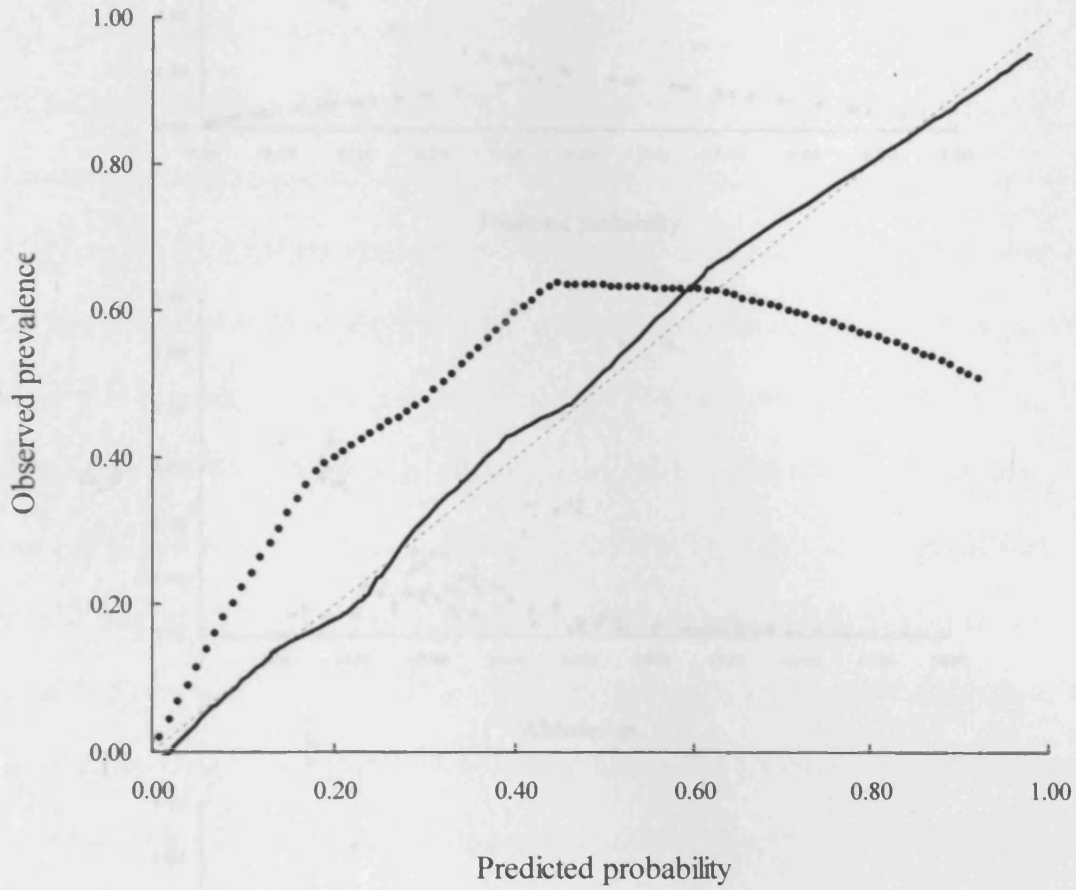


Figure 2. Non-parametric calibration curves for the plumbeous redstart model, both in the training data (heavy solid line) and in the test data set (dotted line), compared to perfect calibration (fine dashed line). Smoothing was performed by the LOESS method, with no iterations and a smoothing neighbourhood contained 60% of the sites.

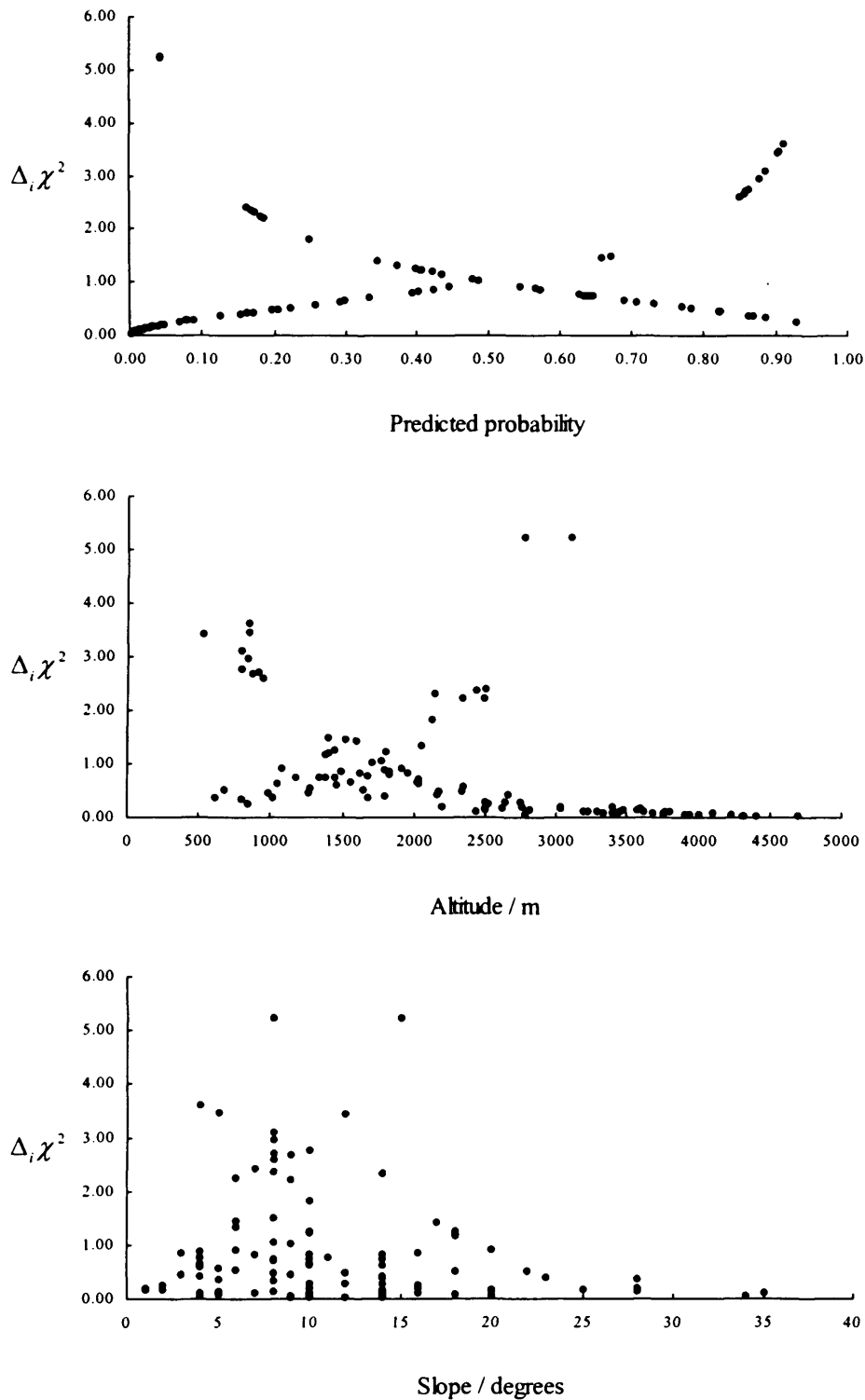


Figure 3. The accuracy of predictions for sites in the test set ($n = 105$) as indicated by their individual contributions to the overall Pearson χ^2 for the Cox calibration model ($\Delta_i \chi^2$). In (a) accuracy is plotted against the predictions (probabilities), in (b) against the altitude of the sites, and in (c) against the slope of the sites.

with no obvious relationship to slope (Fig. 3c). The calibration problems at intermediate probabilities (Fig. 2) reflected higher prevalence of the birds at intermediate altitude, including a cluster of sites between 2000 and 2500-m (Fig. 3b).

3.6.3 CONCLUSIONS FROM TESTING

Altitude and slope were effective predictors of plumbeous redstart occurrence on Himalayan streams, consistent with the bird's behaviour as an altitudinal migrant, occupying lower altitude streams during the winter (Inskipp & Inskipp 1991). The proposed model showed only slight overfitting, consistent with a relatively large sample size compared to model complexity (only two predictors). Transportability was limited, however. Plumbeous redstarts were much less likely to use sites that were below 1000m in the test data than in the training data. This may have reflected different habitat use between the regions sampled or perhaps a differing pattern of migration in the winters of 1995 and 1996, with birds spread across intermediate and low altitudes more evenly than in 1994, when there was a greater concentration at the lowest altitudes. This latter suggestion is also consistent with the calibration problems at intermediate prevalence (Fig. 2). A larger set of test data would be required to resolve these issues. Ultimately, the addition of further predictors to the model – especially direct rather than indirect ones (Vaughan & Ormerod 2003) – may be required to improve transportability. In the meantime, the model would have to be restricted to sites above 1000m and care taken in interpreting the probabilities given to sites of intermediate altitude, although the model accurately placed sites in rank order above 1000m.

3.7 Conclusions

Testing is a vital stage in developing predictive distribution models (Rushton *et al.* 2004). It should aim not only to provide a rigorous test of a model's predictive ability – both accuracy and generalisability – but should do so in a way that provides the greatest information for ecologists and model developers working across the field of distribution modelling, and planners or practitioners dependent on their outputs. Tests of model performance should therefore be tailored to the eventual applications of a model, should employ statistics with the most general meanings, should consider discrimination *and* calibration, and should attempt to diagnose model problems.

Frequently, tests of species distribution models are much more cursory. This may create problems when models are applied beyond the extent of their validation: in different applications or using probabilistic predictions without testing their calibration. Such models could mislead conservation planning (Loiselle *et al.* 2003), inappropriately targeting limited resources, and it may be difficult to defend or justify conservation recommendations based upon them to a sceptical audience. To avoid these problems, a sound, rigorous approach to model evaluation is required.

The first step, immediately following model development, should employ resample testing to assess overfitting and provide a relatively unbiased estimate of model performance. Bootstrapping or jack-knifing are the most efficient approaches.

Assuming distinct test data are not available at this stage, promising performance during resampling may form the justification for pursuing development of the model and for the expense of collecting test data (Verbyla & Litvaitis 1989).

The second stage is to collate test data, either collecting bespoke field data or collecting information from existing databases. The critical issue is to identify the range of possible applications for a model and to ensure that test data represent them. If both training and test data were collecting within a single field season, for example, good test performance is no guarantee that the model will generalise to applications in following seasons. The range of conditions used for testing should be made clear, so that the model is not applied outside of them except with extreme caution. The size of the test set should be sufficient to obtain precise performance estimates: 200–300 sites may be a realistic minimum.

The final stage is the measurement of predictive performance with the available data. For probabilistic models it is essential that calibration be tested as well as discrimination. The *c*-index/non-parametric AUC provides an effective, pure measure of the latter, which is simple to calculate and interpret, and is largely independent of the composition of the test set. Plots based on scatterplot smoothers are an effective approach to assessing calibration, revealing both linear and non-linear patterns. The Cox approach to calibration is useful when overfitting is of interest. It provides a linear description of calibration problems and a basis for shrinking model coefficients to reduce overfitting. In the specific case where >1 model for the same species is being compared with the same data, overall goodness-of-fit measures, which are more powerful than non-parametric measures (e.g. *c*), may be employed to increase the power of inter-model comparisons.

Nominal presence/absence predictions present a greater challenge to general testing, because calibration cannot be completely separated from discrimination. The

problems are acute for dichotomised continuous predictions, rather than those produced by a true classifier, due to the potential bias introduced, and information lost, by applying a threshold (Kramer 1979; Altman *et al.* 1994; Hoehler 2000). In such cases, threshold-independent evaluation should be used. The role of kappa statistics is to detect accuracy beyond chance, rather than revealing how well a model performs overall, and so are only meaningful for inter-model comparisons if prevalence is constant (McPherson *et al.* in press). Information-theoretic measures appear promising and further work should investigate their potential value in ecology. Confusion matrices should always be displayed, so that others can calculate different statistics as deemed necessary to interpret model accuracy.

The conclusion of testing a species distribution model should be clear statements of the conditions under which testing was performed, how accurate it was, where its limits of generalisability were and, hopefully, the causes of those limitations and ways to correct them. Beyond this assessment of performance, broader issues such as costs and complexity can be considered to evaluate the actual value of a model to its potential users.

3.8 Acknowledgements

I would like to thank Dr Ingrid Jüttner for preparation of the raw Himalayan data.

Collection of the Himalayan data was funded by the Darwin Institute for the Survival of Species.

3.9 References

Altman, D.G., Lausen, B., Sauerbrei, W. & Schumacher, M. (1994) Dangers of using “optimal” cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute*, **86**, 829–835.

Altman, D.G. & Royston, P. (2000) What do we mean by validating a prognostic model? *Statistics in Medicine*, **19**, 453–473.

Ash, A. & Schwartz, M. (1999) R^2 : a useful measure of model performance when predicting a dichotomous outcome. *Statistics in Medicine*, **18**, 375–384.

Aspinall, R. (1992) An inductive modelling procedure based on Bayes’ theorem for analysis of pattern in spatial data. *International Journal of Geographical Information Systems*, **6**, 105–121.

Bamber, D. (1975) The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, **12**, 387–415.

Berry, P.M., Dawson, T.P., Harrison, P.A. & Pearson, R.G. (2002) Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global Ecology & Biogeography*, **11**, 453–462.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F. K.A. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300.

Brennan, R.L. & Prediger, D.J. (1981) Coefficient kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, **41**, 687–699.

Brenner, H. & Gefeller, O. (1994) Chance-corrected measures of the validity of a binary diagnostic test. *Journal of Clinical Epidemiology*, **47**, 627–633.

Brier, G.W. (1950) Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**, 1–3.

Carroll, C., Zielinski, W.J. & Noss, R.F. (1999) Using presence-absence data to build and test spatial habitat models for the fisher in the Klamath region, U.S.A. *Conservation Biology*, **13**, 1344–1359.

Charlson, M.E., Ales K.L., Simon, R. & MacKenzie, C.R. (1987) Why predictive indexes perform less well in validation studies: is it magic or methods? *Archives of Internal Medicine*, **147**, 2155–2161.

Chatfield, C. 1995. Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A*, **158**, 419–466.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

- Couto, P. (2003) Assessing the accuracy of spatial simulation models. *Ecological Modelling*, **167**, 181–198.
- Cox, D.R. (1958) Two further applications of a model for binary regression. *Biometrika*, **45**, 562–565.
- Cox, D.R. & Wermuth, N. (1992) A comment on the coefficient of determination for binary responses. *American Statistician*, **46**, 1–4.
- Cumming, G.S. (2000) Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography*, **27**, 441–455.
- Donner, A. (1998) Sample size requirements for the comparison of two or more coefficients of inter-observer agreement. *Statistics in Medicine*, **17**, 1157–1168.
- Efron, B. (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, **78**, 316–331.
- Efron, B. & Gong, G. (1983) A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, **37**, 36–48.
- Efron, B. & Tibshirani, R. (1997) Improvements on cross-validation: the .632+ bootstrap method. *Journal of the American Statistical Association*, **92**, 548–560.

Feinstein, A.R. & Cicchetti, D.V. (1990) High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, **43**, 543–549.

Fielding, A.H. (1999) How should accuracy be measured? *Machine Learning Methods for Ecological Applications* (ed A.H. Fielding), pp. 209–223. Kluwer academic publishers, Boston.

Fielding, A.H. (2002) What are the appropriate characteristics of an accuracy measure? *Predicting Species Occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 271–280. Island Press, Washington.

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

Fielding, A.H. & Haworth, P.F. (1995) Testing the generality of bird-habitat models. *Conservation Biology*, **9**, 1466–1481.

Finn, J.T. (1993) Use of the average mutual information index in evaluating classification error and consistency. *International Journal of Geographical Information Systems*, **7**, 349–366.

Forbes, A.D. (1995) Classification-algorithm evaluation: five performance measures based on confusion matrices. *Journal of Clinical Monitoring*, **11**, 189–206.

Franco, A.M.A., Brito, J.C. & Almeida, J. (2000) Modelling habitat selection of common cranes *Grus grus* wintering in Portugal using multiple logistic regression.

Ibis, **142**, 351–358.

Glas, A.S, Lijmer, J.G., Prins, M.H., Bonsel, G.J. & Bossuyt, P.M.M. (2003) The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical*

Epidemiology, **56**, 1129–1135.

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

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Harrell, F.E. (2001) *Regression Modeling Strategies*. Springer-Verlag, New York.

Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L. & Rosati, R.A. (1982) Evaluating the yield of medical tests. *Journal of the American Medical Association*, **247**, 2543–

2546.

Harrell, F.E., Lee, K.L., Califf, R.M., Pryor, D.B. & Rosati, R.A. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, **3**,

143–152.

Harrell, F.E., Lee, K.L. & Mark, B.B. (1996) Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, **15**, 361–387.

Harrell, F.E., Margolis, P.A., Gove, S., Mason, K.E., Mulholland, E.K., Lehmann, D., Muhe, L., Gatchalian, S., Eichenwald, H.F. (1998) Development of a clinical prediction models for an ordinal outcome: the World Health Organization multicentre study of clinical signs and etiological agents of pneumonia, sepsis and meningitis in young infants. *Statistics in Medicine*, **17**, 909–944.

Hirsch, R.P. (1991) Validation samples. *Biometrics*, **47**, 1193–1194.

Hoehler, F.K. (2000) Bias and prevalence effects of kappa viewed in terms of sensitivity and specificity. *Journal of Clinical Epidemiology*, **53**, 499–503.

Hosmer, D.W. & Lemeshow, S. (2000) *Applied Logistic Regression* (2nd edn). John Wiley & Sons, New York.

Inskipp, C. & Inskipp, T. (1991) *A Guide to the Birds of Nepal*. Christopher Helm, London.

Johnson, C.J., Seip, D.R. & Boyce, M.S. (2004) A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. *Journal of Applied Ecology*, **41**, 238–251.

Justice, A.C., Covinsky, K.E., & Berlin, J.A. (1999) Assessing the generalizability of prognostic information. *Annals of Internal Medicine*, **130**, 515–524.

Kraemer, H.C. (1979) Ramifications of a population model for κ as a coefficient of reliability. *Psychometrika*, **44**, 461–472.

Kraemer, H.C. (2004) Reconsidering the odds ratio as a measure of 2×2 association in a population. *Statistics in Medicine*, **23**, 257–270.

Landis, J.R. & Koch, G.C. (1977) The measurement of observer agreement of categorical data. *Biometrics*, **33**, 159–174.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G. & Williams, P.H. (2003) Avoiding pitfalls in using species distribution models in conservation planning. *Conservation Biology*, **17**, 1591–1600.

Mackillop, W.J. & Quirt, C.F. (1997) Measuring the accuracy of prognostic judgements in oncology. *Journal of Clinical Epidemiology*, **50**, 21–29.

Manel, S., Buckton, S.T. & Ormerod, S.J. (2000) Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan rivers. *Journal of Applied Ecology*, **37**, 756–770.

Manel, S., Dias, J.M., Buckton, S.T. & Ormerod, S.J. (1999a) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734–747.

Manel, S., Dias, J.M. & Ormerod, S.J. (1999b) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, **120**, 337–347.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921–931.

McPherson, J.M., Jetz, W. & Rogers, D.J. (in press) The effects of species' range size on the accuracy of environmental distribution models – ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, in press.

Meggs, J.M., Munks, S.A., Corkrey, R. & Richards, K. (2004) Development and evaluation of predictive habitat models to assist the conservation planning of a threatened lucanid beetle, *Hoplogonus simsoni*, in north-east Tasmania. *Biological Conservation*, **118**, 501–511.

Miller, M.E., Hui, S.L. & Tierney, W.M. (1991) Validation techniques for logistic regression models. *Statistics in Medicine*, **10**, 1213–1226.

Olden, J.D. & Jackson, D.A. (2002) A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, **47**, 1976–1995.

Olden, J.D., Jackson, D.A. & Peres-Neto, P. (2002) Predictive models of fish species distributions: a note on proper validation and chance predictions. *Transactions of the American Fisheries Society*, **131**, 329–336.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.

Pearce, J., Ferrier, S. & Scotts, D. (2001) An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. *Journal of Environmental Management*, **62**, 171–184.

Pearce, J., Venier, L.A., Ferrier, S. & McKenney, D.W. (2002) Measuring prediction uncertainty in models of species distribution. *Predicting Species Occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 383–390. Island Press, Washington.

Pereira, J.M.C. & Itami, R.M. (1991) GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. *Photogrammetric Engineering & Remote Sensing*, **57**, 1475–1486.

Poses, R.M., Cebul, R.D., Collins, M. & Fager, S.S. (1986) The importance of disease prevalence in transporting clinical prediction rules. *Annals of Internal Medicine*, **105**, 586–591.

Pregibon, D. (1981) Logistic regression diagnostics. *The Annals of Statistics*, **9**, 705–724.

Rushton, S.P., Ormerod, S.J. & Kerby, G. (2004) New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**, 193–200.

SAS Institute Inc. (1999) *SAS OnlineDoc® Version 8*. Cary, NC, USA.

Seillier-Moiseiwitsch, F. (1996) Predictive diagnostics for logistic models. *Statistics in Medicine*, **15**, 2149–2160.

Snee, R.D. (1977) Validation of regression models: methods and examples. *Technometrics*, **19**, 415–428.

Steyerberg, E.W., Bleeker, S.E., Moll, H.A., Grobbee, D.E. & Moons, K.G.M. (2003) Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*, **56**, 441–447.

Steyerberg, E.W., Eijkemans, M.J.C., Harrell, F.E. & Habbema, J.D.F. (2000) Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in Medicine*, **19**, 1059–1079.

Steyerberg, E.W., Harrell, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y. & Habbema, J.D.F. (2001) Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, **54**, 774–781.

Suárez-Seoane, S., Osborne, P.E. & Alonso, J.C. (2002) Large-scale habitat selection by agricultural steppe birds in Spain: identifying species-habitat responses using generalized additive models. *Journal of Applied Ecology*, **39**, 755–771.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Thompson, W.D. & Walter, S.D. (1988) A reappraisal of the kappa coefficient. *Journal of Clinical Epidemiology*, **41**, 949–958.

Titus, K., Mosher, J.A. & Williams, B.K. (1984) Chance-corrected classification for use in discriminant analysis: ecological applications. *American Midland Naturalist*, **111**, 1–7.

Vaughan, I.P. & Ormerod S.J. (2003) Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conservation Biology*, **17**, 1601–1611.

Venier, L.A., Pearce, J., McKee, J.E., McKenney, D.W. & Niemi, G.J. (2004) Climate and satellite-derived land cover for predicting breeding bird distribution in the Great Lakes Region. *Journal of Biogeography*, **31**, 315–331.

Verbyla, D.L. & Litvaitis, J.A. (1989) Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environmental Management*, **13**, 783–787.

Wright, A. & Fielding, A.H. (2002) Modeling wildlife distribution within urbanized environments: an example with the Eurasian badger *Meles meles* L. in Britain.

Predicting Species Occurrences: issues of accuracy and scale (eds J.M. Scott, P.J.

Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 255–262. Island Press, Washington.

Wright, J.F. (1995) Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology*, **20**, 181–197.

Wyatt, J.C. & Altman, D.G. (1995) Prognostic models: clinically useful or quickly forgotten? *British Medical Journal*, **311**, 1539–1541.

Yates, J.F. (1982) External correspondence: decompositions of the mean probability score. *Organizational Behavior and Human Performance*, **30**, 132–156.

Yates, J.F., Price, P.C., Lee, J.-W. & Ramirez, J. (1996) Good probabilistic forecasters: the 'consumer's' perspective. *International Journal of Forecasting*, **12**, 41–56.

Zwick, R. (1988) Another look at interrater agreement. *Psychological Bulletin*, **103**, 374–378.

Chapter 4

Simplifying complex data in ecology: improving the interpretability and versatility of principal components analysis

4.1 Summary

1. Ecological data often combine numerous environmental variables with relatively small sample sizes. To analyse such data reliably, preliminary data reduction is often required, and principal components analysis (PCA) is the standard approach. There are concerns about the difficulty of interpreting principal components, however. Moreover, resulting ecological tools (e.g. species distribution models) can be unnecessarily expensive since they require the recording of all the original variables to calculate the principal components, even where some variables prove to be redundant.
2. River Habitat Survey (RHS) data represent an acute challenge for data analysis. They comprise over 100 variables describing river structure, often recorded on 3–6 point ordinal scales. This chapter considers whether (i) the application of a variable clustering step, prior to PCA, could improve upon the interpretability of PCA and overcome cost concerns, and (ii) whether a generalised PCA method developed in psychology could benefit analyses using ordinal and other qualitative data.
3. Agglomerative cluster analysis was used to cluster variables based upon their correlation matrix, and the first principal component of the variables within each cluster formed the cluster score. Where clusters contained ordinal variables, they were scored secondarily by the optimally-scaled PCA method designed for qualitative data. The efficacy of these two approaches, both in terms of interpretability and efficiency, was compared to conventional PCA using all the RHS variables by building distribution models for four Welsh river birds.
4. Twenty-three variable clusters were defined from the initial set of 109 variables, describing habitat factors that were more readily interpreted than the principal

components derived without prior clustering. The predictive performance of the two approaches was similar. Optimally-scaled PCA compromised little interpretability relative to conventional PCA for cluster scoring, yet explained 24% more variance in the first component, and marginally improved the accuracy of distribution models.

5. PCA with prior variable clustering represents a significant advance over conventional PCA for reducing complex data such as RHS. The resulting indices allow ease of interpretation with little loss of predictive power and offer improved utility to researchers, conservation planners and practitioners alike. In this analysis, optimally-scaled PCA provided modest, but consistent, improvement in predictive performance from clusters containing ordinal variables. This technique may allow the benefits of PCA to be generalised to other ecological applications characterised by qualitative data.

Key words: mixed measurement scales; multicollinearity; optimal scaling; qualitative data; scoring; Type I–III errors; variable selection

4.2 Introduction

Ecological data analysis can present major challenges, especially when data are derived from observational studies, rather than experimentally (Jongman, ter Braak & van Tongeren 1995; Manel, Buckton & Ormerod 2000). Observations are frequently made on many variables, due either to the complex multivariate nature of the phenomenon under study, or uncertainty over which factors might be relevant. Multicollinearity between variables is often present (Graham 2003). At the same time, resource limitations in many ecological projects restrict sample sizes, especially if they address large spatial or temporal scales (Williams & Titus 1988; Vaughan & Ormerod 2003). As a consequence, analyses are often performed with relatively few observations compared to the number of variables: frequently below recommended minimum sample sizes, such as 5–10 observations per variable for multiple regression (Harrell *et al.* 1984; Freedman & Pee 1989) or three observations per variable per group for discriminant analysis (Williams & Titus 1988). The results of such analyses are likely to suffer from errors of types I–III: spuriously identifying ‘important’ relationships, overlooking important relationships due to insufficient analytical power and describing relationships in the wrong direction, respectively (Peduzzi *et al.* 1996). The overall accuracy of such analyses may be questionable and the results sensitive to small changes in the data set.

Habitat data are among the most complex of all to appraise, often recording the highly multivariate dimensions of an organism’s niche. The River Habitat Survey (RHS), developed by the Environment Agency to provide a standard method for describing UK waterways, falls into this category as a method for determining habitat variations for organisms along rivers. RHS aims, in particular, to describe comprehensively the

physical habitats along a stretch of watercourse, tailored to incorporate features considered to be important for nature conservation (Raven *et al.* 1997). RHS data sets frequently include well in excess of 100 variables, making it extremely difficult to meet the sample size guidelines discussed above. In addition, multicollinearity occurs because different parts of the survey record similar information. Whilst the RHS has potential value in describing the habitats of riverine species (e.g. Buckton & Ormerod 1997; Brewin, Buckton & Ormerod 1998; Naura & Robinson 1998), it relies upon data reduction both to make analyses more reliable, and to make results easier to interpret.

Principal Components Analysis (PCA) is one of the most widely used statistical methods in ecology and can be very effective at addressing the problems of large numbers of variables, multicollinearity and small sample sizes (Harrell *et al.* 1984; Graham 2003). It has been used widely for RHS-based studies (e.g. Buckton & Ormerod 1997; Brewin, Buckton & Ormerod 1998; Naura & Robinson 1998; Manel *et al.* 1999a, 1999b, 2000, 2001; Hastie *et al.* 2003; Jüttner *et al.* 2003). PCA works on the basis that multicollinearity is evidence that the recorded variables describe a smaller number of latent ecological factors (Tabachnick & Fidell 2001). Following PCA, each underlying latent factor should be described by a principal component (PC), allowing the major structure of a data set – and therefore overall variation – to be described by a smaller number of PCs than original variables (D'Agostino *et al.* 1995). PCs are also constrained to be uncorrelated with one another, eliminating multicollinearity (Graham 2003). The data reduction achieved by using PCA in this way and the removal of collinearity greatly facilitate both the description of complex data sets and subsequent statistical analyses.

Unfortunately, in its conventional form, PCA has two important disadvantages for simplifying ecological data and both stem from the fact that every PC is a function of all of the original variables. The first is the difficulty of interpreting the output of the analysis: ecological information must be separated from 'noise', both in terms of how many of the components represent ecological patterns and which variables contribute meaningfully to each PC (Grossman *et al.* 1991; Jackson 1993; Franklin *et al.* 1995; Peres-Neto, Jackson & Somers 2003). For many applications, including data reduction, the former is the greater challenge (Peres-Neto *et al.* 2003). The second disadvantage of PCA concerns applications for which PCs need to be calculated in future. A common example is their use in species distribution models, where PCs are often used as predictors in place of a greater number of habitat variables (e.g. Osborne & Tigar 1992; Manel *et al.* 1999a, 1999b; Marsden & Fielding 1999; Gutzwiller & Barrow 2001). Future applications of such models require the complete set of habitat variables that were originally recorded to be measured in every case so that the PCs may be calculated, even if it emerges subsequently that many of the raw variables are redundant. This problem clearly leads to unnecessary costs.

When interpretability or cost saving is a priority, variable clustering may be an effective approach to data reduction (Cureton & D'Agostino 1983; Harrell *et al.* 1984; Marshall *et al.* 1994). A diverse range of methods is involved, but all conform to a two-stage process. In the first stage, variables are separated into discrete clusters based upon known or expected relationships between them. Clusters are expected to describe latent ecological factors in the same way as PCs (D'Agostino *et al.* 1995). In the second stage, clusters are scored so that each underlying factor (cluster) is indexed

by a single variable. PCA is one of the most effective methods for achieving this, the first PC for a cluster of variables describing the greatest proportion of a cluster's variance with a single, linear index (Cureton & D'Agostino 1983). It is the only approach to scoring clusters considered here, and in this respect variable clustering is treated as an extension of PCA. By adding the preliminary step of forming discrete clusters, the meaning of the underlying ecological factors should be easier to identify. If a cluster is deemed unimportant, its constituent variables need not be recorded for future applications (Marshall *et al.* 1994). The primary aim of the current paper is to assess whether these potential benefits are manifested with RHS data.

An additional challenge with habitat recording methods is the presence of nominal or ordinal variables. Although quickly and cheaply recorded in the field, they convey less information than continuous variables. In addition, PCA is optimised for variables recorded on continuous scales, although non-metric variables can be used (Hair *et al.* 1995). In RHS, 30–40% of the variables recorded are ordinal over a 3–6 point scale. A priority for RHS-based analyses is therefore to find techniques better suited to ordinal variables and a second aim of this paper is to contribute to this search. A variant of PCA, developed in psychology, is demonstrated as an alternative to conventional PCA for scoring clusters of variables that include ordinal habitat data. It uses an optimal-scaling method to allow data sets consisting of variables that are qualitative, semi-quantitative or recorded on mixed measurement scales, such as RHS, to be analysed (Young, Takane & de Leeuw 1978; Young 1981).

The current study investigates how data reduction using the two extensions of PCA – performing initial variable clustering and using the variant designed for qualitative

data – could benefit ecological studies involving habitat data and other complicated data sets. Two properties of the different data reduction methods are compared: interpretability and analytical efficacy. For the former, a conventional PCA of the full set of RHS variables is compared to the results obtained by preceding PCA with variable clustering. This is examined at a national level, using a sample of nearly 9000 RHS sites from across the UK. Analytical efficacy is compared for conventional PCA and PCA preceded by variable clustering, with the latter employing both conventional PCA for each cluster and optimally-scaled PCA for clusters containing ordinal variables. A case study is used for the comparison, constructing RHS-based models of river bird distributions in mid-Wales and comparing the predictive abilities of models using the three types of predictors.

4.3 Methods

4.3.1 RIVER HABITAT SURVEY

RHS is one of the most comprehensive river habitat recording methodologies ever produced, and summarises features of river habitat over a given 500-m reach in two ways. The first is as a series of 10 ‘spot-checks’ that are equidistant along the reach. At each one, features of the flow, channel, banks and immediate land use are recorded, typically as the dominant type from a pre-determined series of categories. Dominant flow types, for example, are chosen from ten possibilities.

The second part of the survey, described as the ‘sweep-up’, provides overall assessments of the extent of features over the 500-m reach and identifies features that may not have been recorded by the spot-checks (Raven *et al.* 1997). Most sweep-up variables are ordinal, usually being recorded as either absent, present over < 33% of

the reach, or present at $\geq 33\%$ (= 'extensive'). The sweep-up also includes counts of anthropogenic features and a measurement of the channel cross section (Appendix 1.2).

Habitat surveys from 8963 sites were used in the analysis. Most were collected from sites across the UK between 1995 and 2002 and were extracted from the national RHS database (Version 3.3, Environment Agency 2002). Sites with numerous missing variables were excluded, as were those sampled using the 1994 version of the survey, which differed from subsequent years'. The database was augmented by a small subset of sites in mid-Wales ($n = 37$) sampled in 2002 by an accredited RHS surveyor (I.V.).

In common with previous studies (e.g. Naura & Robinson 1998; Manel *et al.* 1999a, 1999b, 2000) some initial simplification of the RHS data preceded analysis. Spot-check variables were summed across the 10 sampling points, to give counts out of 10 for each category (e.g. one variable for each flow type). For spot-check variables describing the river banks, totals for left and right were summed to give counts out of 20. For sweep-up variables recorded on both banks, the maximum value was retained, due to the problems of summing ordinal variables. Further variables were excluded to reduce redundancy between spot-checks and sweep-up, with the former being preferred. Counts of bridges, outfalls and weirs were converted to a five-point ordinal scale to reduce the potential impact of small numbers of very large outliers. In total, 109 variables were entered into the analysis.

4.3.2 PRINCIPAL COMPONENTS ANALYSIS

PCA was performed on the correlation matrix from the complete data set of 8963 observations on 109 variables using PROC PRINCOMP in SAS version 8.2. The number of meaningful PCs was determined using a combination of the broken stick method and a scree plot (Cattell 1966; Jackson 1993).

The contribution of individual variables to the significant PCs was assessed using the bootstrapped eigenvector method recommended by Peres-Neto *et al.* (2003). Five hundred bootstraps were used and variables whose 95% confidence intervals around their loadings did not incorporate zero were considered to contribute significantly to the relevant PC.

4.3.3 VARIABLE CLUSTERING

4.3.3.1 *Cluster formation*

A correlation matrix for all 109 RHS variables was generated using Spearman's ρ coefficients. The correlations were then converted to distances as $1-|\rho|$, and agglomerative hierarchical cluster analysis performed on the distance matrix using PROC CLUSTER in SAS. Spearman's ρ was used in preference to Pearson's r as it describes all monotonic relationships, rather than just linear ones, and this was more appropriate for ordinal variables and their subsequent monotonic transformations (see below).

The identification of clusters is largely subjective and two strategies were used to assist the process. First, where distinct clusters, well separated from the other

variables, were evident, their constituent variables were removed from the distance matrix and the clustering re-run with fewer variables to focus the analysis upon the less distinct parts (Anderberg 1973). Secondly, whilst average linkage was the predominant clustering method used, equivalent analyses were run with single, complete and centroid linkage methods to help resolve clusters more clearly.

4.3.3.2 Scoring clusters

A separate PCA was performed for every variable cluster identified and the first PC used as the cluster score. For clusters containing one or more ordinal variables, equivalent scoring was also performed using the generalised PCA method developed by Young *et al.* (1978). This method alternates PCA with optimal scaling of the raw variables to maximise the variance explained by a specified number of PCs: a single cluster score in this case. Prior to the initialisation of the process, an appropriate type of transformation is selected for each variable, such as the monotonic transformation described below. The procedure begins with a conventional PCA carried out on the variables within a cluster to obtain the first PC. Univariate least-squares regressions are then fitted for each variable within a cluster, fitting the pre-specified transformations to the variables and predicting the first PC (Harrell 2001). PCA is then repeated, using the predicted values of the original PC from each of the individual regressions as the new values of the variables within the cluster (Harrell 2001). The process then iterates, with new regressions carried out to predict the new PC, which in turn form the basis for a subsequent PCA. The variance explained by the first PC usually approaches an asymptote within 10 iterations, and the process stops either when the variance explained by the first PC ceases to increase, or increases by increments below a defined threshold (Young *et al.* 1978; Young 1981; Harrell 2001).

This approach to scaling, optimised for PCA, is described as Maximum Total Variance (MTV) scaling.

MTV optimal scaling was implemented in SAS PROC PRINQUAL. Spot checks and other continuous variables were held constant during the procedure. Ordinal variables were transformed using the primary least squares method devised by Kruskal (1964), widely used for non-metric multidimensional scaling. This broadly maintains the rank order of the ordinal categories (i.e. the transformation is monotonic) but may untie pairs of sites with the same starting value of a variable (Kruskal 1964). This is an appropriate transformation for ordinal variables when the underlying process (i.e. extent of habitat features) varies continuously, but is merely recorded in an ordinal manner (Young 1981). After running the PRINQUAL procedure, conventional PCA was calculated on the transformed variables to attain scores for each cluster.

4.3.4 RIVER BIRD MODELS

To provide a sample application for the alternative data reduction methods, distribution models were built for four species of river birds: common sandpiper *Actitis hypoleucos*, dipper *Cinclus cinclus*, goosander *Mergus merganser* and grey wagtail *Motacilla cinerea*. Distribution data were collected in 1995 from 74 sites in mid-Wales that were also included in the RHS database (Buckton *et al.* 1998). The 37 extra RHS sites surveyed in 2002 were also surveyed for birds following the same protocol as in 1995, giving a total RHS-bird data set of 111 sites.

Models were fitted using logistic regression. Predictors were chosen prior to modelling, based upon the range of clusters that were formed and the published

literature regarding the species' habitat requirements. No further variable selection was performed. The total number of predictors in each model was limited to ensure that around 10 times as many observations of each species were available as predictors in a model so as to minimise overfitting (Harrell 2001). Conventional PCA models used an identical number of predictors to the equivalent models employing variable clustering. Two versions of variable cluster models were fitted. The first used conventional PCA-scored clusters. The second substituted optimally-scaled PCA-scored (OS-PCA) clusters for conventional PCA ones, where possible.

The effects of different predictors on the likelihood of species occurrence were quantified using odds ratios. These describe the change in the odds of a species' occurrence observed between two values of a predictor, holding all of the other predictors constant (Hosmer & Lemeshow 2000). To maximise the interpretability of this process, odds ratios were calculated between the first and third quartiles of each predictor for the Welsh river bird sites (c.f. nationally). The inter-quartile range tends to be a more meaningful change in the predictor than a one-unit change that is the default basis for calculating odds ratios (Harrell 2001). To describe the river habitat at each quartile, the average values of the raw RHS variables within single percentiles of the national RHS data set – centred on the quartiles calculated in the Welsh data – were calculated. This provided a more representative description of the habitat than only using the specific values at the quartiles, because very similar PCA/cluster scores may conceal relatively large variations in the raw habitat variables: the specific values at the quartiles may have been atypical of neighbouring sites.

Three measures were used to assess the predictive performance of the river bird models. First, discriminatory ability was assessed using the *c*-index of Harrell *et al.* (1982), with 95% confidence intervals calculated via bootstrapping. This was the primary focus as it measured the predictive information in the alternative predictors: how well they distinguished occupied from unoccupied sites. Secondly, calibration was assessed using the calibration slope measure (Cox 1958), the degree to which the slope was less than unity indicating the extent to which a model overfitted the training data (Steyerberg *et al.* 2000). Finally, because models built with the different types of predictors were compared with identical data, an overall measure of accuracy was employed to increase the power of the inter-model comparisons over separate discrimination and calibration measures (see Chapter 3). The Brier score (Brier 1950) was used, calculated as the mean value of the squared difference between predicted probability and observed presence/absence at each site. Overfitting was estimated for all three measures using a bootstrap method (Efron 1983).

4.3.4.1 *Waterways Breeding Bird Survey*

To extend the sample size for comparing the two variants of PCA for scoring variable clusters, models were built for a subset of species from the Waterways Breeding Bird Survey. These data are described and analysed in Chapter 5. The same modelling strategy was used as for the Welsh river birds, selecting predictors *a priori*. Models for 23 species used one or more clusters that incorporated some ordinal sweep-up variables, and so were scored using the two variants of PCA. Only the discriminatory abilities of the 23 pairs of models were examined, using only the fit to the training data.

4.4 Results

4.4.1 OVERALL PCA

Individual PCs explained relatively small proportions of the overall variance in the RHS data. Twenty-two PCs were required to describe 50% of the variance and 76 to describe 90%, whilst 32 PCs had eigenvalues greater than one. Following the broken-stick criterion, only the first three PCs were retained, explaining 19% of the variance. This was consistent with a break-point in the scree plot (Fig. 1).

The majority of the 109 variables contributed significantly to each of the three PCs: 105, 97 and 86 for PCs 1–3 respectively. Based upon the variables with the largest eigenvectors (\propto Pearson correlation coefficients with the raw variables) in each case, PC1 represented a trend from upland, rocky channels with high-energy flows to lowland, vegetated channels, often in agricultural areas. PC2 ran from tree-lined channels to rocky channels and open moorland. PC3 described anthropogenic modification. Numerous other variables contributed significantly to these PCs with moderate loadings (Fig. 2), but did not always easily fit the general trends described. Beyond PC3, overall trends on individual PCs were difficult to interpret.

4.4.2 VARIABLE CLUSTERS

Twenty-three clusters and 11 outlying variables were derived from the RHS data. Some were relatively distinct in the initial clustering of all 109 variables (Fig. 3), whereas others were only finally resolved using a combination of clustering methods and the sequential strategy of removing distinct clusters and re-analysing the remainder. Nevertheless, the majority of the arrangements used in the final cluster

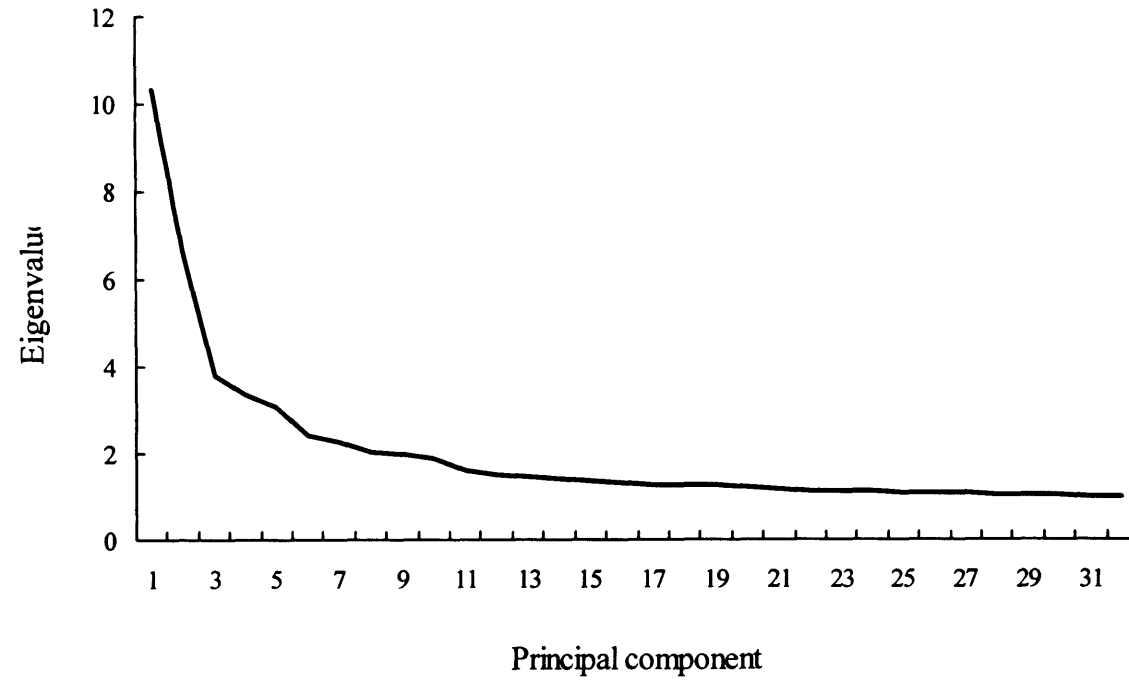


Figure 1. The scree plot from principal components analysis of RHS data. The first 32 components are shown, representing all of those with eigenvalues ≥ 1 .

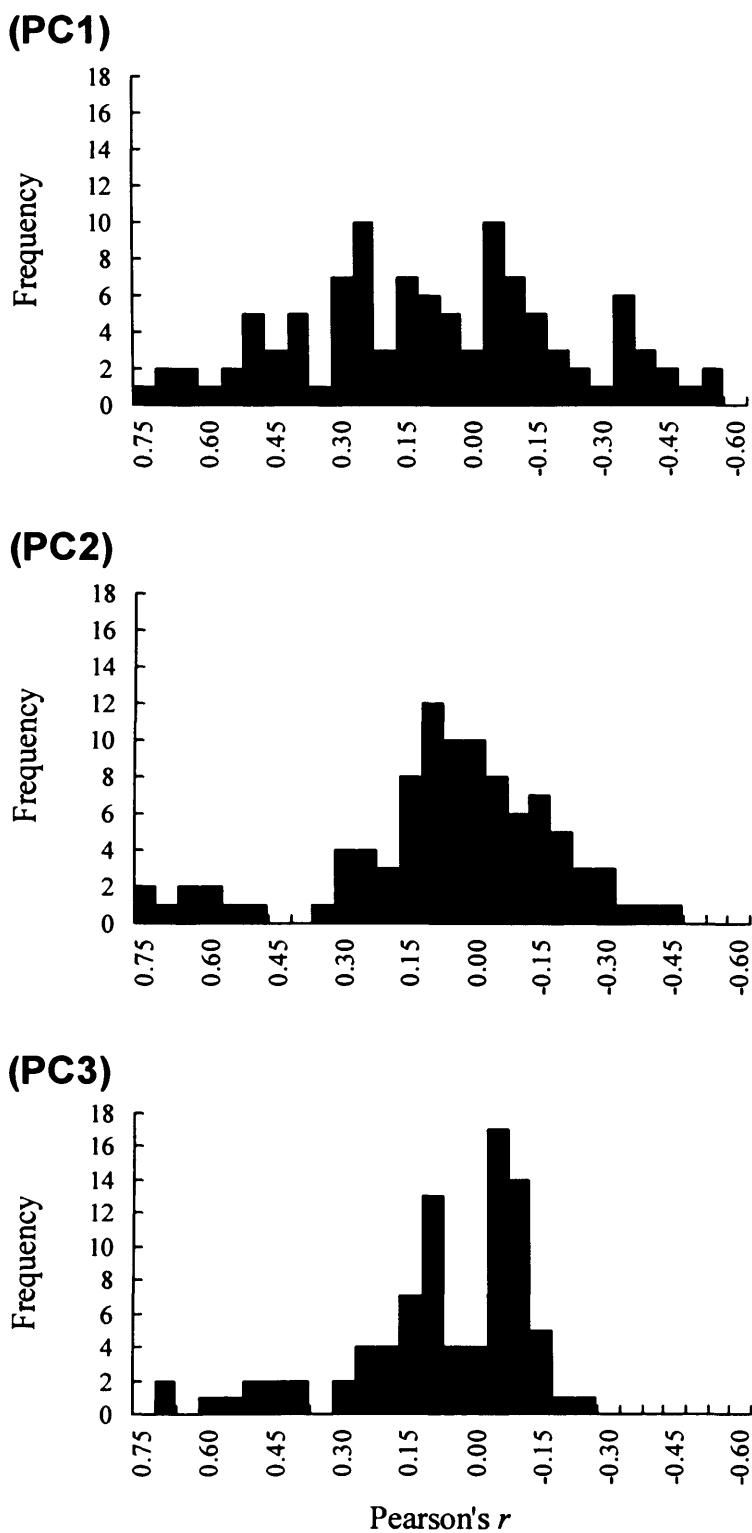


Figure 2. The distribution of Pearson correlation coefficients (\propto eigenvectors) for the variables that contributed significantly to PCs 1–3 ($n = 105, 97$ and 86 for PCs 1–3 respectively).

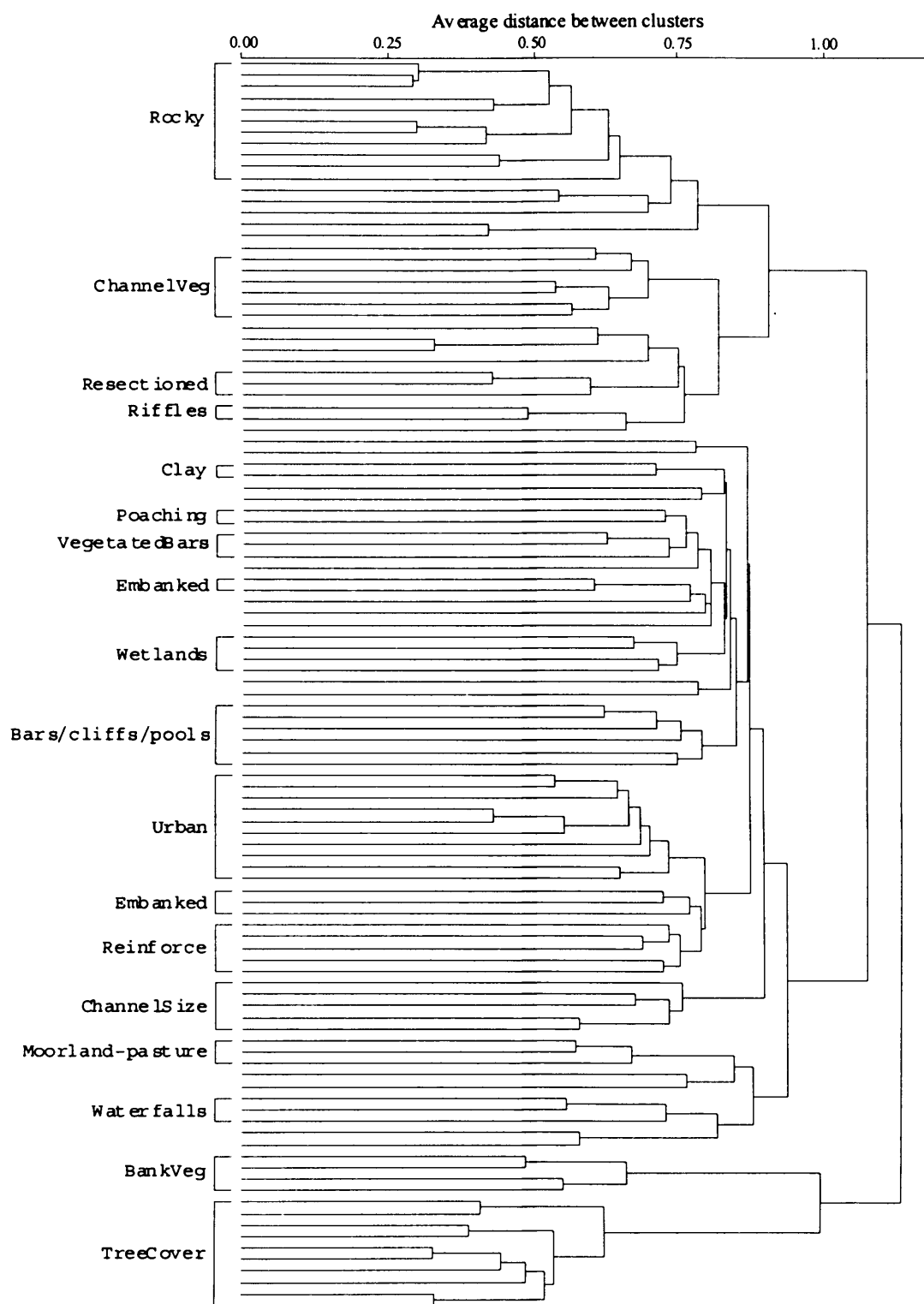


Figure 3. Average-linkage clustering tree for all 109 RHS variables in the full, 8963 site data set. Only 18 variable clusters are shown, with their approximate locations, due to the rearrangements that occurred during analyses of subsets of variables to produce the final clustering solution (see text for details). Details of cluster names and composition are given in Table 1.

solution were evident throughout and no re-arrangement based upon subject-matter knowledge was deemed necessary.

Four general categories of cluster were recognisable (Table 1):

4.4.2.1 Land use

Four clusters and four individual variables described land uses (Table 1).

TreeCover included all of the RHS variables pertaining to deciduous woodland and associated impacts on the channel, such as shading. *BankVeg* described variation in the structural complexity of vegetation within 1m of the channel. It ranged from a predominance of ‘uniform’ vegetation, which consists of a single vegetation type (without scrub or trees) such as arable farmland or grazed/improved grass swards to ‘simple’ vegetation, consisting of two structural types. The latter includes more complex swards with taller herbs amongst grass, or scrub or brambles.

Moorland-pasture was a trend from acidic uplands, with peat soils and moorland, to improved grassland at lower altitudes. *Wetlands* increased as the water table moved closer to the surface (water level nearer to floodplain level), river banks sloped more gently onto the floodplain, and wetlands and unimproved grazing became more extensive.

4.4.2.2 Anthropogenic modification

Six clusters and one individual variable concerned modifications of the river corridor (Table 1). *Urban* was the largest cluster, including RHS variables pertaining to the built environment with associated structures (e.g. bridges, weirs) and potential

Table 1. RHS variable clusters and their constituent variables. Clusters are loosely grouped according to the aspect of the river habitat that they describe. The nature of the different variables is indicated: whether they are recorded in the spot-check or sweep-up parts of the RHS, the range of values they can assume and the direction of their contribution to cluster scores (+ or -). PC and OS denote whether conventional PCA, optimally-scaled PCA or both methods were used to create cluster scores.

Land use					
TreeCover ($X = 10$, PC & OS)			Wetlands ($X = 4$, PC & OS)		
• ‘Complex’ bank-face vegetation	+	spot, 0–20	• Wetlands	+	sweep, 0–2
• ‘Complex’ bank-top vegetation	+	spot, 0–20	• Rough pasture	+	sweep, 0–2
• Broad-leaved woodland	+	sweep, 0–2	• Gentle bank profile	+	sweep, 0–2
• Extent of tree cover	+	sweep, 0–5	• Bank-top height above water level	-	sweep, >0
• Shading of channel	+	sweep, 0–2			
• Overhanging boughs	+	sweep, 0–2	Moorland-pasture ($X = 3$, PC & OS)		
• Exposed bank-side tree roots	+	sweep, 0–2	• Moorland/heath	+	sweep, 0–2
• Exposed underwater roots	+	sweep, 0–2	• Peat banks	+	spot, 0–10
• Fallen trees in channel	+	sweep, 0–2	• Improved grassland	-	sweep, 0–2
• Woody debris	+	sweep, 0–2			
			Single variables		
BankVeg ($X = 4$, PC only)			• Tilled land use	NA	sweep, 0–2
• ‘Uniform’ bank-face vegetation	+	spot, 0–20	• Conifer plantations	NA	sweep, 0–2
• ‘Uniform’ bank-top vegetation	+	spot, 0–20	• Scrub	NA	sweep, 0–2
• ‘Simple’ bank-face vegetation	-	spot, 0–20	• Adjacent open water	NA	sweep, 0–2
• ‘Simple’ bank-top vegetation	-	spot, 0–20			

Anthropogenic

Urban ($X = 10$, PC & OS)

- Urban land use + sweep, 0–2
- Number of outfalls + sweep, 0–4
- Rubbish in channel + spot, 0–10
- Number of bridges + sweep, 0–4
- Number of weirs + sweep, 0–4
- Whole bank reinforced + sweep, 0–2
- Brick banks + spot, 0–20
- Concrete banks + spot, 0–20
- ‘Bare’ bank-top vegetation + spot, 0–20
- Artificial channel substratum + spot, 0–10

Resectioned ($X = 3$, PC & OS)

- Re-sectioned channel + spot, 0–10
- Re-sectioned banks + sweep, 0–2
- Vertical or undercut banks - sweep, 0–2

Embanked ($X = 2$, PC & OS)

- Embankments + spot, 0–20
- Embankments set-back from channel + sweep, 0–2

Reinforce ($X = 5$, PC & OS)

- Bank top reinforced + sweep, 0–2
- Bank toe reinforced + sweep, 0–2
- Builder’s waste on banks + spot, 0–20
- Rip-rap banks + spot, 0–20
- Gabion banks + spot, 0–20

Piling ($X = 3$, PC & OS)

- Wood piling + spot, 0–20
- Sheet piling + spot, 0–20
- Two-stage channel + sweep, 0–2

Poaching ($X = 2$, PC & OS)

- Poaching of channel margin + spot, 0–20
- Composite bank profile + sweep, 0–2

Single variables

- Fords NA sweep, ≥ 0
-

Overall channel characteristics

Rocky ($X = 12$, PC & OS)

- Bedrock banks + spot, 0–20
- Bedrock channel substratum + spot, 0–10
- Exposed bedrock in channel + sweep, 0–2
- Boulder banks + spot, 0–20
- Boulders channel substratum + spot, 0–10
- Exposed boulders in channel + spot, 0–10
- Bryophytes + spot, 0–10
- ‘Chute’ flow type + spot, 0–10
- Cascades + sweep, 0–2
- Rapids + sweep, 0–2
- ‘Broken wave’ flow type + spot, 0–10
- Earth banks - spot, 0–20

Vegetated Bars ($X = 3$, PC only)

- Vegetated mid-channel bars + spot, 0–10
- Vegetated side bars + spot, 0–20
- Number of vegetated point bars + sweep, ≥ 0

Bars/cliffs/pools ($X = 7$, PC & OS)

- Unvegetated mid-channel bars + spot, 0–10
- Number of unvegetated point bars + sweep, ≥ 0
- Eroding cliffs $\geq 0.5\text{m}$ + spot, 0–20
- Stable cliffs $\geq 0.5\text{m}$ + spot, 0–20
- Vertical + toe bank profile + sweep, 0–2
- Number of pools + sweep, ≥ 0
- Marginal dead-water + sweep, 0–2

Channel Veg ($X = 7$, PC only)

- Submerged fine/linear-leaved plants + spot, 0–10
 - Submerged broad-leaved + spot, 0–10
 - Free floating + spot, 0–10
 - Floating-leaved (rooted) + spot, 0–10
 - Reeds/sedges/rushes + spot, 0–10
 - Emergent broad-leaved + spot, 0–10
 - Amphibious + spot, 0–10
-

Overall channel characteristics (*continued*)

ChannelSize (*X* = 4, PC only)

- Water width + sweep, ≥0
- Water depth + sweep, ≥0
- Number of culverts - sweep, ≥0
- No flow - spot, 0–10

Flow types and channel/bank substrata

Waterfalls (*X* = 3, PC & OS)

- Free-fall flow + spot, 0–10
- Chaotic flow + spot, 0–10
- Waterfalls + sweep, 0–2

Cobbles (*X* = 4, PC only)

- Cobble banks + spot, 0–20
- Cobble channel substratum + spot, 0–10
- Unvegetated side bars + spot, 0–20
- Silt substratum - spot, 0–10

Boils (*X* = 2, PC & OS)

- Extent of boils + sweep, 0–2
- 'Upwelling' flow type + spot, 0–10

GSBanks (*X* = 2, PC only)

- Gravel or sand banks + spot, 0–20
 - 'Bare' bank face + spot, 0–20
-

Flow types and channel/bank substrata (*continued*)

Riffles ($X = 2$, PC only)

- Number of riffles + sweep, ≥ 0
- 'Unbroken wave' flow type + spot, 0–10

Clay ($X = 2$, PC only)

- Clay banks + spot, 0–20
- Clay channel substratum + spot, 0–10

Runs ($X = 2$, PC & OS)

- Extent of runs + sweep, 0–2
- 'Rippled' flow type + spot, 0–10

Single variables

- Sand channel substratum NA spot, 0–10
- Gravel or pebble channel substratum NA spot, 0–10
- No perceptible flow NA spot, 0–10

Glides ($X = 2$, PC & OS)

- Extent of glides + sweep, 0–2
- 'Smooth' flow type + spot, 0–10

Others

-
- Filamentous algae NA spot, 0–10
 - Steep bank profile NA sweep, 0–2
 - Mature islands in channel NA spot, 0–10
-

degradation of the channel (outfalls and rubbish). It included completely artificial channels, where the entire banks and channel were brick or concrete. Reinforce was related to Urban, but tended to describe minor reinforcements. Typically only the top or toe of the bank was strengthened and relatively open/loose materials were used, such as rip-rap or gabions.

Resectioned, Piling and Embanked described aspects of flood defence. These clusters represented three types of modification, sometimes used simultaneously, that increase the capacity and velocity of either the existing channel or secondarily on the neighbouring flood plain. Poaching described damage to the banks caused by either humans or livestock.

4.4.2.3 Overall channel characteristics

Five clusters represented intuitive river habitat trends or types, distinct from those concerning land use or modification (Table 1). Rocky described a trend from boulder-strewn channels, with exposed bedrock, high-energy flows and bryophytes covering the rocks, to lowland channels with earth banks. Bars/cliffs/pools captured features of meandering channels, with earth cliffs on the outside of bends and point bars on the apices. Mid-channel bars occur on straighter stretches and pools and deadwater are spaced between the other features. ChannelVeg combined truly aquatic macrophytes with amphibious and emergent types. ChannelSize not only represented a trend towards wider and deeper channels, but also distinguished smaller channels that were regularly forced through culverts and watercourses that had dried up during the summer. VegetatedBars was related to Bars/cliffs/pools, but indicative of more stable areas of deposition.

4.4.2.4 Flow types and substrata

Many of the flow types and channel and bank substrata were not closely associated with other clusters (Table 1). Flow types were not necessarily tied to a particular channel substratum, with glides, for example, possible over fine substrata (e.g. silt or clay), bedrock or more coarse substrata with sufficient water depth and gentle gradients. Most of the flow types formed separate clusters, representing a series from the lowest energy (no perceptible flow) to some of the highest (Boils and Waterfalls; Table 1).

Most of the natural substrata also formed distinct clusters or outlying variables, the exception being bedrock and boulders which were characteristic of upland channels (Rocky).

4.4.3 SCORING CLUSTERS WITH PCA

Fifteen clusters contained ordinal variables and so were scored using OS-PCA as well as conventional PCA (Table 1). On average, optimally-scaled cluster scores explained 24% more of the variance within clusters than the first conventional PC (74% versus 51%). Not surprisingly, the size of the increase correlated with the proportion of variables within a cluster that was transformed ($r = 0.74, n = 15$).

All of the transformed variables had the same general form (Fig. 4). This reflected the monotonic nature of the transformation, broadly maintaining the ordering of sites across the different levels of the variable (e.g. 0, 1, 2), whilst untying pairs of sites within measurement levels as required to optimise the overall scaling.

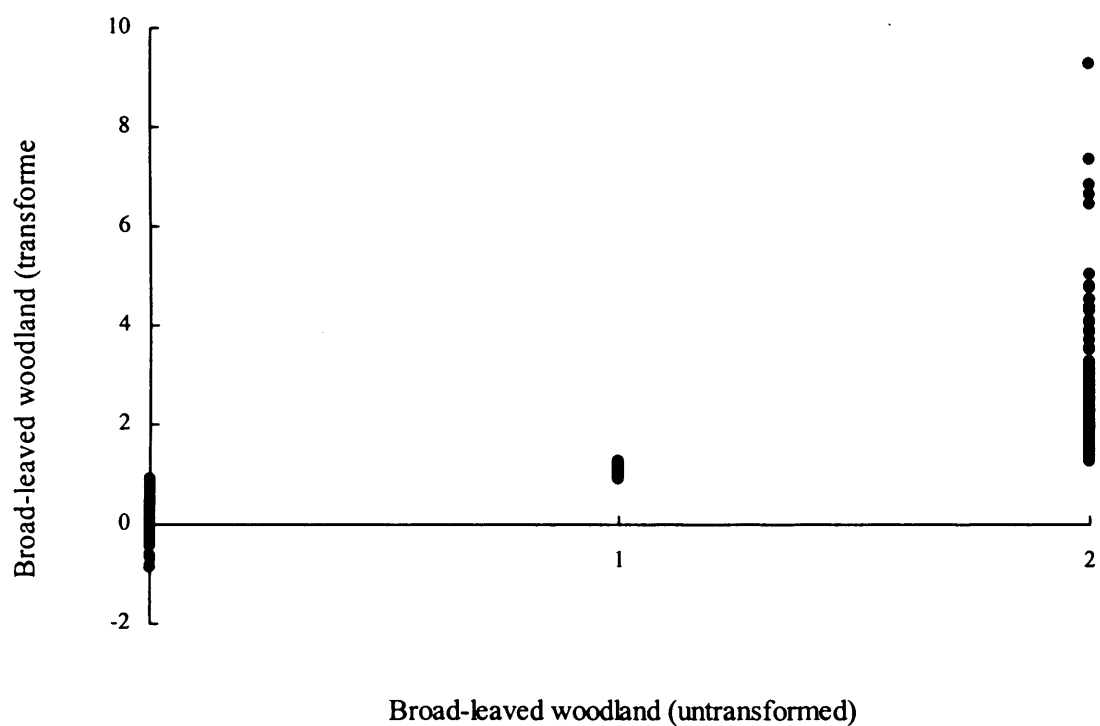


Figure 4. Typical form of the optimised transformation for ordinal sweep-up variables. Levels 0, 1 and 2 of the untransformed variable represent absence, presence at less than 33% of a reach and presence at $\geq 33\%$ of a reach respectively.

4.4.4 RIVER BIRD MODELS

Common sandpiper breeding territories are characterised by expanses of bare shingle alongside the channel, used as foraging habitat by chicks (Yalden 1986).

`Bars/cliffs/pools` was therefore selected as a predictor along with `ChannelSize`, which reflected the limited expanse of bare shingle generally associated with narrower channels and was shown elsewhere to be correlated with common sandpiper occurrence (Yalden 1986). The association between dippers and riffles for foraging is well known (e.g. Round and Moss 1984; Tyler & Ormerod 1994), and there is also a positive association between dippers and deciduous riparian trees which provide nest sites and affect in-stream food abundance (Ormerod *et al.* 1986). `TreeCover` and `Riffles` were therefore chosen as predictors, along with `ChannelSize` to reflect increasing area of foraging habitat (riffles) over the 500m RHS reach as width increases. Goosanders were only recorded at eight sites, restricting models to a single predictor. `ChannelSize` was used, reflecting the only reliable correlate of goosander density in a national survey (Gregory, Carter & Baillie 1997). This cluster was only scored using conventional PCA and so no comparison with OS-PCA was possible. The full PCA model used PC1, because it gave the highest loadings to channel width and depth amongst the three PCs (depth was not a significant contributor to PC3). For grey wagtail, `TreeCover`, `Bars/cliffs/pools` and `Rocky` were used, reflecting published habitat associations and known foraging locations (e.g. Round & Moss 1984; Tyler & Ormerod 1991; Tyler 2003).

All of the resulting river bird models, with the exception of the goosander PCA model, showed marked discriminatory ability, with c in excess of 0.70 (Table 2).

Table 2. Predictive abilities of the distribution models for the four river bird species using the three data reduction methods: overall PCA, variable clustering with conventional PCA scoring and clustering with OS-PCA scoring. OS-PCA was not used for the goosander model. Accuracy is given in the training data, along with the degree of overfitting estimated by bootstrapping. Bootstrapped calibration estimates for goosander proved unstable, due to the small number of occurrences in many bootstraps (frequently $<<8$) and so are not shown.

	Discriminatory ability: <i>c</i> -index			Calibration: calibration slope		Overall accuracy: Brier score		
	Training data (\pm SE)	Optimism	Optimism-corrected	Optimism	Optimism-corrected	Training data	Optimism	Optimism-corrected
Goosander								
PCA	0.64 (\pm 0.12)	0.05	0.60	–	–	0.065	-0.005	0.070
CLUSTER PCA	0.72 (\pm 0.07)	0.02	0.70	–	–	0.065	-0.005	0.070
Common sandpiper								
PCA	0.79 (\pm 0.05)	0.01	0.78	0.04	0.96	0.112	-0.009	0.120
CLUSTER PCA	0.78 (\pm 0.07)	0.01	0.77	0.03	0.97	0.098	-0.007	0.105
CLUSTER OS-PCA	0.84 (\pm 0.07)	0.01	0.83	0.01	0.99	0.092	-0.007	0.098
Grey wagtail								
PCA	0.75 (\pm 0.05)	0.02	0.72	0.11	0.89	0.182	-0.014	0.196
CLUSTER PCA	0.75 (\pm 0.05)	0.02	0.72	0.09	0.91	0.181	-0.012	0.194
CLUSTER OS-PCA	0.75 (\pm 0.05)	0.02	0.73	0.10	0.90	0.178	-0.013	0.191
Dipper								
PCA	0.74 (\pm 0.05)	0.02	0.71	0.11	0.89	0.206	-0.016	0.222
CLUSTER PCA	0.73 (\pm 0.04)	0.03	0.70	0.14	0.86	0.215	-0.016	0.231
CLUSTER OS-PCA	0.74 (\pm 0.05)	0.03	0.72	0.13	0.87	0.210	-0.015	0.226

Overfitting was small, with a maximum over-optimism of 0.03 in c and 0.14 in the calibration slope (Table 2). Common sandpiper models in particular combined high accuracy (bootstrapped $c \geq 0.77$) with low overfitting (Table 2).

The differences in predictive ability between the three types of predictors were small (Table 2). Across all three species, OS-PCA scored clusters produced the models with the greatest discriminatory ability. Overall PCA was usually the second most discriminatory, followed by variable clustering using conventional PCA scoring. The only marked differences between predictors were for common sandpiper, where the optimally scored cluster model had a c -index that was 0.05 larger than overall PCA, and goosander, where ChannelSize produced a c -index 0.08 larger than PC1 (Table 2). Calibration did not show a consistent pattern across the three species, with the estimated degree of overfitting similar for the three predictor types and their rank order differing in each case (Table 2). Variable clustering with OS-PCA produced the best (lowest) Brier scores for common sandpiper and grey wagtail, but was second to overall PCA for dipper. It was consistently better than conventional PCA-scored clusters.

The simplicity of variable clusters relative to the three overall PCs made the resulting models simple to interpret. The efficacy of many of the predictors was confirmed by odds ratios differing significantly from zero (Table 3). Common sandpipers' preferences for wider channels with exposed gravel bars was confirmed, with an increase from one to around five unvegetated bars in a 500m reach in particular being associated with an order of magnitude increase in the odds of their occurrence at a site

Table 3. Structures and interpretation of the variable cluster models. In (a) variable clusters are related back to the raw habitat variables at their first and third quartiles in the Welsh river bird data set. The terms ‘present’ and ‘extensive’ in the descriptions follow RHS conventions, with the former describing variables present over <33% of the 500m RHS reach and the latter referring to occurrence at $\geq 33\%$. In (b) the odds ratios (with 95% confidence intervals) are given for each predictor in each river bird model, describing the effects of moving from the first to third quartile of each variable upon the odds of species occurrence. Only OS-PCA models are shown due to the similarity of results from clustering using conventional PCA.

(a)		
Cluster	Quartile	Description of river habitat
TreeCover	First	Discontinuous tree cover over 500m. Broad-leaved woodland present. No impacts upon channel recorded e.g. shading, overhanging boughs, tree cover on the banks (‘complex’ vegetation).
	Third	Broad-leaved woodland frequently covering $\geq 33\%$ of 500m reach (within 50m of channel). 10–20% of bank length wooded. Extensive shading of channel and overhanging boughs, and all other impacts present e.g. exposed bankside roots.
ChannelSize	First	Channel dimensions approximately 2m wide by 0.15m deep. No culverts or sections of dry watercourse.
	Third	Channel dimensions approximately 6m wide by 0.20m deep. No culverts or sections of dry watercourse.
Bars/cliffs/pools	First	Approximately: three earth/sand cliffs $\geq 0.5\text{m}$ high, one unvegetated point bar, two pools and deadwater present in the margins. No mid-channel bars.
	Third	Approximately: seven earth/sand cliffs $\geq 0.5\text{m}$ high, four unvegetated points bars and one mid-channel bar, five pools and deadwater present in the margins.

Rocky	First	60% of reach with earth banks, 40% with boulders present in the channel. Patchy occurrence of exposed bedrock, cascades, rapids and other sweep-up variables.
	Third	25% of reach with earth banks, 70% of channel with boulders present. Exposed bedrock, rapids and cascades 'extensive'.
Riffles	First	Approximately: 20% of channel with 'unbroken' wave flow, 8-9 riffles.
	Third	Approximately: 30% of channel with 'unbroken' wave flow, 13 riffles.

(b)

Species	TreeCover	ChannelSize	Bars/cliffs/pools	Rocky	Riffles
Goosander	-	1.59 (0.93-2.69)	-	-	-
Common sandpiper		3.73 (1.58-8.78)	11.52 (3.08-43.16)	-	-
Grey wagtail	1.49 (1.08-2.08)	-	0.41 (0.19-0.87)	0.93 (0.41-2.11)	-
Dipper	1.44 (1.07-1.93)	1.79 (1.14-2.79)	-	-	1.19 (0.83-1.71)

(Table 3). Dippers showed preferences for wider, wooded channels. In particular, the odds of their occurring at a site nearly doubled as the river tripled in width over the inter-quartile range but remained shallow, only increasing in depth by 25% (Table 3). Grey wagtails were more likely to occur on wooded sites and less likely on sequences of unvegetated bars, pools and river cliffs, where the odds more than halved over the inter-quartile range (Table 3).

4.4.4.1 Waterways Breeding Bird Survey models

The Waterways Breeding Bird Survey produced outcomes that were consistent with the Welsh river bird models. Models incorporating OS-PCA showed small, but consistent improvements in discrimination (median increase in $c = 0.013$) over those using only conventional PCA scoring (one-sample Wilcoxon test on differences, $W = 216$, $P = 0.018$). Improvements were evident for 18 out of the 23 models.

4.5 Discussion

Ideally, the design of ecological studies should treat the dimensionality of a data set relative to the sample size and the potential for multicollinearity as priority issues (Graham 2003; Vaughan & Ormerod 2003). By maximising the sample size, focusing upon more specific issues, employing all available subject-matter knowledge to limit the number of variables considered and resisting the temptation to entertain numerous variables on a speculative basis, the statistical problems described in the introduction may largely be avoided (Derksen & Keselman 1992; Olden & Jackson 2000; Harrell 2001; Vaughan & Ormerod 2003). For many studies, however, especially retrospective analyses of existing data sets or the use of detailed standard surveys

such as RHS, the potential to implement such measures is greatly restricted. It may then be necessary to use some form of statistical data reduction to stabilise subsequent analyses. The cost of doing so is an inability to describe the relationships between individual independent variables and the dependent one, and also interactions between independent variables. Considering the potential severity of the statistical problems that may otherwise occur, however, this penalty may be considered to be of secondary importance in many studies.

Variable clustering methods may prove very useful for statistical data reduction, especially when ease of interpretation is a priority (Table 4). This was clearly illustrated by the RHS data. The overall PCA produced only three meaningful axes, yet for each one, almost all of the variables were significant contributors. Whilst it proved straightforward to assign a broad description to each one, the complexity of the axes made it virtually impossible to appreciate the full range of habitat changes they described. It was for this reason that odds ratios were not calculated for the overall PCA models as they were for variable-cluster PCA ones (Table 3b). It would have been simple to calculate an equivalent table, but it would have been virtually impossible to gain a thorough understanding of what the changes in a PC score composed of >80 variables represented. Without clear descriptions, the value of interpreting a model, rather than using it solely as a predictive tool, is limited.

RHS variable clusters all described easily understood habitat characteristics or trends. Whilst this was enhanced by the subjective decisions made about where to split the clusters, the basic cluster solution was based entirely upon the data set. No secondary analyses were required to explain the clusters, unlike conventional PCA (Peres-Neto

Table 4. A comparison of conventional PCA and variable clustering (using PCA or other scoring methods) for statistical data reduction in ecology.

	PCA	Variable clustering
Simplicity of method	Simple	Varies with method used: may be simpler or more complex than PCA
Interpretability of output	Requires secondary analyses to identify the number of relevant PCs and which variables contribute to each PC Interpreting PCs, especially beyond PC1, often challenging	No secondary analyses required. With a good clustering solution, all variables in a cluster measure the same underlying factor.
Objectivity	Objective. Hence easily reproducible	Varies with method: from relatively objective methods based upon PCA (e.g. PROC VARCLUS in SAS), to entirely subjective grouping of variables.
Practicality	All variables need to be recorded in future to calculate PCs.	Only constituent variables of clusters actually used for analyses need to be recorded for future applications. Scores for small clusters using simple cluster weightings can be calculated very quickly and easily.
Collinearity	Eliminated from a data set	Reduced, because correlated variables are grouped together, but not eliminated
Efficiency	More efficient at summarising overall variance than variable clustering using PCA for cluster scores. Fewer PCs may suffice for more variable clusters.	

Versatility	Optimised for continuous variables.	Can generalise to any data type by adopting different similarity/distance measures and different algorithms for forming and scoring clusters
Other characteristics		More informative about the data (e.g. examining clustering trees) than PCA Different clustering methods may lead to a different clustering solution.

Sources: Cureton & D'Agostino (1983); Harrell *et al.* (1984); Marshall *et al.* (1994); SAS Institute (1999); Peres-Neto *et al.* (2003)

et al. 2003). The magnitudes of eigenvectors used in calculating cluster scores were homogenous compared to those in the overall PCs, consistent with all variables contributing to the same underlying factor. It is in such situations that the interpretability benefits of clustering variables are maximised (Cureton & D'Agostino 1983). Manual clustering of the type used here, generating clustering trees, afforded an additional benefit relative to overall PCA, in revealing further information about the structure of the data set.

The small size of the clusters ($\bar{x} = 4.3$ variables) ensured that it was easy to relate them back to the actual river habitat, as was evident when interpreting the river bird models. They focused upon more specific aspects of river habitat than the broad trends of the overall PCA. This may make such clusters particularly useful in smaller-scale ecological studies and for applied concerns, where such detailed factors may be of primary importance compared to large-scale biogeographical issues, where broad habitat trends may be more relevant. The selection of 23 clusters plus outlying variables was initially a concern, suggesting relatively poor data reduction. In practice, this did not prove to be a disadvantage because the habitat trends represented by each cluster made it straightforward to select a small number of river bird predictors prior to modelling. Essentially, the use of statistical data reduction made it easier to perform further subject-matter based data reduction. By selecting a small set of predictors in this way prior to modelling, the problems of statistical variable selection were avoided and the confidence intervals around model parameter estimates should be meaningful (Harrell 2001). In essence, variable clustering made it easier to propose specific hypotheses about habitat use that could then be investigated using the fitted models and odds ratios. The benefit of this specificity was evident for

goosanders, where the effects of river size – shown previously to be related to goosander distribution (Gregory *et al.* 1997) – could be investigated directly using `ChannelSize`, which proved to be a better discriminator than any of the three overall PCs were.

The second major benefit of performing variable clustering prior to PCA is the ability to eliminate variables that comprise unused clusters from future data collection (Marshall *et al.* 1994). When using a fixed survey methodology, such as RHS, this benefit would not actually be manifested. It is nevertheless interesting to note that the variable cluster PCA models used 4–29 variables to produce similar performance to the 109 variables used in the overall PCA. Such a reduction in the number of variables required could bring valuable resource savings to ecological applications.

Variable clustering methods may have valuable roles to play in ecology beyond data reduction. Where measurements of variables are unreliable or frequently involve large errors or missing values, a cluster score combining several related variables may be more reliable than using them individually (Churchill 1979). Clusters may also be useful when there is a desire to measure a latent or abstract factor (D'Agostino *et al.* 1995). This is the major reason for the popularity of scored variable clusters, forming 'tests', in the social sciences. Personality traits and opinions are often of interest in such research, but cannot be measured directly. Instead, a series of items in a questionnaire address the latent characteristic, such as depression, forming the variables in a cluster, and by combining the value of the individual items, a measure of the condition may be obtained (e.g. Lindelow, Hardy & Rodgers 1997; Pincus *et al.* 2004). In the RHS analysis, `TreeCover` may provide a better overall measure of the

‘woodiness’ of a channel: a concept that might be difficult to capture in terms only of the length of river bank with trees present or proportion of a channel that is shaded by trees. *TreeCover* attempts to encapsulate the concept, integrating other characteristics associated with trees, such as woody debris in the channel and exposed bankside roots, within a single measure.

Despite its apparent benefits, variable clustering – using PCA or other scoring methods – is not universally better than conventional PCA, having several generic disadvantages (Table 4). The creation of discrete clusters, scored with PCA, tends to be less efficient than conventional PCA at describing the overall variance in a data set (SAS Institute 1999). Although this was not examined directly here, the discriminatory ability of models using variable clusters scored with conventional PCA tended to be lower than their equivalents using overall PCs1–3, with the exception of goosander. Fortunately, this reduction was small. Variable clustering’s other major weakness relative to conventional PCA is its failure to eliminate collinearity between variables (Cureton & D’Agostino 1983). Compared to the absence of correlations between the three overall PCs, the 23 variable cluster and 11 single variable data set had a mean |Pearson r | correlation of 0.08 and a maximum of 0.48 (the same for both scoring methods). Nevertheless, by grouping together and scoring related variables, variable clustering should consistently reduce multicollinearity, even if it cannot eliminate it. The raw RHS variables had the same mean intercorrelation as the clusters, but a larger maximum ($|r| = 0.71$). Greater benefit may have been apparent if much of the obvious redundancy between RHS variables had not been removed prior to analysis (see Appendix 1.3 for details). For example, land use was only retained from the sweep-up section, but was also recorded in the spot-checks. The equivalent

land use types in the two sections of the RHS had a median Spearman's ρ of 0.75, and based on preliminary clustering of a larger set of RHS variables, always clustered together.

4.5.1 SCORING VARIABLE CLUSTERS

The analytical benefits of OS-PCA over conventional PCA for creating cluster scores were slight compared to the interpretability benefits of variable clustering as a whole relative to overall PCA. Nevertheless, across the common sandpiper, dipper and grey wagtail models considered here, and averaged over the 23 Waterways Breeding Bird Survey models, OS-PCA improved the discriminatory ability. Although the improvements were small, they were important for two reasons. First, and most importantly, they indicated that OS-PCA was more effective than conventional PCA at summarising RHS data in variable clusters. Through the optimal scaling procedure, OS-PCA seemed better able to position river sites along the latent gradients identified by the variable clusters. The second reason for highlighting the improvements of OS-PCA is that the transformations were very simple, giving the improvement in discrimination for little increase in cluster complexity.

Generically, OS-PCA's major benefit over conventional PCA is its versatility (Table 5). In essence, it is a generalised version of the latter, encompassing data on any measurement scale – including solely nominal variables – and performing non-linear PCA (Young *et al.* 1978). The greater the proportion of qualitative data, the greater the benefits are likely to be relative to conventional PCA, to which the latter is poorly suited (Hair *et al.* 1995). With RHS, 5–6 point ordinal variables still provide

Table 5. Potential benefits and disadvantages of optimally-scaled PCA.

Potential advantages of optimally-scaled PCA

- Generalises PCA to a much wider range of situations: nominal, ordinal and continuous variables (and any mixture of the three types) may be analysed
- Quantifies differences between observations that are expressed in terms of qualitative variables
- Performs non-linear PCA, including non-monotonic relationships if appropriate transformations (e.g. splines) are used
- Increased efficiency – explains more of the variance in a data set for a given number of PCs
- Optimised PC scale can resolve smaller inter-site differences, improving the discriminatory power of analyses

Potential disadvantages of optimally-scaled PCA

- More complex than PCA and potentially more difficult to interpret
- Requires much greater computational time/resources than traditional PCA
- Estimating missing values may be difficult if complex transformations are used
- In small sample sizes, and when certain categories contain few observations, the optimisation procedure may converge upon a local minimum, rather than finding the desired global optimum

Sources: Young *et al.* (1978); Young (1981); Didow *et al.* (1985); Harrell (2001)

reasonable resolution and ordinal variables were always accompanied by metric ones in the clusters. This may explain why the benefits observed were modest.

In general, the penalties for the versatility of OS-PCA are greater computational time and more complex cluster structures (Table 5). The former was evident in the RHS analysis, for which conventional PCA of the near-9000 site data set took less than 2 seconds, whilst the PRINQUAL procedure ran for up to 10 minutes with large clusters containing many variables to be transformed, such as `TreeCover`.

4.6 Conclusions

Where data reduction is required and interpretability a primary concern, variable clustering may provide major benefits for PCA. Based upon the current study, it appears to represent a significant advance for the analysis of RHS data. More precise hypotheses could be formulated and analysed than using the broad patterns described by overall PCA. At the same time, the process of clustering resulted in little loss of predictive ability (conventional PCA scoring), suggesting that one of the major shortcomings of clustering – that of reduced analytical efficiency (Table 4) – may not prove to be major weakness for RHS-based analyses.

Conventional PCA is an effective way of scoring variable clusters, maximising the within-cluster variance that is explained by a single index (Cureton & D'Agostino 1983). OS-PCA allows this process to be generalised to mixed measurement scales and to describe non-linear habitat trends with variable clustering. For RHS data, OS-PCA afforded a small increase in the efficacy of subsequent analyses for little increase in cluster complexity.

Only one of the possible methods for variable clustering was used here, being treated as an extension of PCA. Both clustering and scoring methods are diverse and can be tailored for different applications. Here the desire was to examine the clustering solutions that were proposed by the data. Alternatively, clusters could be formulated largely or entirely on subject matter knowledge, as is often the case in medicine, for the formulation of clinical indices used for disease prognosis (e.g. Harrell *et al.* 1984; Marshall *et al.* 1994; Mulla *et al.* 2003). Another option is to use a fully automated technique for clustering and scoring, such as the VARCLUS procedure in SAS, which is based on a derivative of PCA with oblique rotation of components (SAS Institute 1999). For the scaling of ordinal RHS data, monotonic transformations were used and Spearman's ρ coefficients used to measure similarity so that any form of monotonic relationships between variables would be detected. In other situations, clustering and scoring could be generalised to entertain non-monotonic relationships (Harrell 2001). The overall versatility of clustering as a data reduction method suggests a wide range of possible applications to be explored in ecology.

4.7 Acknowledgements

The Welsh survey data were collected under combined funding from the Environment Agency, Welsh Assembly and Countryside Council for Wales.

4.8 References

Anderberg, M.R. (1973) *Cluster Analysis for Applications*. Academic Press, New York.

Brewin, P.A., Buckton, S.T. & Ormerod, S.J. (1998) River habitat surveys and biodiversity in acid-sensitive rivers. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 501–514.

Brier, G.W. (1950) Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**, 1–3.

Buckton, S.T., Brewin, P.A., Lewis, A., Stevens, P. & Ormerod, S.J. (1998) The distribution of dippers, *Cinclus cinclus* (L.), in the acid-sensitive region of Wales, 1984–95. *Freshwater Biology*, **39**, 387–396.

Buckton, S.T. & Ormerod, S.J. (1997) Use of a new standardized habitat survey for assessing the habitat preferences and distribution of upland river birds. *Bird Study* **44**, 327–337.

Cattell, R.B. (1966) The scree test for the number of factors. *Multivariate Behavioral Research*, **1**, 245–276.

Churchill, G.A. (1979) A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, **16**, 64–73.

Cox, D.R. (1958) Two further applications of a model for binary regression.

Biometrika, **45**, 562–565.

Cureton, E.E. & D'Agostino, R.B. (1983) *Factor analysis: an applied approach*.

Lawrence Erlbaum Associates, Hillsdale, New Jersey.

D'Agostino, R.B., Belanger, A.J., Markson, E.W., Kelly-Hayes, M. & Wolf, P.A

(1995) Development of health risk appraisal functions in the presence of multiple indicators: the Framingham study nursing home institutionalization model. *Statistics in Medicine*, **14**, 1757–1770.

Derksen, S. & Keselman, H.J. (1992) Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables.

British Journal of Mathematical and Statistical Psychology, **45**, 265–282.

Didow, N.M., Keller, K.L., Barksdale, H.C. & Franke, G.R. (1985) Improving measure quality by alternating least squares optimal scaling. *Journal of Marketing Research*, **22**, 30–40.

Efron, B. (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, **78**, 316–331.

Fox, P.J.A., Naura, M. & Scarlett, P. (1998) An account of the derivation and testing of a standard field method, River Habitat Survey. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 455–475.

Franklin, S.B., Gibson, D.J., Robertson, P.A., Pohlmann, J.T. & Fralish, J.S. (1995)

Parallel analysis – a method for determining significant principal components.

Journal of Vegetation Science, **6**, 99–106.

Freedman, L.S. & Pee, D. (1989) Return to a note on screening regression equations.

The American Statistician, **43**, 279–282.

Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression.

Ecology, **84**, 2809–2815.

Gregory, R.D., Carter, S.P. & Baillie, S.R. (1997) Abundance, distribution and habitat use of breeding Goosanders *Mergus merganser* and Red-breasted Mergansers *Mergus serrator* on British rivers. *Bird Study*, **44**, 1–12.

Grossman, G.D., Nickerson, D.M. & Freeman, M.C. (1991) Principal component analyses of assemblage structure data: utility of tests based on eigenvalues. *Ecology*, **72**, 321–347.

Gutzwiller, K.J. & Barrow, W.C. (2001) Bird-landscape relations in the Chihuahuan desert: coping with uncertainties about predictive models. *Ecological Applications*, **11**, 1517–1532.

Hair, J.F., Anderson, R.E., Tatham, R.L. & Black, W.C. (1995) *Multivariate Data Analysis*, 4th edn. Prentice Hall, Eaglewood Cliffs, New Jersey.

Harrell, F.E. (2001) *Regression Modeling Strategies*. Springer-Verlag, New York.

Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L. & Rosati, R.A. (1982) Evaluating the yield of medical tests. *Journal of the American Medical Association*, **247**, 2543–2546.

Harrell, F.E., Lee, K.L., Califf, R.M., Pryor, D.B. & Rosati, R.A. (1984) Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, **3**, 143–152.

Hastie, L.C., Cooksley, S.L., Scougall, F., Young, M.R., Boon, P.J. & Gaywood, M.J. (2003) Characterization of freshwater pearl mussel (*Margaritifera margaritifera*) riverine habitat using River Habitat Survey data. *Aquatic Conservation: marine and freshwater ecosystems*, **13**, 213–224.

Hosmer, D.W. & Lemeshow, S. (2000) *Applied Logistic Regression* (2nd edn). John Wiley & Sons, New York.

Jackson, D.A. (1993) Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, **74**, 2204–2214.

Jongman, R.H.G., ter Braak, C.J.F. & van Tongeren, O.F.R (1995) *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge.

Jüttner, I., Sharma, S., Dahal, B.M., Ormerod, S.J., Chimonides, P.J. & Cox, E.J.

(2003) Diatoms as indicators of stream quality in the Kathmandu Valley and Middle Hills of Nepal and India. *Freshwater Biology*, **48**, 2065–2084.

Kruskal, J.B. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1–27.

Lindelow, M., Hardy, R. & Rodgers, B. (1997) Development of a scale to measure symptoms of anxiety and depression in the general UK population: the psychiatric symptom frequency scale. *Journal of Epidemiology and Community Health*, **51**, 549–557.

Manel, S., Buckton, S.T. & Ormerod, S.J. (2000) Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan rivers. *Journal of Applied Ecology*, **37**, 756–770.

Manel, S., Dias, J.M., Buckton, S.T. & Ormerod, S.J. (1999a) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734–747.

Manel, S., Dias, J.M. & Ormerod, S.J. (1999b) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, **120**, 337–347.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921–931.

Marsden, S. & Fielding, A.H. (1999) Habitat associations of parrots on the Wallacean islands of Buru, Seram and Sumba. *Journal of Biogeography*, **26**, 439–446.

Marshall, G., Grover, F.L., Henerson, W.G. & Hammermeister, K.E. (1994) Assessment of predictive models for binary outcomes: an empirical approach using operative death from cardiac surgery. *Statistics in Medicine*, **13**, 1501–1511.

Mulla, N.F., Osher, A.P., Beeson, W.L., Kuhn, M.A. & Larsen, R.L. (2003) Scoring system to determine the need for balloon atrial septostomy for restrictive interatrial communication in infants with hypoplastic left heart syndrome. *The Journal of Heart and Lung Transplantation*, **22**, 883–888.

Naura, M. and Robinson, M. (1998) Principles of using River Habitat Survey to predict the distribution of aquatic species: an example applied to the native white-clawed crayfish *Austropotamobius pallipes*. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 515–527.

Olden, J.D & Jackson, D.A. (2000) Torturing data for the sake of generality: how valid are our regression models? *Écoscience*, **7**, 501–510.

Ormerod, S.J., Allinson, N., Hudson, D. & Tyler, S.J. (1986) The distribution of breeding dippers (*Cinclus cinclus* (L), Aves) in relation to stream biology in upland Wales. *Freshwater Biology*, **16**, 501–507.

Osborne, P.E. & Tigar, B.J. (1992) Interpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa. *Journal of Applied Ecology*, **29**, 55–62.

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. & Feinstein, A.R. (1996) A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, **49**, 1373–1379.

Peres-Neto, P.R., Jackson, D.A. & Somers, K.M. (2003) Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. *Ecology*, **84**, 2347–2363.

Pincus, T., Williams, A.C. de C., Vogel, S. & Field, A. (2004) The development and testing of the depression, anxiety and positive outlook scale (DAPOS). *Pain*, **109**, 181–188.

Raven, P.J., Fox, P., Everard, M., Holmes, N.T.H & Dawson, F.H. (1997) River Habitat Survey: a new system for classifying rivers according to their habitat quality. *Freshwater Quality: Defining the Indefinable?* (eds P.J. Boon & D.L. Howell), pp. 215–234. The Stationery Office, Edinburgh.

Round, P.D. & Moss, M. (1984) The waterbird populations of three Welsh rivers.

Bird Study, **31**, 61–68.

SAS Institute Inc. (1999) *SAS OnlineDoc® Version 8*. Cary, NC, USA.

Steyerberg, E.W., Eijkemans, M.J.C., Harrell, F.E. & Habbema, J.D.F. (2000)

Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in Medicine*, **19**, 1059–1079.

Tabachnick, B.G. & Fidell, L.S. (2001) *Using Multivariate Statistics* (4th edn). Allyn & Bacon, Boston, USA.

Tyler, S.J. (2003) Grey wagtail. *The Migration Atlas: movements of the birds of Britain and Ireland*. (eds C.V. Wernham, M.P. Toms, J.H. Marchant, J.A. Clark, G.M. Siriwardena & S.R. Baillie), pp. 480–482. T. & A.D. Poyser, London.

Tyler, S.J. & Ormerod, S.J. (1991) The influence of stream acidification and riparian land-use on the breeding biology of Grey wagtail *Motacilla cinerea* in Wales. *Ibis*, **133**, 286–292.

Tyler, S.J. & Ormerod, S.J. (1994) *The Dippers*. T. & A.D. Poyser, London.

Vaughan, I.P. & Ormerod, S.J. (2003) Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conservation Biology*, **17**, 1601–1611.

Williams, B.K. & Titus, K. (1988) Assessment of sampling stability of ecological applications of discriminant analysis. *Ecology*, **69**, 1275–1285.

Yalden, D.W. (1986) The habitat and activity of Common Sandpiper *Actitis hypoleucos* breeding by upland rivers. *Bird Study*, **33**, 214–222.

Young, F.W. (1981) Quantitative analysis of qualitative data. *Psychometrika*, **46**, 357–388.

Young, F.W., Takane, Y. & de Leeuw, J. (1978) The principal components of mixed measurement level multivariate data: an alternating least squares method with optimal scaling features. *Psychometrika*, **43**, 279–281.

Chapter 5

**River birds and their habitats: linking the Waterways
Breeding Bird Survey and the UK River Habitat Survey**

5.1 Summary

1. The Waterways Breeding Bird Survey (WBBS) was launched in 1998 as a possible successor to the Waterways Bird Survey. In addition to detecting spatial and temporal patterns among river birds, the WBBS could allow improved understanding of their habitat needs both to guide river management and to help interpret population change. These aims could be advanced considerably if the WBBS was linked to the Environment Agency's River Habitat Survey (RHS), currently the most advanced standard method for river habitat recording and analysis.
2. The links between river bird presence/absence and RHS data were investigated using data from the first three years of the WBBS (1998–2000). Regression models related 28 WBBS species to RHS data that were reduced to easily interpreted habitat indices by variable clustering. Small numbers of indices were selected for each species' model, based on postulated habitat preferences, to avoid the problems associated with statistical variable selection and to permit more valid tests of predictive ability.
3. The predictive ability of river bird models varied widely. Models for waders were often less accurate than those for species directly reliant upon the channel, such as waterfowl. RHS predictors describing channel characteristics (e.g. flow types or vegetation) tended to be more effective in predicting bird distributions than predictors describing floodplain habitats (e.g. neighbouring wetlands). A combination of odds ratios and RHS habitat indices allowed the likelihood of each species' occurrence to be related directly to the nature of the river habitat, and results are given for all 28 species. Models using map-derived variables had

similar predictive abilities, but lacked the clear ecological interpretation of RHS models.

4. Further work is required to specify models more completely and to test their transportability before the full potential of RHS in predicting bird distributions can be assessed. Nevertheless, RHS appears to provide a valuable basis for predicting the occurrences of a wide range of bird species and for describing their habitat preferences in a manner suited to conservation applications, including population monitoring via the WBBS.

Key words: bioindicators; discriminatory ability; floodplain habitats; habitat indices; hierarchical partitioning; map-derived predictors; odds ratios

5.2 Introduction

The Waterways Breeding Bird Survey (WBBS) was launched by the British Trust for Ornithology (BTO) in 1998 as a possible successor to the Waterways Bird Survey (WBS), which has been the standard monitoring programme for UK river birds since 1974 (Carter 1989). WBBS has two potential advantages over the WBS. The first is its efficiency: by employing transect counts, rather than territory mapping, WBBS requires fewer repeat visits to a site during a field season and less data processing. This would allow a larger number of randomly selected sites to be monitored, improving geographical coverage and providing population indices more representative of the whole UK (Marchant *et al.* 2002). These were major arguments in favour of the Breeding Bird Survey over its predecessor, the Common Birds Census (Greenwood *et al.* 1995).

The second potential benefit of WBBS over WBS is its integration with River Habitat Survey (RHS). Developed by the Environment Agency, this is a standardised method for describing river habitat across the UK, and is currently the most comprehensive and best-developed tool for this purpose (Raven *et al.* 1997; Fox, Naura & Scarlett 1998). RHS may thus provide greater insight into the habitat requirements of river birds than the simple habitat recording that accompanies WBS (Langston, Marchant & Gregory 1997). Similarly, it may be more sensitive to, and describe more precisely, changes in river habitats associated with changes in river bird populations identified by the WBBS.

Surprisingly little work has so far been carried out to examine the efficacy of RHS for describing the habitat requirements of river birds or, equally, how well RHS can

predict river bird distributions. Initial studies were encouraging, however, linking river birds and RHS in upland Wales and the Himalayan mountains (Buckton & Ormerod 1997; Manel *et al.* 1999a, 1999b). Here, the relationships between RHS and river birds are analysed from the first three years of the WBBS (1998–2000). In essence, this represents a development from earlier studies in several key respects (Buckton & Ormerod 1997). Not only does this study extend across the UK and to a wider range of riparian birds, but it also advances the analytical methods used (Chapters 3 & 4). The aims are twofold: to assess the potential to link river birds to RHS at a UK-wide extent and to quantify the relationships between river birds and aspects of habitat known, or strongly suspected, to affect their distributions. The former is essentially concerned with the predictive ability of models, whilst the latter is essentially explanatory.

5.3 Methods

5.3.1 WATERWAYS BREEDING BIRD SURVEY

Between 1998 and 2000, 144 randomly selected river and canal sites across the UK were surveyed in at least one breeding season by volunteer fieldworkers (Marchant *et al.* 2002). For each site, 1–10 contiguous 500m transects alongside the waterway were surveyed twice during the breeding season and all birds seen or heard recorded separately for each section into three distance classes from the channel (0–25m, 25–100m, >100m), or as ‘in-flight’.

Data were supplied by the BTO for 28 species in the three distance categories (Table 1). The data for individual species were pooled within 500m transects and a species

Table 1. Species considered in the analysis of the WBBS, along with their prevalence in the final 161 site data set.

Species		Prevalence / %
Mute swan	<i>Cygnus olor</i>	17
Greylag goose	<i>Anser anser</i>	17
Canada goose	<i>Branta canadensis</i>	17
Shelduck	<i>Tadorna tadorna</i>	37
Mandarin duck	<i>Aix galericulata</i>	23
Gadwall	<i>Anas strepera</i>	9
Teal	<i>Anas crecca</i>	14
Mallard	<i>Anas platyrhynchos</i>	61
Shoveler	<i>Anas clypeata</i>	22
Pochard	<i>Aythya ferina</i>	20
Goosander	<i>Mergus merganser</i>	9
Moorhen	<i>Gallinula chloropus</i>	40
Coot	<i>Fulica atra</i>	25
Oystercatcher	<i>Haematopus ostralegus</i>	16
Ringed plover	<i>Charadrius hiaticula</i>	20
Lapwing	<i>Vanellus vanellus</i>	17
Snipe	<i>Gallinago gallinago</i>	25
Curlew	<i>Numenius arquata</i>	24
Common sandpiper	<i>Actitis hypoleucos</i>	17
Kingfisher	<i>Alcedo atthis</i>	15
Sand martin	<i>Riparia riparia</i>	37
Yellow wagtail	<i>Motacilla flava</i>	38
Grey wagtail	<i>Motacilla cinerea</i>	34
Pied wagtail	<i>Motacilla alba</i>	34
Dipper	<i>Cinclus cinclus</i>	20
Sedge warbler	<i>Acrocephalus schoenobaenus</i>	17
Reed warbler	<i>Acrocephalus scirpaceus</i>	16
Reed bunting	<i>Emberiza schoeniclus</i>	18

was recorded as present if it was detected in any distance category at any stage over the three years. Species were otherwise considered absent.

5.3.2 RIVER HABITAT SURVEY

RHS is a standard method for describing UK waterways over a 500m reach for applications including river classification and habitat quality assessments (Raven *et al.* 1998). The majority of the data are collected in the field during a two-stage process. In the first part, aspects of the channel structure, hydrology, vegetation and banks are recorded at 10 equidistantly spaced 'spot-checks'. These generally record data in the form of the predominant type of feature from a list of pre-defined categories, such as predominant bank material from 14 possible types (e.g. bedrock, earth, sand or clay). For the current analysis, each category was treated as an individual variable, representing the number of spot checks within the 500m at which the feature was recorded. In the second stage of the survey – the so-called 'sweep-up' – the overall extent of different channel and bank features, and various land uses within 50m of the channel are described as one of three categories: absent, present or extensive ($\geq 33\%$ of the 500m reach). These ordinal variables were coded as 0/1/2 for the analysis and the maximum value used when the variable was recorded on both banks. Aspects of tree cover, channel dimensions and anthropogenic modifications are also recorded during the sweep-up. For a detailed description of RHS see Raven *et al.* (1997).

RHS was carried out at 109 of the 144 WBBS sites between 1998 and 2000, covering more than 600 WBBS transects. Neighbouring 500m transects on a waterway were unlikely to be independent for the purposes of subsequent statistical analysis. In

addition, the agreement between the boundaries of individual RHS and WBBS transects was imperfect. Consequently, where five or fewer transects were surveyed on a waterway, a single, central one was selected for analysis. If six or more were surveyed, the transects at the extreme upstream and downstream ends were selected, ensuring a minimum inter-site spacing of 2km. The resulting sample size was 161.

5.3.3 INDICES OF RIVER HABITAT

RHS data are challenging to analyse because of high dimensionality (109 variables in this analysis), the limited information conveyed by the three-point ordinal sweep-up variables and multicollinearity between groups of variables. Variable clustering was used to tackle these problems, summarising the data set as a series of indices created in a two-stage process. In the first stage, related variables were placed into discrete clusters by generating a Spearman's ρ correlation matrix for all 109 variables, converting them to distances as $1-|\rho|$ and using the matrix as a basis for a cluster analysis in SAS PROC CLUSTER (SAS Institute Inc.). Twenty-three clusters and 11 outlying variables were identified, as described previously (Chapter 4). In the second stage, clusters were scored so that a single index could represent all of the constituent variables within a cluster. Where clusters consisted only of spot-check variables, PCA was performed on the clustered variables and the first component used as the cluster score. Where ordinal sweep-up variables were present, an extension of PCA was used that was designed for qualitative/semi-quantitative data (Young, Takane & de Leeuw 1978).

To increase the reliability of the variable clustering process and to maximise the representativeness of the resulting habitat indices, 9069 sites from across the UK –

including the WBBS transects – were extracted from the national RHS database (version 3.3; Environment Agency 2002). For the current analysis, 12 variable clusters and three individual variables were used (Table 2; see Table 1 in Chapter 4 for full cluster structures).

5.3.4 RIVER BIRD-HABITAT MODELS

Logistic regression was used to relate the likelihood of individual species' occurrences to river habitat. Models were fitted using the LOGISTIC procedure in SAS and linearity assumptions checked as recommended by Hosmer & Lemeshow (2000) along with partial residual plots (Landwehr, Pregibon & Shoemaker 1984). The Pearson residuals were checked for spatial autocorrelation. Correlograms using Moran's *I* statistic were calculated for inter-site distances of up to 200km using 20km intervals (Sokal & Oden 1978). The residuals for the Canada goose, grey wagtail and pochard models showed short-range positive autocorrelation (<40–60km) that was significant at $p=0.05$ and so caution is required in their interpretation. In particular, confidence intervals may be too narrow and estimates of optimism in prediction accuracy (see below) too small.

Typically, some form of statistical variable selection, such as a stepwise procedure, is used to select variables in regression models. These methods can be unreliable because they may be affected disproportionately by data idiosyncrasies such as outlying points and multicollinearity (Derksen & Keselman 1992). In addition, statistical variable selection tends to result in estimates of regression coefficients that are biased towards more extreme values, exaggerating the relationships between predictor variables and species' occurrences (Steyerberg, Eijkemans & Habbema

Table 2. Descriptions of the river habitat indices and individual variables used in the river bird models at their first and third quartiles amongst the WBBS sites (except where specified otherwise). Average conditions are described within a single percentile centred on these points from the 9069 site RHS data set. Following RHS conventions, the term 'present' implies occurrences over <33% of the 500m reach; 'extensive' at $\geq 33\%$.

Habitat index / variable	First quartile	Third quartile
TreeCover	Open channels with only a few isolated trees	Broad-leaved woodland present or extensive within 50m of channel. Extensive encroachment upon the channel (e.g. shading, overhanging boughs).
BankVeg	80–90% of banks covered with vegetation of two structural types e.g. some tree cover, mixed herbs and grasses	60–80% of banks covered with vegetation of a single structural type e.g. grassland, crops, fringing reeds
ChannelSize	Around 5m wide and 0.3m deep	Around 15–20m wide and 0.5–0.6m deep
Moor–pasture	Extensive improved grassland, no moorland/peat banks	Extensive moorland/heath. No improved grassland or peat banks ¹ .
Urban	No anthropogenic features	Moderate anthropogenic development: some reinforcement of the whole bank present and a single bridge over the channel.
Bars/cliffs/pools	One unvegetated point bar and one pool. No river cliffs or marginal deadwater.	One unvegetated point bar and one pool. Two earth/sand cliffs ≥ 0.5 m. Marginal deadwater present.
Wetlands	Water level around 2.5m below floodplain. Rough pasture present. No wetlands within 50m of channel.	Water level 0.8m below floodplain. Rough pasture and wetlands present ² .

ChannelVeg	Channel vegetation absent	20% of reach with emergent and amphibious vegetation. Small number of fully aquatic species – up to one spot check record.
Rocky	>50% of reach with earth banks. Exposed bedrock present. No boulders.	50% earth banks. Boulders present in channel over 50% of reach. Exposed bedrock present.
Riffles	Riffles absent	Around 10% of reach with riffles (\cong three riffles over 500m)
Glides	Glides present.	Extensive glides, laminar flow over 80% of reach
GSBanks	All banks vegetated and no underlying gravel or sand.	10% of banks gravel or sand. Bare bank face at one or more spots checks
No perceptible flow	River actively flowing throughout 500m reach (unless dried up)	50% of reach without perceptible water movement
Adjacent open water	No open water within 50m of channel	Open water present within 50m of channel.
Tilled land use	No tilled land within 50m of channel.	Tilled land adjacent to $\geq 33\%$ of reach

¹80th percentile.

²90th percentile.

1999). Variables were therefore chosen prior to model development using the published literature concerning species' habitat preferences (Cramp & Simmons 1977, 1980, 1983; Cramp 1985, 1988, 1992; Cramp & Perrins 1994). Regression models were then fitted directly to the proposed variables with no further variable selection. The total number of variables in a model was restricted so that there were approximately 10 observations of the rarer event (presence or absence) for every variable in the model, as recommended by Harrell (2001). Whilst important predictors may have been overlooked, biases in coefficient estimates and confidence intervals should have been minimised, permitting more accurate quantification of the links between river birds and habitat, and a more valid overall assessment of RHS. This strategy also avoided the potential problems of performing significance testing in the presence of spatial autocorrelation, where the 'p-values' for different predictors may be artificially low (Legendre 1993).

To help to place the RHS models in context, equivalent models were built for the 28 river birds using map-derived variables. Three predictors were used – altitude, slope and geographical location – to match the most frequent number of predictors used in RHS models. Geographical location was described by the first principal component of the National Grid Reference Northings and Eastings of the sites, explaining 77% of the variance that was within the two coordinates. Results are reported for all 28 models and also just for the 23 species whose RHS models used three or more variables.

The predication accuracy of the 56 models was tested using bootstrapping. Efron's (1983) bootstrapping method was used to estimate the extent to which models

overfitted the training data, in terms of both discrimination and calibration. The former considers how accurately a model separates occupied from unoccupied waterways and was assessed using the concordance index of Harrell *et al.* (1982). This is the fundamental component of predictive ability (Harrell *et al.* 1984) and so was the primary concern of model evaluation. The *c*-index equates to the area under the receiver operating characteristic curve statistic, and describes the probability that if two sites are selected at random – one occupied, the other unoccupied – the model would correctly assign the higher likelihood of occurrence to the former (Harrell *et al.* 1982). The *c*-index is one of the few statistics that can be meaningfully compared between different species and different studies, being independent of species' prevalence (c.f. overall prediction success, kappa or goodness-of-fit measures; Miller, Hui & Tierney 1991; McPherson *et al.*, in press). Here, it allowed meaningful comparison amongst the 28 species and allows this study to form a benchmark against which future ones can be compared. Henceforth, c_{train} denotes discrimination in the training data, whilst c_{boot} , the bootstrap-corrected estimate of discrimination.

Calibration describes the accuracy of the probability labels. As the models were not applied to new data, the actual calibration accuracy was of limited relevance.

However, when using bootstrapping, calibration tests can be a sensitive way of identifying overfitting – manifested as exaggerated probabilities towards the extremes of the probability scale, underestimating occurrence at low probabilities and overestimating it at high probabilities (Steyerberg *et al.* 2000). A logistic regression model was fitted between the predicted probabilities, logit transformed, and observed presence/absence, with the degree to which the slope coefficient was below unity indicating the degree of overfitting (Cox 1958; Steyerberg *et al.* 2000). Only the

bootstrap-estimated calibration slope is reported, as the slope is always unity in the data set to which the model is calibrated.

Two methods were used to interpret the RHS-based models. First, odds ratios were calculated to quantify how changes in individual variables, holding the other variables in a model constant, affected the likelihood of a species occurring on a waterway. Odds ratios were calculated between the two values of a variable described in Table 2, rather than the default of an increase of one unit along a predictor's scale, so as to maximise interpretability. Secondly, a hierarchical partitioning protocol (Chevan & Sutherland 1991) was adopted to estimate the independent contribution that each predictor made to the overall discriminatory ability of a model by tackling the confounding effect of multicollinearity. Although collinearity was low in the RHS data following variable clustering, most cluster scores were correlated up to $|r| = 0.20$, and *Rifflés* and *Rocky* $r = 0.51$. Hierarchical partitioning averaged the increase in discrimination observed when a particular variable was included into a model, over the simpler model without it, across all possible model pairings including and excluding the variable (Chevan & Sutherland 1991). This is achieved with hierarchies of complexity, from a model containing only the predictor of interest compared to the intercept-only model, to the full model compared to a model lacking only the variable of interest. The results are given as the percentage independent contribution given by each of the predictors in a model to the overall discrimination. Full details of the process are given by Chevan & Sutherland (1991) and Mac Nally (2000).

5.4 Results

Between two and six variables were selected for the RHS-based models (Table 3), according to species' prevalence in the data set (Table 1). Models showed moderate overall predictive ability ($\overline{c_{train}} = 0.72$), varying widely from little improvement over chance (e.g. coot $c_{train} = 0.61$) up to $c_{train} = 0.85$ for sedge warbler (Table 3). The estimated overfitting in c was small ($\bar{x} = 0.03$), resulting in a mean bootstrapped discrimination ability of 0.69 (range = 0.55–0.84). For the majority of models, the calibration slope also indicated little overfitting (slopes > 0.80).

Map-derived models had similar predictive performance to RHS-based models, with $\overline{c_{train}} = 0.71$ and $\overline{c_{boot}} = 0.69$ (Table 3). When the five less-prevalent species were removed, map-derived models were marginally more discriminating than RHS models, producing better predictive performance in 15 out of 23 models and improving c by an average of 0.01. The accuracy of the two types of model for each species were broadly correlated (c_{boot} $r = 0.58$, $p = 0.0017$, $n = 28$ or $r = 0.61$, $p = 0.0018$, $n = 23$).

The accuracy of RHS-based models showed some slight relationships to species' taxonomy and/or ecology (Table 3). Models for the six wader species tended to have relatively poor discriminatory abilities ($\overline{c_{boot}} = 0.64$), whereas models for the 13 waterfowl species tended to perform well ($\overline{c_{boot}} = 0.71$), with the particular exceptions of coot and greylag goose, where $c_{boot} < 0.60$ (Table 3). Models for the three wagtail species were relatively poor discriminators of suitable habitat ($\overline{c_{boot}} = 0.64$), contrasting with the relatively good accuracy for two of the three small

Table 3. The predictive performance of the river bird models built from the WBBS data set, using both RHS data and map-derived variables. Discriminatory ability is estimated both in the training data and via bootstrapping. For RHS-based models, the list of predictors selected is given, along with their odds ratios ($\pm 95\%$ confidence intervals) calculated over the ranges indicated in Table 2, and their independent contributions to the overall discriminatory abilities, estimated via hierarchical partitioning.

Species	RHS-based models						Map-derived models	
	C_{train}	C_{boot}	Calibration slope	Variables	Odds ratio	Contribution to $c / \%$	C_{train}	C_{boot}
Mute swan	0.81	0.78	0.90	Urban	1.10 (0.44–2.74)	5.3	0.83	0.82
				ChannelVeg	2.65 (1.50–4.67)	71.4		
				No perceptible flow	2.21 (1.21–4.04)	23.3		
Greylag goose	0.63	0.57	0.60	BankVeg	1.00 (0.50–1.97)	12.4	0.67	0.63
				Urban	2.08 (0.85–5.08)	51.2		
				No perceptible flow	0.62 (0.28–1.40)	36.4		
Canada goose	0.75	0.72	0.90	BankVeg	3.33 (1.54–7.21)	65.5	0.81	0.79
				Moor-pasture	0.01 (0.00–15.37)	31.9		
				Urban	1.06 (0.44–2.55)	2.6		
Shelduck	0.71	0.69	0.87	TreeCover	3.79 (1.91–7.92)	70.2	0.65	0.65
				ChannelSize	1.40 (0.97–2.03)	10.1		
				Adjacent open water	1.31 (0.47–3.65)	0.8		
				No perceptible flow	0.64 (0.32–1.29)	18.9		
Mandarin duck	0.77	0.74	0.89	TreeCover	0.18 (0.08–0.40)	61.6	0.68	0.66
				Urban	0.52 (0.22–1.26)	26.3		
				ChannelVeg	0.61 (0.32–1.18)	9.8		
				No perceptible flow	0.92 (0.46–1.85)	2.4		
Gadwall	0.72	0.70	0.99	TreeCover	5.43 (1.35–21.94)	98.4	0.60	0.54
				No perceptible flow	0.97 (0.28–3.35)	1.6		

Teal	0.84	0.83	0.84	ChannelVeg	3.57 (1.90–6.70)	76.9	0.77	0.74
				No perceptible flow	1.74 (0.94–3.21)	23.1		
Mallard	0.80	0.79	0.88	BankVeg	0.80 (0.41–1.56)	2.3	0.81	0.81
				Urban	1.97 (0.88–4.40)	16.6		
				ChannelVeg	0.95 (0.47–1.92)	19.5		
				Rocky	0.26 (0.14–0.49)	45.9		
				Adjacent open water	4.24 (0.61–29.71)	8.7		
				No perceptible flow	1.60 (0.70–3.64)	7.0		
Shoveler	0.71	0.67	0.82	TreeCover	6.42 (1.99–20.71)	47.2	0.74	0.73
				BankVeg	1.01 (0.42–2.41)	17.6		
				ChannelVeg	1.28 (0.69–2.38)	9.9		
				No perceptible flow	3.03 (1.55–5.95)	25.3		
Pochard	0.77	0.75	0.89	BankVeg	0.27 (0.13–0.60)	23.5	0.79	0.77
				ChannelVeg	4.49 (2.19–9.20)	67.3		
				No perceptible flow	1.43 (0.76–2.69)	9.2		
Goosander	0.73	0.70	0.92	TreeCover	4.50 (1.23–16.47)	53.3	0.70	0.62
				ChannelSize	1.62 (1.01–2.60)	46.7		
Moorhen	0.77	0.74	0.83	TreeCover	3.48 (1.54–7.87)	19.2	0.83	0.81
				Urban	1.61 (0.78–3.31)	16.7		
				ChannelVeg	2.78 (1.57–4.94)	46.9		
				No perceptible flow	2.17 (1.15–4.07)	15.7		
				Adjacent open water	1.15 (0.42–3.14)	1.5		
Coot	0.61	0.55	0.65	TreeCover	0.70 (0.30–1.64)	30.9	0.64	0.61
				BankVeg	0.79 (0.38–1.67)	14.4		
				ChannelSize	1.42 (1.00–2.02)	64.2		
				ChannelVeg	0.89 (0.49–1.61)	-9.5		
Oystercatcher	0.67	0.62	0.72	Moor-pasture	0.67 (0.27–1.67)	7.6	0.79	0.78
				Bars/cliff/pools	1.44 (0.77–2.68)	24.0		
				Wetlands	1.66 (0.95–2.90)	68.4		

Ringed plover	0.64	0.61	0.80	BankVeg	0.76 (0.42–1.37)	26.8	0.57	0.53
				Bars/cliffs/pools	1.54 (0.88–2.69)	42.3		
				Wetlands	0.84 (0.68–1.05)	30.9		
Lapwing	0.63	0.59	0.64	BankVeg	1.55 (0.77–3.13)	32.6	0.65	0.60
				Moor-pasture	0.18 (0.02–1.58)	63.6		
				Tilled land	0.72 (0.24–2.14)	3.9		
Snipe	0.68	0.64	0.66	BankVeg	1.72 (0.91–3.23)	15.3	0.75	0.73
				Moor-pasture	0.07 (0.01–0.77)	63.1		
				Wetlands	0.95 (0.76–1.20)	8.0		
				Tilled land	0.43 (0.16–1.18)	13.6		
Curlew	0.75	0.72	0.68	TreeCover	0.31 (0.14–0.66)	47.1	0.79	0.77
				Moor-pasture	0.94 (0.78–1.13)	0.5		
				Wetlands	1.24 (0.88–1.76)	14.0		
				Tilled land	0.14 (0.04–0.52)	38.3		
Common sandpiper	0.71	0.68	0.89	BankVeg	2.89 (1.35–6.18)	60.7	0.75	0.72
				ChannelSize	1.37 (0.96–1.98)	33.7		
				Bars/cliffs/pools	1.29 (0.69–2.41)	5.6		
Kingfisher	0.64	0.61	0.86	ChannelSize	1.36 (0.95–1.96)	82.4	0.64	0.60
				Bars/cliffs/pools	1.21 (0.64–2.32)	17.6		
Sand martin	0.70	0.66	0.64	TreeCover	0.97 (0.48–1.93)	4.9	0.64	0.62
				Moor-pasture	0.10 (0.01–0.98)	34.1		
				ChannelSize	1.16 (0.82–1.64)	10.2		
				Bars/cliffs/pools	1.41 (0.84–2.36)	8.4		
				Glides	2.17 (1.06–4.46)	39.3		
				GSBanks	1.20 (0.88–1.64)	3.2		

Yellow wagtail	0.64	0.59	0.66	BankVeg	1.76 (0.88–3.52)	46.4	0.56	0.54
				Moor-pasture	0.49 (0.16–1.54)	19.2		
				Wetlands	1.02 (0.83–1.26)	10.2		
				Tilled land	1.83 (0.78–4.30)	24.2		
Grey wagtail	0.70	0.64	0.81	TreeCover	1.15 (0.60–2.20)	11.4	0.62	0.59
				Urban	2.88 (1.43–5.81)	71.1		
				Rocky	0.84 (0.54–1.30)	2.9		
				Riffles	1.20 (0.83–1.74)	14.6		
Pied wagtail	0.72	0.67	0.74	TreeCover	0.43 (0.18–1.02)	40.7	0.58	0.53
				BankVeg	1.36 (0.67–2.74)	25.5		
				Moor-pasture	0.31 (0.09–1.04)	28.6		
				Bars/cliffs/pools	1.13 (0.67–1.91)	3.2		
				Urban	1.17 (0.58–2.37)	2.0		
Dipper	0.81	0.79	0.87	TreeCover	2.33 (0.97–5.58)	11.8	0.69	0.65
				Rocky	1.82 (1.16–2.85)	39.1		
				Riffles	2.19 (1.43–3.36)	49.1		
Sedge warbler	0.85	0.84	0.98	BankVeg	0.57 (0.24–1.33)	20.5	0.80	0.76
				ChannelVeg	4.21 (2.26–7.86)	79.5		
Reed warbler	0.71	0.69	0.91	BankVeg	1.32 (0.58–3.03)	22.9	0.86	0.84
				ChannelVeg	1.79 (0.98–3.28)	54.0		
				No perceptible flow	2.05 (1.15–3.65)	23.1		
Reed bunting	0.78	0.76	0.93	BankVeg	2.90 (1.22–6.91)	36.8	0.76	0.74
				ChannelVeg	2.14 (1.20–3.82)	47.9		
				Wetlands	0.79 (0.62–1.00)	15.2		

passerines utilising fringing wetland vegetation (reed and sedge warblers, and reed bunting). These patterns meant that models for lowland river species tended to be more accurate than those for upland rivers, due to the numerical dominance of waterfowl species in the former, and waders in the latter. The accuracy of models for more specialised river species (common sandpiper, dipper and grey wagtail) was variable. They were frequently poorer than models for species utilising fringing vegetation (e.g. sedge warbler) or slow-flowing waterways as alternatives to stillwaters (most waterfowl), but better than models for species utilising habitats in the wider floodplain (most waders).

The majority of species-habitat relationships were modelled in the expected direction (Table 3). Where this was not the case, both the odds ratios and hierarchical partitioning results usually indicated that only a weak effect was modelled, such as the apparently negative relationship between grey wagtails and rocky, upland streams (Table 3). An exception was the unexpected negative relationship between Mandarin duck occurrence and `TreeCover`, which afforded the majority of the model's discriminatory ability (Table 3).

5.5 Discussion

Birds are an important component of river biodiversity, yet their habitat preferences have only rarely been the focus of quantitative study (Rushton, Hill & Carter 1994). In the UK, despite the relatively detailed qualitative knowledge of species' habitat requirements (e.g. Snow & Perrins 1998), the number of quantitative studies has been limited and they have generally focused upon small numbers of species over extents ranging from regional studies (e.g. Buckton & Ormerod 1997) down to one or two

catchments (e.g. Yalden 1986). National studies have usually focused upon the links between river birds and map-derived features such as geographical position, geology, altitude and slope (e.g. Marchant & Hyde 1980; Hill 1991). There appears to have been only one previous UK-wide study of the links between birds and descriptions of river habitats recorded in the field, which was based upon the simple habitat recording used by the WBS (Rushton *et al.* 1994).

The current study therefore provides for the first time detailed habitat data from waterways across the UK that are related to a wide range of river birds. As the analysis forms part of an evaluation of the WBBS in its pilot form, the over-riding concern was to evaluate how well it linked with RHS. Specifically, the aims were to assess the efficacy with which RHS described habitat characteristics related to species' occurrences and in turn the ability to predict species' distributions from RHS data. In addition, the practicality of RHS as a tool for river bird conservation was also important.

5.5.1 PRAGMATIC PROPERTIES OF RHS

RHS has a range of pragmatic benefits over the bespoke habitat recording used by WBS. It is a standard methodology, helping to link species' habitat preferences to other aspects of river classification and management in a common format (Langston *et al.* 1997). The national RHS database contains over 15 000 sites from across the UK, which allows individual sites to be placed in a broader, regional or national context (Raven *et al.* 1997). RHS's major advantage, however, is in the level of detail it contains, creating the possibility of describing species' habitat preferences in much greater detail than was possible using WBS (Langston *et al.* 1997). Unfortunately, for

this same reason – the number of variables it contains – RHS can be difficult to analyse and interpret, and this is a major pragmatic barrier to its widespread use for describing species' habitat preferences. To realise its potential, some form of data reduction is required. Previously, PCA has been used to simplify RHS data for modelling species' distributions (e.g. Buckton & Ormerod 1997; Naura & Robinson 1998). Ordination was also used by Rushton *et al.* (1994) to obtain two axes that summarised WBS habitat data. The problem in these approaches is that the meaning of ordination axes is frequently unclear and secondary analyses are often required to assist their interpretation (Peres-Neto *et al.* 2003). Although overall trends in habitat may be intuitive, such as that from upland to lowland channels (e.g. Rushton *et al.* 1994; Chapter 4), it is difficult to understand the full range of changes in habitat that actually occur. Much of the detail conveyed by RHS may, in effect, be hidden when combined in ordination axes, restricting its benefits relative to WBS habitat recording. These problems could seriously limit the application of RHS for conservation objectives.

Fortunately, variable clustering appeared to be an effective solution to this problem. All of the variables within a variable cluster relate to a common habitat factor(s) (Table 2), eliminating the need for secondary interpretative analyses (Cureton & D'Agostino 1983). Clusters can address more specific aspects of habitat than ordination axes, such as the degree of tree cover or extent of vegetation in a channel, compared to broad upland-lowland differences, for example. In addition, cluster scores were easily related back to the observed habitat, frequently summed up within single sentences (Table 2). The benefits of this approach, combined with RHS, were evident when interpreting river bird models. For example, a previous national study

identified a relationship between goosander distribution and channel size (Gregory, Marchant & Carter 1997). Here, although the 95% confidence intervals around the odds ratios were large, the effect of approximately tripling river width (from 5m to 15–20m) and doubling depth (from 0.3m to 0.5–0.6m) was directly quantified, as was the effect of increasing tree cover between equally well-defined habitat descriptions (Tables 2 & 3). Similarly, the increased likelihood of moorhen occurrence between channels devoid of vegetation and those with 20% of the reach fringed with emergent vegetation could be quantified (Table 3). Information of this sort should prove much more valuable to conservation and management efforts.

5.5.2 LINKING RIVER BIRDS TO RHS

The approach to distribution modelling used in the current study was a compromise aimed at allowing unbiased effect estimation, whilst permitting the overall predictive ability of models to be assessed. The former, quantifying the links between river habitat and the likelihood of species' occurrences, was considered a priority in assessing the value of RHS for river bird studies. By avoiding statistical variable selection, relying instead upon the published literature, relatively unbiased characterisations of species-habitat relationships were possible (Harrell 2001). The penalty for this was that the highest predictive ability for each species was unlikely to be realised, as useful predictors were probably overlooked. Nevertheless, a useful assessment of predictive potential was still possible. In addition, the process of creating an hypothesis for each species – selecting habitat factors expected to influence distribution – and then testing them – measuring the predictive accuracy that was achieved – represented a systematic way of assessing the efficacy of RHS-derived predictors.

Considering overall predictive performance first, the similar performance achieved by RHS-based models compared to map-derived ones was an encouraging result. The latter are well-proven river bird predictors (e.g. Marchant & Hyde 1980; Round & Moss 1984; Ormerod, Boilstone & Tyler 1985; Hill 1991; Rushton *et al.* 1994), reflecting the important roles of physical geography and fluvial geomorphology in shaping river habitats (Richards, Johnson & Host 1996; Davies, Norris & Thoms 2000). RHS indices represent habitat factors to which birds should respond more directly and which, in some cases, could be manipulated for management. In general, models based upon such variables are expected to generalise to new locations more successfully than those using indirect, map-derived ones (Vaughan & Ormerod 2003).

It is difficult to identify a level of overall accuracy above which individual models could be considered useful. Such thresholds are arbitrary and depend upon the destined use for a model. Nevertheless, concordance indices of 0.70–0.75 have been suggested as a threshold above which models could benefit conservation (Pearce & Ferrier 2000; Elith & Burgman 2002). Further support for such figures can be gleaned from analogous systems, such as daily weather forecasting. Swets (1988) carried out a review of several such fields, finding that models in routine use typically had *c*-indices of 0.70 and above. On this basis, even though models were not optimised for prediction, 25–50% of the 28 river bird species investigated could be usefully modelled by RHS (based upon bootstrap testing).

The potential to link RHS and river birds might be expected considering the similar scale (extent) of RHS reaches to many river birds' territories and the responsiveness

of many birds to habitat structures of the types covered by RHS (Buckton & Ormerod 1997). However, it is only by testing the predictive performance of RHS-based models that such expectations can be assessed. By hierarchically partitioning the *c*-index, it was possible to examine discriminatory ability at the level of individual predictors in addition to overall accuracy – to my knowledge, the first time that hierarchical partitioning has been used in this way (c.f. overall goodness-of-fit, Mac Nally 2000; Gibson *et al.* 2004). This allows the value of different parts of the RHS to be examined.

The efficacy of RHS's descriptions of river vegetation was apparent from the discriminatory power of `TreeCover`, `BankVeg` and `ChannelVeg` in many models (Table 3). The model for coot was an exception, with all three clusters contributing to very poor overall accuracy, although the map-derived model was similarly poor, suggesting more generic problems in predicting coot distribution. Bank vegetation was of particular interest because RHS uses highly simplistic variables to describe it, using a four-point scale that describes the number of distinct structural types amongst the vegetation, rather than any detailed botanical information such as species composition, density or height (Environment Agency 1997). `BankVeg` played an important role in predicting the distribution of Canada geese, pochards, common sandpipers, sedge warblers and reed buntings (Table 3).

Different flow types are manifestations of a suite of important ecological factors within the channel, such as the substratum size, depth and velocity of flow (Padmore 1997). It was therefore encouraging that flow-related variables were effective predictors in many cases, such as 'no perceptible flow' for several waterfowl,

`Riffles` for dippers and `Glides` for sand martins. Habitat indices representing other fundamental channel characteristics were also effective, such as `ChannelSize` for goosanders and `Rocky` for mallards and dippers. The results for `Urban` were equivocal, being a valuable predictor for grey wagtails, presumably capturing their association with artificial structures such as bridges, weirs and walls – partly for nest sites (Cramp 1988) – but showing little predictive ability with most other species for which it was used (Table 3).

Consideration of some of the models with poor overall discrimination suggests that RHS may be less successful in describing some of the other aspects of river habitat. The sparse coverage of floodplain habitats has been highlighted previously as a possible weakness of RHS (McEwan, Brazier & Gordon 1997). Only distances up to 50m from the channel are considered, and the only information recorded is the land use into very general categories, such as rough pasture, improved grassland and wetlands (Environment Agency 1997). The weakness of models for predicting wader distributions may reflect this because, with the exception of common sandpiper, their associations with rivers tend to be mediated through the suitability of floodplain habitats, rather than a direct link to the river itself (Marchant & Hyde 1980). `Moor-pasture` and `Wetlands` were generally poor discriminators. The 50m corridor limit may also compromise the predictive potential of neighbouring open water as a predictor for waterfowl which use river channels as part of wider wetland systems in large river valleys (Cramp & Simmons 1977).

Another concern was the weakness of the models for common sandpiper ($c_{boot} = 0.68$) and grey wagtail ($c_{boot} = 0.64$), both of which are river specialists. In a previous study

in mid-Wales (Chapter 4) better discrimination ($c_{boot} \geq 0.73$) was obtained for both species using similar sets of RHS clusters. This suggests that RHS could be a useful predictor for them, but that the models used here need further development.

The current study therefore provides evidence that large parts of the RHS could be valuable for describing many species' habitat requirements and, equally, predicting their distributions. Before strong conclusions can be drawn, however, the transportability of models – and in turn bird-habitat relationships – to different watercourses across the UK and to different years needs testing. Nevertheless, considering that only a small number of variables were used in most models and that the data were not used to select the predictors that gave the best fitting models, the overall level of predictive performance was reasonably good. Furthermore, additional compromises were made during model development to cope with the small sample size ($n = 161$). Only linear models were fitted, providing an adequate fit for the WBBS data set, whereas ultimately species-habitat relationships are likely to show non-linearity, but may require large data sets to reliably detect and model them. Also, potential interactions between predictors were not considered, yet could be important. For example, the relationship between `Bars/cliffs/pools` and common sandpiper occurrence – and therefore its efficacy as a predictor – may be modulated by the predominant substratum: similar numbers of exposed bars may affect the likelihood of sandpiper occurrence differently depending upon whether they are shingle or mud. The predominance of stony materials comprising bars in mid-Wales compared to UK wide may explain the relative efficacy of `Bars/cliffs/pools` as a predictor there (Chapter 4) compared to the WBBS data.

5.5.3 APPLICATIONS OF RHS MODELS

The UK is relatively unusual in having well developed censusing networks for birds, removing many of the potential applications for which distribution models are used elsewhere in the world (Fielding & Haworth 1995). However, although they may not be required to estimate basic distribution data regionally or nationally, river bird models could be useful at the level of individual sites, both in the assessment of water quality and the effects of habitat management. These will become particularly pressing over the next 10–15 years as the EU Water Framework Directive (2000/60/EC) requires ‘Programmes of Measures’ that allow rivers across Europe to attain ‘good ecological status’. Water quality, river morphology and hydrology will all figure strongly.

River birds have a widely recognised potential as indicators of river quality (Ormerod & Tyler 1993). One way in which their indicator value could be further realised is illustrated by the RIVPACS system for assessing water quality using macroinvertebrates (Wright 1995). Predictive models for individual species or communities based upon the physical environment are developed at sites assumed to be unpolluted. Subsequent applications then compare the observed invertebrate community to that predicted, with absences taken as a possible indication of degraded water quality (Wright 1995). Two vital requirements for such a system are suitable organisms and a sufficiently accurate and transportable model. Dippers have particular potential to be employed in this way, being entirely reliant upon the river channel for food and largely sedentary in the UK (Ormerod & Tyler 1993). This potential has been confirmed empirically, with correlative evidence of the effects of acidification and other pollutants upon their distribution (e.g. Ormerod *et al.* 1986; Buckton *et al.* 1998; Sorace, Colombari &

Cordiner 1999), and chemicals such as heavy metals and polychlorinated biphenyls detected in their tissues and eggs (e.g. Ormerod & Tyler 1990; Ormerod, Tyler & Jüttner 2000; O'Halloran *et al.* 2003). The current study provides evidence for the second requirement, indicating that dipper distribution can be successfully predicted based upon the physical habitat: a c_{boot} of nearly 0.80 was achieved using only a simple linear model with three predictors. Models with c -indices around 0.80 tend to make meaningful predictions at the individual site level, rather than only when considered over sets of sites (Harrell 2001), and this would be vital for any form of bioindication. The results of this study also suggest other species that may be worth investigating, where c_{boot} approaches or exceeds 0.80 (Table 3).

River engineering and other forms of management can have major effects upon river bird communities (e.g. Taylor 1984; Raven 1986). The ability to predict in advance how any changes to the habitat may positively or negatively affect species' distributions could therefore be very valuable in planning the extent and form of any management intervention. Although the potential to relate RHS to habitat modifications was not considered here, due to the limited sample size and consequent restrictions upon numbers of predictors, RHS provides detailed descriptions of anthropogenic modifications (Environment Agency 1997). Numerous variables describe re-profiling and reinforcement of both channel and banks, as well as other developments and impacts, such as outfalls, litter in the channel, bridges and weirs. An important aim for subsequent studies may be to examine how these relate to species' distributions, exploring the potential for developing management support tools.

5.6 Conclusions

An analysis of data collected during the first three years of the WBBS provides evidence that:

1. RHS forms a valuable basis for predicting the distributions of a range of bird species associated with river habitat at a UK-wide level. It is most effective with species directly associated with the channel and fringing vegetation, rather than floodplain habitats.
2. Following from (1), RHS appears to capture habitat features that are associated with the likelihood of species' occurrences. In turn, this suggests that RHS may be useful for identifying changes in river habitats that are associated with changing river bird populations, as indicated by the WBBS.
3. Deriving habitat indices from RHS data by variable clustering appears to be an effective solution to some of the challenges of RHS and may have generic value for other RHS-based studies.

Further development is required, using a larger data set and independent test data, to fully assess the value of RHS for river bird conservation.

5.7 Acknowledgements

I wish to thank the numerous volunteers who collected the WBBS data over 1998–2000 and the BTO for making the data available for analysis. In particular, I am grateful to David Noble and Dave Leech for helping with the transfer of data. The Environment Agency helped to fund the 1998–2000 WBBS and RHS data collection.

5.8 References

- Buckton, S.T., Brewin, P.A., Lewis, A., Stevens, P. & Ormerod, S.J. (1998) The distribution of dippers, *Cinclus cinclus* (L.), in the acid-sensitive region of Wales, 1984–95. *Freshwater Biology*, **39**, 387–396.
- Buckton, S.T. & Ormerod, S.J. (1997) Use of a new standardized habitat survey for assessing the habitat preferences and distribution of upland river birds. *Bird Study* **44**, 327–337.
- Carter, S.P. (1989) The Waterways Bird Survey of the British Trust for Ornithology: an overview. *Regulated Rivers: Research & Management*, **4**, 191–197.
- Chevan, A. & Sutherland, M. (1991) Hierarchical partitioning. *The American Statistician*, **45**, 90–96.
- Cox, D.R. (1958) Two further applications of a model for binary regression. *Biometrika*, **45**, 562–565.
- Cramp, S. (ed) (1985) *Handbook of the Birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Volume IV. Terns to Woodpeckers*. Oxford University Press, Oxford.
- Cramp, S. (ed) (1988) *Handbook of the Birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Volume V. Tyrant Flycatchers to Thrushes*. Oxford University Press, Oxford.

Cramp, S. (ed) (1992) *Handbook of the Birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Volume VI. Warblers*. Oxford University Press, Oxford.

Cramp, S. & Simmons, K.E.L. (eds) (1977) *Handbook of the Birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Volume I. Ostrich to Ducks*. Oxford University Press, Oxford.

Cramp, S. & Simmons, K.E.L. (eds) (1980) *Handbook of the Birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Volume II. Hawks to Bustards*. Oxford University Press, Oxford.

Cramp, S. & Simmons, K.E.L. (eds) (1983) *Handbook of the Birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Volume III. Waders to Gulls*. Oxford University Press, Oxford.

Cramp, S. & Perrins 1994 (eds) (1994) *Handbook of the Birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Volume IX. Buntings and New World Warblers*. Oxford University Press, Oxford.

Cureton, E.E. & D'Agostino, R.B. (1983) *Factor analysis: an applied approach*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Davies, N.M., Norris, R.H. & Thoms, M.C. (2000) Prediction and assessment of local stream habitat features using large-scale catchment characteristics. *Freshwater Biology*, **45**, 343–3693

Derksen, S. & Keselman, H.J. (1992) Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, **45**, 265–282.

Elith, J. & Burgman, M. (2002) Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. *Predicting Species Occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 303–313. Island Press, Washington.

Efron, B. (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, **78**, 316–331.

Environment Agency (1997) *River Habitat Survey: Field Survey and Guidance Manual*. Environment Agency, Bristol.

Fielding, A.H. & Haworth, P.F. (1995) Testing the generality of bird-habitat models. *Conservation Biology*, **9**, 1466–1481.

Fox, P.J.A., Naura, M. & Scarlett, P. (1998) An account of the derivation and testing of a standard field method, River Habitat Survey. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 455–475.

Gibson, L.A., Wilson, B.A., Cahill, D.M. & Hill, J. (2004) Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology*, **41**, 213–223.

Greenwood, J.J.D., Baillie, S.R., Gregory, R.D., Peach, W.J. & Fuller, R.J. (1995) Some new approaches to conservation monitoring of British breeding birds. *Ibis*, **137** (Suppl. 1), S16–S28.

Gregory, R.D., Carter, S.P. and Baillie, S.R. (1997) Abundance, distribution and habitat use of breeding Goosanders *Mergus merganser* and Red-breasted Mergansers *Mergus serrator* on British rivers. *Bird Study*, **44**, 1–12.

Harrell, F.E. (2001) *Regression Modeling Strategies*. Springer-Verlag, New York.

Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L. & Rosati, R.A. (1982) Evaluating the yield of medical tests. *Journal of the American Medical Association*, **247**, 2543–2546.

Harrell, F.E., Lee, K.L., Califf, R.M., Pryor, D.B. & Rosati, R.A. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, **3**, 143-152.

Hill, M.O. (1991) Patterns of species distribution in Britain elucidated by canonical correspondence analysis. *Journal of Biogeography*, **18**, 247–255.

Hosmer, D.W. & Lemeshow, S. (2000) *Applied Logistic Regression* (2nd edn). John Wiley & Sons, New York.

Landwehr, J.M., Pregibon, D. & Shoemaker, A.C. (1984) Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association*, **79**, 61–71.

Langston, R.H.W., Marchant, J.H. & Gregory, R.D. (1997) Waterways Bird Survey: evaluation of population monitoring and appraisal of future requirements. *Freshwater Quality: Defining the Indefinable?* (eds P.J. Boon & D.L. Howell), pp. 282–289. The Stationery Office, Edinburgh.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659–1673.

Marchant, J.H. & Hyde, P.A. (1980) Aspects of the distribution of riparian birds on the waterways in Britain and Ireland. *Bird Study*, **27**, 183–202.

McEwen, L.J., Brazier, V. & Gordon, J.E. (1997) Evaluating the geomorphology of fresh waters: an assessment of approaches. *Freshwater Quality: Defining the Indefinable?* (eds P.J. Boon & D.L. Howell), pp. 258–281. The Stationery Office, Edinburgh.

McPherson, J.M., Jetz, W. & Rogers, D.J. (in press) The effects of species' range size on the accuracy of environmental distribution models – ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, in press.

Mac Nally, R. (2000) Regression and model-building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – 'predictive' and 'explanatory' models. *Biodiversity and Conservation*, **9**, 655–671.

Manel, S., Dias, J.M., Buckton, S.T. & Ormerod, S.J. (1999a) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734–747.

Manel, S., Dias, J.M. & Ormerod, S.J. (1999b) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, **120**, 337–347.

Marchant, J.H., Noble, D.G., Leech, D.I. & Freeman, S.N. (2002) *River Habitat Survey and Waterways Breeding Bird Survey 1998–2000: final report*. R&D Technical Report W1-043/TR, BTO Research Report 291.

Miller, M.E., Hui, S.L. & Tierney, W.M. (1991) Validation techniques for logistic regression models. *Statistics in Medicine*, **10**, 1213–1226.

Naura, M. and Robinson, M. (1998) Principles of using River Habitat Survey to predict the distribution of aquatic species: an example applied to the native white-

clawed crayfish *Austropotamobius pallipes*. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 515–527.

O'Halloran, J., Irwin, S., Harrison, S., Smiddy, P. & O'Mahony, B. (2003) Mercury and organochlorine content of Dipper *Cinclus cinclus* eggs in south-west Ireland: trends during 1990-1999. *Environmental Pollution*, **123**, 85–93.

Ormerod, S.J., Allinson, N., Hudson, D. & Tyler, S.J. (1986) The distribution of breeding dippers (*Cinclus cinclus* (L), Aves) in relation to stream biology in upland Wales. *Freshwater Biology*, **16**, 501–507.

Ormerod, S.J., Boilstone, M.A. & Tyler, S.J. (1985) Factors influencing the abundance of breeding Dippers *Cinclus cinclus* in the catchment of the river Wye, mid-Wales. *Ibis*, **127**, 332–340.

Ormerod, S.J. & Tyler, S.J. (1990) Environmental pollutants in the eggs of Welsh dippers *Cinclus cinclus* – a potential monitor of organochlorine and mercury contamination in upland rivers. *Bird Study*, **37**, 171–176.

Ormerod, S.J. & Tyler, S.J. (1993) Birds as indicators of changes in water quality. *Birds as Monitors of Environmental Change* (eds R.W. Furness & J.J.D. Greenwood) pp 179–216. Chapman & Hall, London.

Ormerod, S.J., Tyler, S.J. & Jüttner, I. (2000) Effects of point-source PCB contamination on breeding performance and post-fledging survival in the dipper *Cinclus cinclus*. *Environmental Pollution*, **109**, 505–513.

Padmore, C.L. (1997) Biotopes and their hydraulics: a method for defining the physical component of freshwater quality. *Freshwater Quality: Defining the Indefinable?* (eds P.J. Boon & D.L. Howell), pp. 251–257. The Stationery Office, Edinburgh.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.

Peres-Neto, P.R., Jackson, D.A. & Somers, K.M. (2003) Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. *Ecology*, **84**, 2347–2363.

Raven, P.J. (1986) Changes in the breeding bird population of a small clay river following flood alleviation works. *Bird Study*, **33**, 24–35.

Raven, P.J., Fox, P., Everard, M. Holmes, N.T.H & Dawson, F.H. (1997) River Habitat Survey: a new system for classifying rivers according to their habitat quality. *Freshwater Quality: Defining the Indefinable?* (eds P.J. Boon & D.L. Howell), pp. 215–234. The Stationery Office, Edinburgh.

Raven, P.J., Holmes, N.T.H., Dawson, F.H. & Everard, M. (1998) Quality assessment using River Habitat Survey data. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **8**, 477–499.

Richards, C., Johnson, L.B. & Host, G.E. (1996) Landscape-scale influences on stream habitat and biota. *Canadian Journal of Fisheries and Aquatic Science*, **53** (Suppl. 1), 295–311.

Round, P.D. & Moss, M. (1984) The waterbird populations of three Welsh rivers. *Bird Study*, **31**, 61–68.

Rushton, S.P., Hill, D. & Carter, S.P. (1994) The abundance of river corridor birds in relation to their habitat: a modelling approach. *Journal of Applied Ecology*, **31**, 313–328.

Snow, D.W. & Perrins, C.M. (eds) (1998) *The Birds of the Western Palearctic. Concise edition*. Oxford University Press, Oxford.

Sokal, R.R. & Oden, N.L. (1978) Spatial autocorrelation in biology 1. Methodology. *Biological Journal of Linnean Society*, **10**, 199–228.

Sorace, A., Colombari, P. & Cordiner, E. (1999) Bird communities and extended biotic index (EBI) in some tributaries of the Tiber river. *Aquatic Conservation: marine and freshwater ecosystems*, **9**, 279–290.

Steyerberg, E.W., Eijkemans, M.J.C. & Habbema, J.D.F. (1999) Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology*, **52**, 935–942.

Steyerberg, E.W., Eijkemans, M.J.C., Harrell, F.E. & Habbema, J.D.F. (2000) Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in Medicine*, **19**, 1059–1079.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Taylor, K. (1984) The influence of watercourse management on Moorhen breeding biology. *British Birds*, **77**, 144–148.

Vaughan, I.P. & Ormerod, S.J. (2003) Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conservation Biology*, **17**, 1601–1611.

Wright, J.F. (1995) Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology*, **20**, 181–197.

Yalden, D.W. (1986) The habitat and activity of Common Sandpiper *Actitis hypoleucos* breeding by upland rivers. *Bird Study*, **33**, 214–222.

Young, F.W., Takane, Y. & de Leeuw, J. (1978) The principal components of mixed measurement level multivariate data: an alternating least squares method with optimal scaling features. *Psychometrika*, **43**, 279–281.

Chapter 6

General discussion

6.1 Introduction

This thesis has considered three basic themes: (i) methodological developments for species distribution modelling (Chapters 2–4), (ii) the value of RHS as a basis for describing species' habitats and predicting their distributions – focusing upon river birds – (Chapters 4 & 5), and (iii) an analysis of the Waterways Breeding Bird Survey (Chapter 5). Each of these themes will be considered in turn.

6.2 Species distribution modelling

Distribution models have developed rapidly over the last 20 years and are now abundant in the ecology and conservation literature. A wide range of applications is identified for them (Table 1 in Chapter 1), suggesting great potential to benefit conservation and management. Actual applications of models are little documented, however (e.g. Ferrier *et al.* 2002), making it difficult to assess whether this *potential* translates into practice. This may simply reflect a general reporting bias against models in use, rather than during development. Conversely, it may indicate a more worrying phenomenon, whereby the upbeat assessments and/or ambitious hopes often voiced for models by their developers are not translated into practice. This would be a manifestation of the 'widening gulf' between scientists and practitioners described by Wiens (2002). For the wider benefit of distribution modelling, assessments of the value placed in models and/or their predictions by the proposed end users should be performed as a matter of urgency, both to resolve this issue and guide future model developments. The surveys of Flashpohler, Bub & Kaplin (2000) and Ormerod *et al.* (2002) provide a model.

In lieu of such assessments of distribution modelling, a series of criteria can be identified that may affect the perceived value of models to planners and practitioners (Table 1). Three of the four are essentially practical issues, which often appear to receive scant consideration in modelling studies when compared to predictive performance. This may reflect a poor understanding by model developers of practitioners' requirements (Salwasser 1986). Performance and practicality are twin themes that need to be considered in improving distribution models, and both were concerns for the modelling developments made in the current project.

Table 1. Factors that may influence the perceived value of species distribution models in practice.

Performance	Prediction accuracy, generalisability, precision of predictions (size of confidence intervals), accuracy with which species-environment relationships are quantified (e.g. species response curves)
Relevance	Does the model address an important issue? e.g. does it predict wintering habitat use when breeding habitat is the primary conservation concern? Does a model predict the correct type of dependent variable for the intended application? e.g. abundance vs. presence/absence
Credibility	Does the model have a clear ecological rationale or does it function as a 'black-box'? Does it rely upon indirect predictors or variables that directly affect distribution? Does any part of the model appear to contradict 'accepted wisdom' about a species' habitat preferences?
Practicality	Are readily collected variables used? Is the number of variables small to minimise the cost of applying the model? Is the model simple to use, requiring minimal technical expertise – perhaps integrated into special software?

6.2.1 IMPROVING THE VALUE AND UPTAKE OF DISTRIBUTION MODELS

Methodological developments to increase prediction accuracy are the most obvious ways in which distribution models could be improved. Chapter 1 (Table 2) highlighted many possible issues and Chapters 2–4 addressed some of them. Such developments continue at a rapid rate across the distribution modelling literature, but it often appears that fundamental issues (e.g. data collection) are neglected in favour of seemingly more fashionable ones such as alternative modelling algorithms (e.g. neural networks, genetic programming). Examining novel methods is undoubtedly important because of the new possibilities they create. It could be argued, however, that the novelty itself attracts much of the research effort towards them (Van Horne 2002). Progressive development of fundamental issues such as data collection (Chapter 2) and model testing (Chapter 3) may prove more profitable to distribution modelling in the long term. The latter occupies a uniquely important position, in fact, as it provides the basis upon which other methodological developments are evaluated. It is interesting that many of the studies comparing alternative modelling algorithms, for example, have done so using evaluation methods that are fundamentally flawed, such as resubstitution testing and overall prediction success (e.g. Franklin 1998). The former in particular is a serious problem, for when taken to an extreme it may only reveal which of a series of modelling methods overfits the training data to the greatest degree. The failure to address such basic issues may misguide the future development of distribution modelling.

Other fields of science, where parallel applications to species distribution modelling occur, can be productive sources of ideas for how to increase model performance: a fact confirmed throughout this thesis. Frequently, researchers in different disciplines

approach similar problems with different emphases, leading to expertise in different areas. In social sciences, variable clustering and data reduction have been important research topics for several decades and offer valuable solutions to ecological problems (e.g. Chapter 4). In medicine, much effort has been applied to testing predictive models, due to the pressure exerted by the moral aspects of prediction errors.

Geography has been valuable through the development of geographical information systems as an environment for predictive modelling and aspects of spatial statistics. Given these benefits, it is surprising how little evidence there appears to be in ecology of searching 'laterally' for such ideas, despite articles specifically making such links (e.g. Fig. 1 in Manel, Williams & Ormerod 2001). It seems likely that in future, ecological modelling will continue to benefit from techniques developed elsewhere.

Developments in pragmatic or operational issues could be as valuable to distribution modelling as improved accuracy. Increasingly accurate models could greatly benefit conservation, yet if their complexity and expense of development and use also increase, the chances of their being applied in practice is likely to decrease. It is in this respect that the findings in Chapter 4 are so important. Variable clustering is very simple, yet it transformed PCA with RHS data, making analyses more specific, creating more meaningful habitat indices and, more generically, saving the need to record eliminated variables for future modelling work. Improvements of this type in model credibility to practitioners and overall practicality should greatly benefit the value of distribution models.

In the same way as for methodological development, testing has an important role to play in encouraging the uptake of models. Scepticism amongst potential end users

should be reduced if a model is thoroughly tested in situations accurately resembling future applications. In some cases, it may be useful to test models against expert opinion – perhaps that of sceptical practitioners – to see whether models really do afford benefits. Even the way in which models are portrayed could affect their uptake: describing them as tools for *supporting* decision making, rather than implying that they could take decision making away from practitioners.

Species distribution modelling will undoubtedly continue to develop over the coming years, as the underlying rationale/theory develops, new methods appear and conservation priorities evolve. To maximise future success, developments must involve closer collaboration between all interested parties, including ecologists, model developers, planners and practitioners. Ultimately, although model development and testing is generally expensive, models are intended as pragmatic tools aiming to either increase knowledge of habitat requirements or permit predictions when resources are insufficient for detailed autecological studies. Consequently, future modelling developments should be subject to pragmatic evaluation, asking whether proposed models (and developments thereof) represent good value-for-money in addressing clearly defined aims.

6.3 An appraisal of RHS for species distribution modelling

Species distribution models could benefit the conservation of a wide range of riverine taxa in the UK. Explanatory models could help to reveal species' habitat requirements, informing river management. Predictive models could estimate distribution data for remote regions of the UK or for taxa that are difficult to survey (e.g. small mammals, many invertebrates), or be used to make initial site appraisals

where distribution data are lacking (e.g. in response to planning applications), aid habitat quality assessments or estimate the effects of alternative management strategies. The range of applications reflects both the biodiversity of rivers and the range of parties involved in their conservation. It is hoped that RHS will to a large extent be able to fulfil these modelling roles, but to do so must satisfy pragmatic concerns and prove its efficacy at a task for which it was not specifically developed.

Practical benefits of using RHS for modelling are clear. It uses a simple, repeatable and largely objective method that has been widely tested and refined over the last decade, with over 15 000 surveys now carried out across the UK. This provides a wealth of information, including a clear context for future surveys, in an easily accessible database. The categorisations used for different features are readily identifiable, facilitating the reliability of the survey, whilst making it easy to envisage the habitat described by RHS data (e.g. Chapters 4 & 5). For example, using RHS, the channel substratum at a site might be described in terms of four spots-checks of cobble substratum and six with boulder substratum, compared to an average substratum size coupled with a measure of variability. These practical benefits have led to the wide range of roles for RHS, from river description and classification to setting objectives for habitat management (e.g. Walker, Diamond & Naura 2002).

Currently, less is known about RHS's efficacy for modelling, predominantly because it is harder to prove than its practicality and fewer attempts have been made to investigate it. It is a major challenge to capture the full range of variation observed across the tens of thousands of kilometres of natural and artificial waterways in the UK within a single survey method. RHS has to incorporate both the wide range of

variation in hydrology, geomorphology and vegetation structure, and the diverse ways in which humans modify waterways. Its efficacy for applications such as river classification (e.g. Newson *et al.* 1998) and the intuitive fluvio-geomorphological and/or ecological patterns that emerge indicate that RHS is an effective descriptor, at least from a human perspective.

For distribution modelling, however, RHS is required to extend beyond its intuitive, descriptive role: it has to capture the aspects of habitat variation to which species respond and do so in a way that is meaningful to them (Beutel & Beeton 1999). Only then can RHS form a valuable basis for species distribution modelling and its full potential as a conservation tool be realised. Studies such as the current one are vital to test for this ability, where it is manifested as good predictive performance and modelled species-environment relationships that are consistent with existing ecological knowledge (where it is available). The results of Chapter 5 are encouraging in both of these respects. Many of the modelled relationships fitted published qualitative knowledge and predictive performance based upon them was often good: no attempt was made to find the ‘best fitting model at all costs’ (e.g. using stepwise regression on the full set of variable clusters).

A great deal of further work is needed before strong conclusions can be drawn about the value of RHS for species distribution modelling. Whilst a range of studies, including the present one, has provided initial evidence of RHS’s potential, they do not represent a comprehensive assessment. Relatively few taxa have been examined (mainly birds and some invertebrates), all of the studies have used small sample sizes (<200 sites and often low species’ prevalence), and testing – even where independent

data have been used (e.g. Manel *et al.* 1999a, 1999b) – has been rudimentary. Once again, appropriate model evaluation is crucial, not least because the transportability of species' relationships to the RHS also needs to be examined. Hopefully the model evaluation and variable clustering methods in this thesis will benefit future studies: these need large sample sizes and a wider range of river biota.

Aside from more thorough assessments of RHS modelling, a priority for future work is to investigate the possible weaknesses in RHS that have emerged. Concerns about coverage of the floodplain (Chapter 5) have been expressed previously by geomorphologists (e.g. McEwan, Brazier & Gordon 1997). Also, apparent weaknesses in predictors strongly believed to be related to habitat preferences need to be investigated (e.g. common sandpipers and unvegetated bars; Chapter 5). It is imperative to establish whether such problems arise from aspects of the modelling process (e.g. Type II errors resulting from small data sets), misconceptions about species' ecology or a failure by the RHS (via variable clustering in this instance) to capture habitat characteristics as well as it could.

Ultimately, to fully prove the value of RHS for species' conservation, it will be necessary to move beyond simple, correlative modelling, to identify direct, causal links to species' distributions. This will require different analytical approaches, such as comparisons of species occurrences before and after management to identify positive or negative impacts. If the 'ecological validity' of RHS could be proven in this manner, its full conservation potential could be realised, forming the basis of management prescriptions. Similarly, RHS could then form the basis of mechanistic, behavioural models, which should be more effective than static, empirical types at

predicting the effects of habitat change, such as those that would be caused by alternative management strategies (Sutherland 1996). A great deal more work is required to reach this stage, but on the basis of this and similar studies, the capacity of RHS to act in such a way seems entirely plausible.

6.4 Waterways Breeding Bird Survey and RHS

Birds make an important contribution to riverine biodiversity (Buckton 1998). They are important indicators of environmental quality and change, being easy to survey and sensitive to both the physical structure of the river environment and to pollutants (Ormerod & Tyler 1993). In the UK, the Waterways Bird Survey (WBS) reveals that river bird populations have changed dramatically in the last 30 years. Some species have shown massive declines, such as an 89% drop in yellow wagtail *Motacilla flava* numbers (1975–2000), whilst others have increased dramatically, such as a near tripling in mallard *Anas platyrhynchos* populations over the same period (Marchant 2004). These changes may in turn reflect major changes in aspects of the river environment. To understand them, and to promote river conservation in the UK, effective schemes for monitoring both birds and their habitats are required. The development of the Waterways Breeding Bird Survey (WBBS) presents a good opportunity to examine more effective alternatives to the simplistic WBS habitat recording, and the comprehensive nature of the current standard methodology – RHS – makes it an obvious candidate.

Preliminary investigations linking bird distributions to RHS are encouraging for a range of species (Buckton & Ormerod 1997; Manel *et al.* 1999a, 1999b; Chapters 4 & 5). Conversely, for others, the links so far discovered with RHS are insufficient for

practical use. Models with *c*-indices of around 0.6, for example, are performing only marginally better than chance and so are likely to provide little increase in understanding of habitat preferences or explanation behind population changes. More work is urgently required to establish the causes: poor model specification, response to factors outside RHS (e.g. floodplain features or water chemistry) or RHS acting as a poor descriptor of physical habitat, as discussed above. Similarly, greater resources – essentially numbers of training and test data – are needed to develop the preliminary models (Chapter 5) further, to provide a more thorough assessment of RHS as a support tool for WBBS. Models with discriminatory abilities of 0.75–0.80 and above, using small numbers of habitat indices with clear ecological rationales to link them to species' distributions, would make RHS very valuable for river bird conservation.

Aside from contributing to an appraisal of WBBS, the analysis of Chapter 5 was significant in its own right as the first UK-wide study relating river bird distributions to detailed habitat data. As such it represents a benchmark against which future studies can be compared, in which the developments in Chapters 3 & 4 were vital. Likelihoods of occurrence were related to easily-understood habitat trends and general evaluation statistics mean that the improvements made by future studies can be easily quantified (c.f. overall prediction success or goodness-of-fit, Rushton, Hill & Carter 1994; Buckton & Ormerod 1997). Two further research directions also emerged. The first is to pursue the use of dippers for bio-indication, following their relatively high predictability based on RHS and their known responses to water quality. The second is to make a fuller assessment of the links between river birds and indices of anthropogenic habitat modification: six were derived from the RHS, but were little used in Chapter 5.

6.5 References

- Beutel, T.S. & Beeton, R.J.S. (1999) Building better wildlife-habitat models. *Ecography*, **22**, 219–223.
- Buckton, S.T. (1998) *Spatio-temporal patterns in the distribution and ecology of river birds*. Unpublished PhD Thesis, University of Wales, Wales.
- Buckton, S.T. & Ormerod, S.J. (1997) Use of a new standardized habitat survey for assessing the habitat preferences and distribution of upland river birds. *Bird Study* **44**, 327–337.
- Ferrier, S., Watson, G., Pearce, J. & Drielsma, M. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species level modelling. *Biodiversity and Conservation*, **11**, 2275–2307.
- Flashpohler, D.J., Bub, B.R. & Kaplin, B.A. (2000) Application of conservation biology research to management. *Conservation Biology*, **14**, 1898–1902.
- Franklin, J. (1998) Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science*, **9**, 733–748.
- McEwen, L.J., Brazier, V. & Gordon, J.E. (1997) Evaluating the geomorphology of fresh waters: an assessment of approaches. *Freshwater Quality: Defining the*

Indefinable? (eds P.J. Boon & D.L. Howell), pp. 258–281. The Stationery Office, Edinburgh.

Manel, S., Dias, J.M., Buckton, S.T. & Ormerod, S.J. (1999a) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734–747.

Manel, S., Dias, J.M. & Ormerod, S.J. (1999b) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, **120**, 337–347.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921–931.

Marchant, J. (2004) Monitoring waterways birds (and mammals). *BTO News*, **253**, 14–15.

Newson, M.D., Clark, M.J., Sear, D.A. & Brookes, A. (1998) The geomorphological basis for classifying rivers. *Aquatic Conservation: marine and freshwater ecosystems*, **8**, 415–430.

Ormerod, S.J., Barlow, N.D., Marshall, E.P.J. & Kerby, G. (2002) The uptake of applied ecology. *Journal of Applied Ecology*, **39**, 1–7.

- Ormerod, S.J. & Tyler, S.J. (1993) Birds as indicators of changes in water quality. *Birds as Monitors of Environmental Change* (eds R.W. Furness & J.J.D. Greenwood) pp 179–216. Chapman & Hall, London.
- Rushton, S.P., Hill, D. & Carter, S.P. (1994) The abundance of river corridor birds in relation to their habitat: a modelling approach. *Journal of Applied Ecology*, **31**, 313–328.
- Salwasser, H. (1986) Modeling habitat relationships of terrestrial vertebrates – the manager’s viewpoint. *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates* (eds J. Verner, M.L. Morrison & C.J. Ralph) pp 419–424. University of Wisconsin Press, Madison.
- Sutherland, W.J. (1996) *From Individual Behaviour to Population Ecology*. Oxford University Press, Oxford.
- Van Horne, B. (2002) Approaches to habitat modeling: the tensions between pattern and process and between specificity and generality. *Predicting Species Occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 63–72. Island Press, Washington.
- Walker, J., Diamond, M. & Naura, M. (2002) The development of Physical Quality Objectives for rivers in England and Wales. *Aquatic Conservation: marine and freshwater ecosystems*, **12**, 381–390.

Wiens, J.A. (2002) Predicting species occurrences: progress, problems and prospects.
Predicting Species Occurrences: issues of accuracy and scale (eds J.M. Scott, P.J.
Heglund, M.L. Morrison, J.B. Hafler, M.G. Raphael, W.A. Wall & F.B. Samson),
pp. 739–749. Island Press, Washington.

Appendix 1.1

Reproduction of the River Habitat Survey forms for the 1997 revision of the survey (as used with the Waterways Breeding Bird Survey).

A BACKGROUND MAP-BASED INFORMATION

Altitude (m)	Slope (m/km)	Flow category (1 - 10)
Solid geology code	Drift geology code	Planform category
Distance from source (km)	Significant tributary ?	Navigation ?
Height of source (m)	Water Quality Class	

B FIELD SURVEY DETAILS

Site Number : _____ Mid-site Grid Reference : _____ River : _____

Date/...../..... Time Surveyor name

Accredited Surveyor ? No Yes If yes, state code





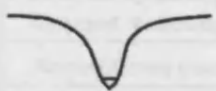
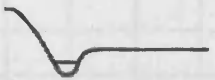
Adverse conditions affecting survey ? No Yes If yes, state

Bed of river visible ? No partially entirely (tick one box)

Duplicate photographs : general character ? No Yes (tick one box)

Site surveyed from : left bank right bank channel (tick as appropriate)

C PREDOMINANT VALLEY FORM (tick one box only)

	<input type="checkbox"/> shallow vee		<input type="checkbox"/> concave/bowl (If U-shaped glacial valley - add "U")
	<input type="checkbox"/> deep vee		<input type="checkbox"/> symmetrical floodplain
	<input type="checkbox"/> gorge		<input type="checkbox"/> asymmetrical floodplain

Terraced valley floor ? No Yes

D NUMBER OF RIFFLES, POOLS AND POINT BARS (indicate total number)

Riffles	Unvegetated point bars
Pools	Vegetated point bars

1997 RIVER HABITAT SURVEY : TEN SPOT-CHECKS

Spot-check 1 is at : upstream end downstream end of site (tick one box)

E PHYSICAL ATTRIBUTES (to be assessed across channel within 1m wide transect)

¹ = one entry only

1 2 3 4 5 6 7 8 9 10

LEFT BANK

Ring EC or SC if composed of sandy substrate

Material ¹ NV, BE, BO, CO, GS, EA, PE, CL, CC, SP, WP, GA, BR, RR, BW

Bank modification(s) NK, NO, RS, RI, PC(B), BM, EM

Bank feature(s) NV, NO, EC, SC, PB, VP, SB, VS

CHANNEL

GP- ring either G or P if predominant

Channel substrate ¹ NV, BE, BO, CO, GP, SA, SI, CL, PE, AR

Flow type ¹ FF, CH, BW, UW, CF, RP, UP, SM, NP, NO

Channel modification(s) NK, NO, CV, RS, RI, DA, FO

Channel feature(s) NV, NO, RO, MB, VB, MI, TR

RIGHT BANK

Ring EC or SC if composed of sandy substrate

Material ¹ NV, BE, BO, CO, GS, EA, PE, CL, CC, SP, WP, GA, BR, RR, BW

Bank modification(s) NK, NO, RS, RI, PC(B), BM, EM

Bank feature(s) NV, NO, EC, SC, PB, VP, SB, VS

↑ Enter channel substrates not occurring in spot-checks but present in > 1% whole site.

F BANKTOP LAND USE AND VEGETATION STRUCTURE (to be assessed over a 10m wide transect)

Land use : choose one from BL, CP, OR, MH, SC, TH, RP, IG, TL, WL, OW, SU

LAND USE WITHIN 5m OF LEFT BANKTOP

LEFT BANKTOP (structure within 1m) B/U/S/C

LEFT BANK FACE (structure) B/U/S/C

RIGHT BANK FACE (structure) B/U/S/C

RIGHT BANKTOP (structure within 1m) B/U/S/C

LAND USE WITHIN 5m OF RIGHT BANKTOP

G CHANNEL VEGETATION TYPES (to be assessed over a 10m wide transect : use E (> 33% area) or ✓ (present)

NONE

Liverworts/mosses/lichens

Emergent broad-leaved herbs

Emergent reeds/sedges/rushes

Floating-leaved (rooted)

Free-floating

Amphibious

Submerged broad-leaved

Submerged linear-leaved

Submerged fine-leaved

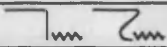



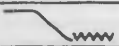

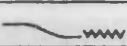

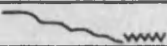

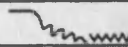
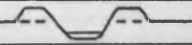
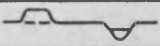
Filamentous algae

Use end "catch-all" column for types not occurring in spot checks as well as overall assessment over 500m (use E or) ✓ ↑

H LAND USE WITHIN 50m OF BANKTOP Use E (> 33% banklength) or ✓ (present)

	L	R		L	R
Broadleaf/mixed woodland (BL)			Rough pasture (RP)		
Coniferous plantation (CP)			Improved/semi-improved grass (IG)		
Orchard (OR)			Tilled land (TL)		
Moorland/heath (MH)			Wetland (eg bog, marsh, fen) (WL)		
Scrub (SC)			Open water (OW)		
Tall herbs /rank vegetation (TH)			Suburban/urban development (SU)		
			Rock and scree (RS)		

I BANK PROFILES Use E (> 33% banklength) or ✓ (present)

Natural/unmodified	L	R	Artificial/modified	L	R
Vertical/undercut 			Resectioned 		
Vertical + toe 			Reinforced - whole bank 		
Steep (>45°) 			Reinforced - top only 		
Gentle 			Reinforced - toe only 		
Composite 			Artificial two-stage 		
			Poached 		
			Embanked 		
			Set-back embankments 		

J EXTENT OF TREES AND ASSOCIATED FEATURES

TREES (tick one box per bank)			ASSOCIATED FEATURES (tick one box per feature)		
	Left	Right	None	Present	E (>33%)
None	<input type="checkbox"/>	<input type="checkbox"/>	Shading of channel	<input type="checkbox"/>	<input type="checkbox"/>
Isolated/scattered	<input type="checkbox"/>	<input type="checkbox"/>	Overhanging boughs	<input type="checkbox"/>	<input type="checkbox"/>
Regularly spaced, single	<input type="checkbox"/>	<input type="checkbox"/>	Exposed bankside roots	<input type="checkbox"/>	<input type="checkbox"/>
Occasional clumps	<input type="checkbox"/>	<input type="checkbox"/>	Underwater tree roots	<input type="checkbox"/>	<input type="checkbox"/>
Semi-continuous	<input type="checkbox"/>	<input type="checkbox"/>	Fallen trees	<input type="checkbox"/>	<input type="checkbox"/>
Continuous	<input type="checkbox"/>	<input type="checkbox"/>	Coarse woody debris	<input type="checkbox"/>	<input type="checkbox"/>

K EXTENT OF CHANNEL FEATURES (tick one box per feature)

	None	Present	E(>33%)		None	Present	E (>33%)
Waterfall(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Marginal deadwater	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cascade(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Exposed bedrock	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rapid(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Exposed boulders	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Riffle(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Unvegetated mid-channel bar(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Run(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Vegetated mid-channel bar(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Boil(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Mature island(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Glide(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Unvegetated side bar(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pool(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Vegetated side bar(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ponded Reach(es)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Discrete silt deposit(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
				Discrete sand deposit(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

L CHANNEL DIMENSIONS *(to be measured at one site on a straight uniform section, preferably across a riffle)*

LEFT BANK	CHANNEL	RIGHT BANK
Banktop height (m)	Bankfull width (m)	Banktop height (m)
Is banktop height also bankfull height? (Y or N)	Water width (m)	Is banktop height also bankfull height? (Y or N)
Embanked height (m)	Water depth (m)	Embanked height (m)

If trashline is lower than banktop break in slope, indicate: height above water (m) = _____

Bed material at site is: consolidated (compact) unconsolidated (loose) unknown

Location of measurement is: riffle run or glide other

M ARTIFICIAL FEATURES *(indicate total number or tick appropriate box)*

None	Major	Intermediate	Minor		Major	Intermediate	Minor
<input type="checkbox"/>	Weirs				Revetments		
	Sluices				Outfalls		
	Culverts				Fords		
	Bridges				Deflectors		
					Other (state)		

Is water impounded by weir/dam? No Yes, <33% of site >33% of site

N EVIDENCE OF RECENT MANAGEMENT *(tick appropriate box(es))*

None Dredging Mowing Weed-cutting

Enhancement Other (state).....

O FEATURES OF SPECIAL INTEREST *(use / or E (> 33% length))*

None <input type="checkbox"/>	Waterfalls > 5m high <input type="checkbox"/>	Artificial open water <input type="checkbox"/>	Bog <input type="checkbox"/>	Fringing reed-bank <input type="checkbox"/>
Braided/side channels <input type="checkbox"/>	Natural open water <input type="checkbox"/>	Carr <input type="checkbox"/>	Floating mat <input type="checkbox"/>	
Debris dams <input type="checkbox"/>	Water meadow <input type="checkbox"/>	Marsh <input type="checkbox"/>	Other (state).....	
Leafy debris <input type="checkbox"/>	Fen <input type="checkbox"/>	Flush <input type="checkbox"/>		

P CHOKED CHANNEL *(tick one box)*

Is 33% or more of the channel choked with vegetation? No Yes

Q NOTABLE NUISANCE PLANT SPECIES *(Use / or E (> 33% length))*

None Giant Hogweed Himalayan Balsam Japanese Knotweed Other (state).....

R OVERALL CHARACTERISTICS *(Circle appropriate words, add others as necessary)*

Major impacts: landfill - tipping - litter - sewage - pollution - drought - abstraction - mill - dam - road - rail - industry - housing - mining - quarrying - overdeepening - afforestation - fisheries management - silting

Land Management: set-aside - buffer strip - headland - abandoned land - parkland - MoD

Animals: otter - mink - water vole - kingfisher - dipper - grey wagtail - sand martin - heron - dragonflies/damselflies

Other significant observations:

S ALDERS *(tick appropriate box(es))*

Alders? None Present Extensive Diseased Alders? None Present Extensive

PHYSICAL ATTRIBUTES (SECTION E)

BANKS		CHANNEL	
<p>Predominant bank material</p> <p>NV = not visible</p> <p>BE = bedrock BO = boulder CO = cobble GS = gravel/sand EA = earth (crumbly) CL = sticky clay</p> <p>CC = concrete SP = sheet piling WP = wood piling GA = gabion BR = brick/laid stone RR = rip-rap BW = builders' waste</p>	<p>Bank modifications</p> <p>NK = not known NO = none</p> <p>RS = resectioned RI = reinforced PC = poached PC(B) = poached (bare) BM = berm EM = embanked</p> <p>Bank features</p> <p>NV = not visible (eg far bank) NO = none</p> <p>EC = eroding earth cliff (ring if sandy substrate) SC = stable earth cliff (ring if sandy substrate)</p> <p>PB = unvegetated point bar VP = vegetated point bar</p> <p>SB = unvegetated side bar VS = vegetated side bar</p>	<p>Predominant substrate</p> <p>NV = not visible</p> <p>BE = bedrock BO = boulder CO = cobble GP = gravel/pebble (ring G or P if predominant) SA = sand SI = silt/mud CL = clay PE = peat AR = artificial</p> <p>Predominant flow (see below)</p> <p>FF = freefall CH = chute BW = broken standing waves (white-water) UW = unbroken standing wave CF = chaotic flow RP = rippled UP = upwelling SM = smooth NP = no perceptible flow NO = No flow (dry)</p>	<p>Channel modifications</p> <p>NK = not known NO = none</p> <p>CV = culverted RS = resectioned RI = reinforced DA = dam/weir FO = ford (man-made)</p> <p>Channel features</p> <p>NV = not visible NO = none</p> <p>RO = exposed bedrock/boulders MB = unvegetated mid channel bar VB = vegetated mid-channel bar MI = mature island TR = urban debris (trash)</p>

FLOW TYPES

ASSOCIATED CHANNEL FEATURES




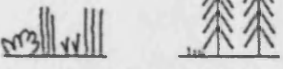




FF: Free fall	clearly separates from back-wall of vertical feature ~ associated with waterfalls.
CH: Chute	low curving fall in contact with substrate.
BW: Broken standing waves	white-water tumbling wave must be present ~ associated with <i>rapids</i> .
UW: Unbroken standing waves	upstream facing wavelets which are not broken ~ associated with <i>riffles</i>
CF: Chaotic flow	a mixture of 3 or more 'rough' flow types on no organised pattern.
RP: Rippled	no waves, but general flow direction is downstream with disturbed rippled surface ~ associated with <i>runs</i> .
UP: Upwelling	heaving water as upwellings break the surface ~ associated with <i>boils</i> .
SM: Smooth	preceptible downstream movement is smooth (no eddies) ~ associated with <i>glides</i> .
NP: No perceptible flow	no net downstream flow ~ associated with <i>pools, ponded reaches and marginal deadwater</i> .
NO: No flow	dry.

1997 RIVER HABITAT SURVEY: SPOT-CHECK KEY Page 2 of 2

LAND USE WITHIN 5m OF BANKTOP (SECTION F)

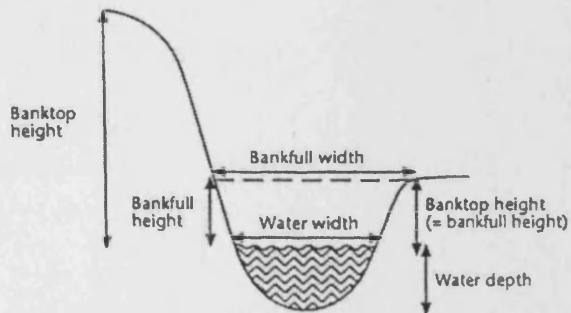
BL = Broadleaf/mixed woodland	SC = Scrub	TL = Tilled land
CP = Coniferous/plantation	TH = Tall herbs	WL = Wetland
OR = Orchard	RP = Rough pasture	OW = Open water
MH = Moorland/heath	IG = Improved grass	SU = Suburban/urban
		RS = Rock & scree

BANKTOP AND BANKFACE VEGETATION STRUCTURE To be assessed within a 10m wide transect (SECTION F)

bare	B	bare earth/rock etc.	vegetation types
uniform 	U	predominantly one type (no scrub or trees)	 bryophytes  short herbs/ creeping grasses
simple 	S	two or three vegetation types	 tall herbs/ grasses  scrub/brambles etc.
complex 	C	four or more types	 saplings and trees

Channel dimensions guidance (Section L)

- Select location on uniform section.
- If riffle is present, measure there.
If not, measure at straightest and shallowest point.
- Bankt p = first major break in slope above which cultivation or development is possible.
- Bankfull = point where river first spills onto flood plain.



Appendix 1.2

Complete listing of variables in the 1997 revision of the River Habitat Survey. Descriptions derived from RHS manual (Environment Agency 1997)

Background information

BACKGROUND MAP-BASED INFORMATION

- *Site altitude* / m
- *Slope* / m km⁻¹
- *Flow category* – from the 1985 Regional River Quality Objective maps and their equivalents in Scotland and Northern Ireland. A 1–10 scale according to annual mean discharge rate / cumecs.
- *Solid geology code* / British Geological Survey codes
- *Drift geology code* / British Geological Survey codes
- *Planform category* – an overall description of the waterway course assessed over 2.5km from the 1:50 000 scale Ordnance Survey map. Nine categories: *Natural planforms* ('straight', 'sinuous', 'irregular meanders', 'regular meanders', 'multi-thread' – split into 'anastomosing' and 'braided') and *Modified planforms* ('straightened/realigned', 'navigation', 'mill channel', 'water meadow').
- *Distance from source* / km
- *Significant tributary*: yes/no. 'Yes' if tributary of similar size to river enters the site.
- *Navigation*: yes/no. Derived from OS-Nicholson (1997).
- *Height of source* / m.
- *Water Quality Class* – taken from the 1995 Water Quality map.

FIELD SURVEY DETAILS

- *Site reference number*
- *Mid-site grid reference* – six-figure grid reference from Ordnance Survey map.
- *River name*
- *Date and time of survey*
- *Surveyor name and accreditation number*
- *Adverse conditions that may affect survey* – factors such as very high or low water level/flow and high turbidity.
- *Bed of river visible*: no/partially/entirely
- *Photographs of site*
- *Survey carried out from*: left bank/right bank/channel
- *SERCON survey* – whether extra information was collected for the System for Evaluating Rivers for CONservation.

Sweep-up – assessed over full 500m

PREDOMINANT VALLEY FORM

- Selection from six shapes: 'shallow vee', 'deep vee', 'gorge', 'concave/bowl', 'symmetrical floodplain', 'asymmetrical floodplain'.
- *Terraced valley floor* / yes/no

NUMBER OF RIFFLES, POOLS AND POINT BARS

- *Riffles* – count of discrete patches of riffle over 500m reach
- *Pools* – count of discrete pools over 500m reach
- *Unvegetated point bars* – count of point bars over 500m reach with <50% perennial plant cover.
- *Vegetated point bars* – count of point bars over 500m reach with ≥50% perennial plant cover.

Spot-checks – assessed every 50m

PHYSICAL ATTRIBUTES – ASSESSED OVER 1M SECTION

- Bank (left and right banks recorded separately):
 - *predominant material* – one from: ‘not-visible’, bedrock, boulders, cobbles, gravel/sand, earth, peat, clay, concrete, sheet piling, wood piling, gabions, bricks, rip-rap, builders’ waste
 - *bank modifications* – up to two from: ‘not-known’, none, resectioned, reinforced, poached, berm, embanked
 - *bank features* – up to two from: ‘not-visible’, none, eroding earth/sand cliffs ≥ 0.5m, stable earth/sand cliffs ≥ 0.5m, unvegetated point bar, vegetated point bar, unvegetated side bar, vegetated side bar
- Channel:
 - *predominant substratum* – one from: ‘not-visible’, bedrock, boulders, cobbles, gravel/pebbles, sand, silt, clay, peat, artificial
 - *predominant flow type* – one from: free-fall, chute, broken standing wave, unbroken standing wave, chaotic flow, rippled, upwelling, smooth, no perceptible flow, no flow (dry river bed)
 - *channel modifications* – up to two from: ‘not-known’, none, culvert, resectioned, reinforced, dam/weir, ford
 - *channel features* – up to two from: ‘not-visible’, none, exposed bedrock/boulders, unvegetated mid-channel bar, vegetated mid-channel bar, mature island, urban debris

BANKTOP LAND USE AND VEGETATION STRUCTURE – ASSESSED OVER 10M LENGTH OF BANK; LEFT AND RIGHT BANKS RECORDED SEPARATELY

- *Predominant land use within 5m of banktop* – one from: broad-leaved or mixed woodland, coniferous plantation, orchard, moorland/heath, scrub, tall herbs/rank vegetation, rough/unimproved pasture, improved/semi-improved grassland, tilled land, wetland, open water, suburban/urban development
- *Banktop* – predominant vegetation structure within 1m, from: bare (<50% vegetation cover), uniform (predominantly one vegetation type), simple (predominantly 2–3 vegetation types), complex (≥4 vegetation types).
- *Bank face* – predominant vegetation structure, from: bare, uniform, simple, complex

CHANNEL VEGETATION TYPE – ASSESSED OVER 10M LENGTH OF CHANNEL

- *None*
- *Bryophytes* – absent, present (<33% area), extensive (≥33% area)
- *Emergent broad-leaved herbs* – absent/present/extensive
- *Emergent reeds/sedges/rushes* – absent/present/extensive
- *Floating-leaved (rooted)* – absent/present/extensive
- *Free-floating* – absent/present/extensive
- *Amphibious* – absent/present/extensive
- *Submerged broad-leaved* – absent/present/extensive
- *Submerged linear-leaved* – absent/present/extensive
- *Submerged fine-leaved* – absent/present/extensive
- *Filamentous algae* – absent/present/extensive

Sweep-up – assessed over full 500m

LAND USE WITHIN 50M OF BANKTOP – LEFT AND RIGHT BANKS RECORDED SEPARATELY

- *Broad-leaved or mixed woodland* – absent, present, extensive
- *Coniferous plantation* – absent/present/extensive
- *Orchard* – absent/present/extensive
- *Moorland/heath* – absent/present/extensive
- *Scrub* – absent/present/extensive
- *Tall herbs/rank vegetation* – absent/present/extensive
- *Rough pasture* – absent/present/extensive
- *Improved/semi-improved grassland* – absent/present/extensive
- *Tilled land* – absent/present/extensive
- *Wetland* – absent/present/extensive
- *Open water* – absent/present/extensive
- *Suburban/urban development* – absent/present/extensive
- *Rock and scree* – absent/present/extensive

BANK PROFILES – LEFT AND RIGHT BANKS RECORDED SEPARATELY

- Natural/unmodified profiles
 - *Vertical/undercut* – absent/present/extensive
 - *Vertical & toe* – absent/present/extensive
 - *Steep (>45 °)* – absent/present/extensive
 - *Gentle* – absent/present/extensive
 - *Composite* – absent/present/extensive
- Artificial/modified
 - *Resectioned* – absent/present/extensive
 - *Reinforced (whole bank)* – absent/present/extensive
 - *Reinforced (top only)* – absent/present/extensive
 - *Reinforced (toe only)* – absent/present/extensive
 - *Artificial two-stage* – absent/present/extensive
 - *Poached* – absent/present/extensive
 - *Embanked* – absent/present/extensive
 - *Set-back embankments* – absent/present/extensive

EXTENT OF TREES AND ASSOCIATED FEATURES

- *Extent of trees* – left and right banks recorded separately. Six point scale: ‘none’, ‘isolated/scattered’, ‘regularly spaced – single’, ‘occasional clumps’, ‘semi-continuous’, ‘continuous’
- *Shading of channel* – absent/present/extensive
- *Overhanging boughs* – absent/present/extensive
- *Exposed bankside roots* – absent/present/extensive
- *Underwater tree roots* – absent/present/extensive
- *Fallen trees* – absent/present/extensive
- *Coarse woody debris* – absent/present/extensive

EXTENT OF CHANNEL FEATURES

- *Waterfalls* – absent/present/extensive
- *Cascades* – absent/present/extensive
- *Rapids* – absent/present/extensive
- *Riffles* – absent/present/extensive
- *Runs* – absent/present/extensive
- *Boils* – absent/present/extensive
- *Glides* – absent/present/extensive
- *Pools* – absent/present/extensive
- *Ponded reaches* – absent/present/extensive
- *Marginal deadwater* – absent/present/extensive
- *Exposed bedrock* – absent/present/extensive
- *Exposed boulders* – absent/present/extensive
- *Unvegetated mid-channel bars* – absent/present/extensive
- *Vegetated mid-channel bars* – absent/present/extensive
- *Mature islands* – absent/present/extensive
- *Unvegetated side bars* – absent/present/extensive
- *Vegetated side bars* – absent/present/extensive
- *Discrete silt deposits* – absent/present/extensive
- *Discrete sand deposits* – absent/present/extensive

CHANNEL DIMENSIONS

- Bank heights (left and right banks separately):
 - *Banktop height* – vertical distance from water level to first major break in slope (metres)
 - *Banktop height same as bankfull height: yes/no*
 - *Embanked height* – extra height added by embankments (metres)
- Channel dimensions:
 - *Bankfull width / m*
 - *Water width / m*
 - *Water depth / m*

- *Trashline* – height above water level, where applicable (metres)
- *Bed material at site of measurement* – from: consolidated, unconsolidated, unknown
- *Flow type at site of measurement* – from: riffle, run/glide, other

ARTIFICIAL FEATURES

- *Weirs* – count in each of three categories (minor/intermediate/major)
- *Sluices* – count
- *Culverts* – count
- *Bridges* – count in each of three categories (minor/intermediate/major)
- *Revetments* – count in each of three categories (minor/intermediate/major)
- *Outfalls* – count
- *Fords* – count
- *Deflectors* – count
- *Other*
- *Water impounded by weir or dam* – no/<33% of reach/≥33% of reach

EVIDENCE OF RECENT MANAGEMENT

- As appropriate from: none, dredging, enhancement, mowing, weed-cutting, other

FEATURES OF SPECIAL INTEREST

- As appropriate from: none, waterfalls > 5m high, braided/side channels, debris dams, leafy debris, artificial open water, natural open water, water meadow, fen, bog, carr, marsh, flush, fringing reed-bank, floating mat, other. Recorded as present/extensive.

CHOKED CHANNEL

- ≥33% of channel choked with vegetation: yes/no

NOTABLE NUISANCE SPECIES

- As appropriate from: none, giant hogweed, Himalayan balsam, Japanese knotweed, other.

OVERALL CHARACTERISTICS – SELECTION OF APPROPRIATE KEYWORDS

- *Major impacts* – as appropriate from: landfill, tipping, litter, sewage, pollution, drought, abstraction, mill, dam, road, rail, industry, housing, mining, quarrying, overdeepening, afforestation, fisheries management, silting
- *Land management* – as appropriate from: set-aside, buffer strip, headland, abandoned land, parkland, MoD
- *Animals* - as appropriate from: otter, mink, water vole, kingfisher, dipper, grey wagtail, sand martin, heron, dragonflies/damselflies
- *Other significant observations*

ALDERS

- *Occurrence of alders* – absent/present/extensive
- *Occurrence of diseased alders* – absent/present/extensive

References

Environment Agency (1997) *River Habitat Survey: Field Survey and Guidance Manual*. Environment Agency, Bristol.

OS-Nicholson (1997) *Nicholson Inland Waterways Map of Great Britain*. HarperCollins, London.

Appendix 1.3

Preparation of River Habitat Survey data for the analyses in Chapters 4 & 5. Systematic description of preparing the 109 variable data set, following the order used on RHS forms (Appendices 1.1 & 1.2)

General points

1. Individual categories in the spot-check sections (e.g. 'bedrock bank material', 'broken-wave flow') formed individual spot-check variables, by summarising the number of spot checks at which that characteristic was recorded (max = 10). Where left and right banks were recorded separately, totals were summed to give counts out of 20.
2. Most sweep-up variables were recorded on a 3-point ordinal scale: 'absent', 'present' (found over <33% of 500m reach) and 'extensive' (present over ≥33% of 500m reach). This was recorded as 0, 1 and 2 respectively. For variables recorded on both banks the maximum value was recorded, due to the problems of attempting to sum ordinal variables.
3. Where the same habitat characteristic was recorded in both spot-checks and sweep-up variables, one of the pair was eliminated. Spot-checks were generally preferred, due to the greater resolution of the 0–10 or 0–20 scale compared to the 3-point ordinal sweep-ups: this was considered more important than the risk of missing features between the 50m sample spacing that would be recorded by the sweep-up.
4. For some spot-check variables, such as channel modifications, 'none' (NO), 'not-visible' (NV) and 'not-known' (NK) could be recorded. These provided little information and so were removed.
5. Where fewer than 10 spot-checks were completed, or certain spot-check values deleted (see (4)), the spot-check variables were standardised to give equivalent counts out of 10 (or 20).
6. Pre-1997 records in the national RHS database include a maximum of one feature in each of 'bank modification', 'channel modification', 'bank features' and 'channel features' spot-checks. Subsequently, up to two features were recorded, but only sporadically. Only the first feature was used to provide backward compatibility and overcome inconsistent recording.

In the following list, the number of variables retained from each section (X) is given in parentheses.

Background information

BACKGROUND MAP-BASED INFORMATION

- All 11 variables excluded. Only actual habitat information collected in the field was used. ($X=0$)

FIELD SURVEY DETAILS

- All variables excluded. Only quality control variables – none describe river habitat. ($X=0$)

Sweep-up

PREDOMINANT VALLEY FORM

- Excluded. ($X=0$)

NUMBER OF RIFFLES, POOLS AND POINT BARS

- All four variables included. Used in preference to the relevant spot-check flow types or sweep-up variables describing their extent due to the greater resolution of a direct count (c.f. 0–10 count for spot-check or 3-point ordinal sweep-up). ($X=4$)

Spot-checks

PHYSICAL ATTRIBUTES

- Bank (left and right banks combined):
 - *predominant material* – all categories retained, except NV. ($X=14$)
 - *bank modifications* – ‘resectioned’, ‘reinforced’ and ‘berm’ excluded, as their sweep-up equivalents are generally more detailed, distinguishing between different types of modification. Maximum of one feature recorded. NO excluded. ($X=2$)
 - *bank features* – ‘unvegetated point bar’ and ‘vegetated point bar’ excluded as actual counts are used in sweep-up. NV & NO excluded. ($X=4$)
- Channel:
 - *predominant substratum* – all categories retained, except NV and ‘peat’ – due to its extremely low prevalence amongst the RHS sites (<1%). ($X=8$)
 - *predominant flow type* – all categories retained. ($X=10$)
 - *channel modifications* – only ‘resectioned’ included. ‘Reinforced’ was combined with ‘*predominant substratum* artificial’. Overall counts of ‘culverts’, ‘dams/weirs’ and ‘fords’ are included in the sweep-up (greater resolution). ($X=1$)
 - *channel features* – all categories included, except NV and NO, due to improved resolution over sweep-up equivalents. ($X=5$)

BANKTOP LAND USE AND VEGETATION STRUCTURE

- *Land use within 5m of banktop* – all excluded. Sweep-up 50m land use used instead – lower resolution but better lateral coverage. ($X=0$)
- *Banktop* – all included. ($X=4$)
- *Bank face* – all included. ($X=4$)

CHANNEL VEGETATION TYPE

- Recorded as present or absent at each spot check, rather than absent/present/extensive. NO excluded. ‘Submerged fine-leaved’ and ‘submerged linear-leaved’ categories combined to provide backward compatibility to pre-1997 RHS. ($X=9$)

Sweep-up

LAND USE WITHIN 50M OF BANKTOP

- All categories retained except ‘Tall-herbs’, ‘Orchard’ and ‘Rock and scree’ to provide backward compatibility. ($X=10$)

BANK PROFILES

- All categories retained except ‘Poached’ and ‘Embanked’, for which spot-checks were preferred. ($X=11$)

EXTENT OF TREES AND ASSOCIATED FEATURES

- All categories retained. ($X=7$)

EXTENT OF CHANNEL FEATURES

- ‘Riffles’ and ‘Pools’ excluded – overall counts used instead. ‘Ponded reaches’, ‘Discrete sand deposits’ and ‘Discrete silt deposits’ excluded for backward compatibility. ‘Unvegetated mid-channel bars’, ‘Vegetated mid-channel bars’, ‘Mature islands’, ‘Unvegetated side bars’, ‘Vegetated side bars’ and ‘Exposed boulders’ excluded in favour of spot-check equivalents. ($X=8$)

CHANNEL DIMENSIONS

- Only ‘Banktop height’ (mean of left and right banks), ‘Water width’ and ‘Water depth’ included. Other variables were only recorded sporadically. ($X=3$)

ARTIFICIAL FEATURES

- Simplified to allow backward compatibility, as 1997 survey included more detail of anthropogenic modifications than previous years. Number of culverts, fords, outfalls, weirs and bridges retained. The latter three were converted to 5-point ordinal scales to cope with the small numbers of large outliers. For each type, absence was recorded as zero, one as one, 2–3 as two, 4–5 as three and >6 as four. ($X=5$)

EVIDENCE OF RECENT MANAGEMENT

- Excluded. ($X=0$)

FEATURES OF SPECIAL INTEREST

- Excluded. ($X=0$)

CHOKED CHANNEL

- Excluded. ($X=0$)

NOTABLE NUISANCE SPECIES

- Excluded. ($X=0$)

OVERALL CHARACTERISTICS

- Excluded. ($X=0$)

ALDERS

- Excluded. ($X=0$)

Appendix 2

**The 161 sites selected for the Waterways Breeding Bird
Survey analysis (Chapter 5)**

River or canal	National Grid Reference*	WBBS transect number	RHS site number
Many Burns River	H382497	1	22910
Many Burns River	H399510	6	22916
Glenlark River	H576873	1	22918
Glenlark River	H589888	6	22923
Ballinderry River	H666802	2	22925
Maldie Burn	NC255345	2	21375
Hamra River	NG194471	3	22929
Allt Coire Sgoireadail	NG953065	1	22931
Allt Coire Sgoireadail	NG973084	7	22938
Allt Coire nan Eiricheallach	NG999041	3	21380
Allt a' Choire Dhomhain	NH157286	4	21387
Allt Cam Ban	NH358498	2	21391
River Findhorn	NH704168	1	21392
River Findhorn	NH669143	10	21401
River Nairn	NH685348	1	21402
River Nairn	NH675321	10	21411
River Ness (non-tidal part)	NH657433	3	21414
Am Beanaidh	NH925029	2	22939
Am Beanaidh	NH917000	10	22948
Water of Buchat	NJ326188	1	22949
Water of Buchat	NJ389158	10	22967
South Ugie Water	NK040478	2	22969
South Ugie Water	NK053483	8	22973
Dubh Lighe	NM964787	1	22974
Dubh Lighe	NM939792	7	22980
River Kingie	NN040978	1	21425
River Kingie	NN002965	10	21434
Allt a' Cham Dhoire	NN071871	1	21417
Allt a' Cham Dhoire	NN042863	6	21424
Allt an Stacain	NN157217	2	22982
River Spean	NN187835	1	21435
River Spean	NN219816	9	23044
Allt Feith Thuill	NN394730	2	21444
Allt Coire Ardair	NN464889	1	21452
Allt Coire Ardair	NN442884	6	21447
Allt a' Chrannaig	NN484879	2	21454
River Spey	NN637938	1	21469

River Spey	NN599938	10	21460
Unnamed, feeds into aqueduct	NN685861	2	21458
Milton Burn	NN743986	1	21471
Milton Burn	NN722959	10	21480
Buckny Burn/Lunan Burn	NO086455	1	22994
Buckny Burn/Lunan Burn	NO062477	10	22985
Baddoch Burn	NO130828	3	21583
River Dee	NO212914	2	21586
Dean Water	NO333478	1	23009
Dean Water	NO290458	10	22995
Kenly Water	NO549119	3	21591
Bonny Water	NS799789	7	21609
Crook Burn	NS973062	1	21617
Crook Burn	NS984041	6	21611
Blackadder Water	NT827531	1	23244
Blackadder Water	NT856543	10	23237
Cross Water of Luce	NX193743	1	21628
Cross Water of Luce	NX180769	10	21618
River Bleng	NY083028	2	23287
Kershope Burn	NY483828	1	23291
Kershope Burn	NY519848	10	23300
Lewis Burn	NY632882	3	23220
River Skerne	NZ293205	1	23101
River Skerne	NZ300193	6	23106
River Wear	NZ285441	1	23301
River Wear	NZ300456	7	23307
Skelton Beck	NZ664207	3	23369
Leeds & Liverpool Canal	SD364070	1	23041
Leeds & Liverpool Canal	SD369089	6	23036
Eagley Brook	SD722129	2	23108
River Wenning	SD747671	1	23247
River Wenning	SD718676	8	23254
Rochdale Canal	SD892040	1	23111
Rochdale Canal	SD884077	10	23120
River Wharfe	SE002633	1	23217
River Wharfe	SD981657	8	23214
River Cover	SE043806	1	23308
River Cover	SE024793	6	23313
River Swale	SE320893	1	23357
River Swale	SE335879	8	23355

New River Ancholme	SE972167	1	23255
New River Ancholme	SE975207	9	23263
Afon Eiddew	SH953246	3	23123
Afon Banwy neu Einion	SJ112074	2	22595
Afon Iwrch	SJ133270	1	23125
Afon Iwrch	SJ127298	7	23132
Afon Tanat	SJ187239	1	23133
Afon Tanat	SJ223240	10	23142
Manchester Ship Canal	SJ464774	3	23145
River Severn	SJ670035	1	23314
River Severn	SJ637044	8	23321
River Weaver	SJ649526	1	23148
River Weaver	SJ664551	10	23157
Shropshire Union Canal	SJ872108	9	23266
River Noe	SK156866	6	23364
River Maun	SK585647	2	23161
Fosdyke Navigation	SK906750	1	23171
Fosdyke Navigation	SK882746	6	23166
River Brant	SK943593	3	23174
Afon Teifi	SN651569	3	23010
Lower Clydach River	SN677037	3	23178
River Clydach	SN739008	1	22469
Afon Rhywni	SO123057	2	23223
Afon Rhywni	SO136044	9	23015
Grwyne Fawr	SO242298	5	23016
Afon Mynwy	SO476175	1	23225
Afon Mynwy	SO468198	8	23023
River Teme	SO631687	1	22445
River Teme	SO654688	7	22451
River Rea	SO664819	1	23181
River Rea	SO664786	9	23198
River Severn	SO719976	1	23199
River Severn	SJ707001	8	23206
River Frome	SO785055	1	23207
River Lew	SS532054	1	23339
River Taw	SS693104	9	22904
Afon Elai	ST037811	1	23184
Afon Elai	ST035823	6	22907
River Tone	ST083212	3	21695
River Otter	ST167016	2	23337

River Frome	ST787466	4	23191
River Avon	ST958805	2	22414
River Avon	ST961830	1	22415
River Avon	ST979822	6	22420
River Kennet	SU290713	3	22649
River Thames/Isis	SU511970	2	23046
River Enborne	SU563642	2	22333
River Loddon	SU737669	3	21486
River Rother	SU963200	1	23029
River Rother	SU978192	6	23229
Virginia Water (outflow)	SU980685	1	23268
River Camel	SX084742	1	23030
River Camel	SX074718	9	23034
River Lyd	SX466833	3	23049
River Otter	SY095961	1	23327
River Otter	SY113987	7	23333
River Axe	SY264944	3	23057
Relief Channel	TF602035	1	23035
River Ivel	TL183415	3	22642
River Ivel	TL223371	2	22646
King's Dike (Drain)	TL225964	1	23051
King's Dike (Drain)	TL247966	6	23056
Forty Foot or Vermuden's Drain	TL342880	1	23345
Forty Foot or Vermuden's Drain	TL318880	6	23340
Twenty Foot River (Drain)	TL318964	1	23346
Twenty Foot River (Drain)	TL346985	8	23353
Mildenhall Drain	TL652820	2	23060
River Lark	TL734738	1	23069
River Lark	TL760728	7	23063
Landermere	TM192240	1	21736
River Brent	TQ146813	1	23070
River Brent	TQ241894	2	23072
River Darent	TQ523623	2	23077
River Medway	TQ538440	3	23075
River Roding	TQ544996	1	23277
River Roding	TQ520982	8	23079
River Medway	TQ738537	1	23088
River Medway	TQ706528	9	23278
Cliffe Fleet	TQ739785	2	23089
River Rother (non-tidal part)	TQ937238	2	23092

Great Stour	TR033442	3	23098
New Sewer	TR086272	1	23096
New Sewer	TR061264	7	23282
Great Stour	TR161594	2	23100

*Grid references are taken from the National RHS database, Version 3.3.

Appendix 3.1

Published version of Chapter 2, ‘Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data’.

***Conservation Biology*, 17, 1601–1611.**

Improving the Quality of Distribution Models for Conservation by Addressing Shortcomings in the Field Collection of Training Data

I. P. VAUGHAN* AND S. J. ORMEROD

Catchment Research Group, School of Biosciences, Cardiff University, P.O. Box 915, Cardiff CF10 3TL, United Kingdom

Abstract: Conservation biology can benefit greatly from models that relate species' distributions to their environments. The foundation of successful modeling is a high-quality set of field data, and distribution models have specialized data requirements. The role of a distribution model may be primarily predictive or, alternatively, may emphasize relationships between an organism and its habitat. For the latter application, the environmental variables recorded should have direct, biological relationships with the organism. Interacting species may be valuable predictors and can improve understanding of distribution patterns. Sampling should cover the full range of environmental conditions within the study region, with samples stratified across major environmental gradients to ensure thorough coverage. Failure to sample correctly can lead to erroneous organism-environment relationships, affecting predictive ability and interpretation. Sampling ideally should examine a series of spatial scales, increasing the understanding of organism-environment relationships, identifying the most effective scales for predictive modeling and complementing the spatial hierarchies often used in conservation planning. Consideration of statistical issues could benefit most studies. The ratio of sample sites to environmental variables considered should ideally exceed a ratio of 10:1 to improve the analytical power and reliability of subsequent modeling. Presence and/or absence models may suffer bias if training data detect the study organism at an atypical proportion of sites. We considered different strategies for spatial autocorrelation and recommend it be included wherever possible for the benefits in biological realism, predictive accuracy, and model versatility. Finally, we stress the importance of collecting independent evaluation data and suggest that, as with the training data, a systematic approach be used to ensure broad environmental coverage, rather than relying on a random selection of test sites.

Key Words: analytical power, distribution modeling, environmental space, model evaluation, sampling scale, species prediction

Mejorando la Calidad de los Modelos de Distribución para la Conservación Enfocando en las Limitaciones en la Recolección a Campo de Datos de Capacitación

Resumen: La biología de la conservación puede obtener grandes beneficios de modelos que relacionan las distribuciones de especies con sus ambientes. El fundamento del modelado exitoso es un conjunto de datos de campo de alta calidad, y los modelos de distribución tienen requerimientos especializados de datos. El papel de un modelo de distribución puede ser primordialmente predictivo o, alternativamente, puede enfatizar las relaciones entre el organismo y su hábitat. Para esta última aplicación, las variables ambientales registradas deben tener relaciones biológicas directas con el organismo. Las especies interactuantes pueden ser predictores valiosos y pueden mejorar el entendimiento de los patrones de distribución. El muestreo debe abarcar todo el rango de condiciones ambientales en la zona de estudio con muestras estratificadas a lo largo de gradientes ambientales importantes para asegurar una cobertura minuciosa. Los errores de muestreo pueden conducir a relaciones organismo/ambiente erróneas, afectando la capacidad predictiva e interpretación. Idealmente, el muestreo debería examinar una serie de escalas espaciales, que aumenta el entendimiento de las

*email vaughanip@cardiff.ac.uk

Paper submitted August 12, 2002; revised manuscript accepted February 26, 2003.

relaciones organismo/ambiente, identifica las escalas más efectivas para el modelado predictivo y complementa las jerarquías espaciales usadas a menudo en la planificación de la conservación. La consideración de temas estadísticos podría beneficiar a la mayoría de los estudios. La proporción entre los sitios de muestreo y las variables ambientales consideradas idealmente debe exceder 10:1 para mejorar el poder analítico y la confiabilidad del modelo posterior. Los modelos de presencia/ausencia pueden tener sesgos si los datos de preparación detectan al organismo en estudio en una proporción atípica de sitios. Consideramos diferentes estrategias de autocorrelación espacial y recomendamos que se incluyan cuando sea posible en beneficio del realismo biológico, la precisión predictiva y la versatilidad del modelo. Finalmente, enfatizamos la importancia de coleccionar datos de evaluación independientes y sugerimos que, así como los datos de preparación, se utilice un enfoque sistemático para asegurar una cobertura ambiental amplia, en lugar de confiar en una selección aleatoria de sitios de muestreo.

Palabras Clave: escala de muestreo, espacio ambiental, evaluación de modelo, modelado de distribución, potencia analítica, predicción de especies

Introduction

Relationships between organisms and their environment are fundamental to ecology (Begon et al. 1996). In the past, detailed autoecology was the foundation for studying such relationships, generally over limited spatial extents. As the scope of conservation biology grows to cover huge spatial scales, capture global environmental change, and make itself relevant to remote regions of the world, models to relate organisms to their environments are needed to fulfil this role (Fielding & Haworth 1995; Gaston & Blackburn 1999; Manel et al. 1999, 2000). Once they are carefully validated, distribution models can provide useful insights into species-environment relationships and represent valuable tools in conservation planning (e.g., Mladenoff et al. 1999; Franco et al. 2000; Guisan & Theurillat 2000; Clewenger et al. 2002; Woolf et al. 2002). Such models are now being applied in conservation, from studies concerned with single species to those predicting whole communities and ecosystem attributes such as diversity and vegetation physiognomy (Franklin 1995). Two recent reviews examine the diversity of these modeling techniques and their applications (Franklin 1995; Guisan & Zimmermann 2000).

With so much at stake biologically and the need to target conservation resources as efficiently as possible, it is imperative that the methodology for distribution modeling be optimized. A good set of training data is the basic foundation for successful distribution modeling, yet there is a surprising paucity of guidance for such data collection, especially compared with the wealth of general fieldwork texts. Here we address this shortcoming, discussing a series of field-survey ideas that could improve distribution modeling. Problems with training data could be contributing to a range of weaknesses observed in distribution models (Table 1). Further support for this idea comes from studies showing similar predictive performance for radically different model types trained with the same data set (e.g., discriminant analysis and logistic regression compared with artificial neural networks;

Manel et al. 1999). Our hope is that by illustrating some of the problems and potential solutions, the quality of the training data can be improved, thereby underpinning analytical success.

Clarification of Model Aims

All distribution models attempt to quantify the relationships between organisms and their environments. A dichotomy exists, however, between emphasis placed primarily on the absolute predictive ability of the model and one on interpretation of the structure of the model—exploring the organism-environment relationship (Ripley 1994; Mac Nally 2000). The nature of the study dictates whether prediction, explanation-interpretation, or, as is more commonly the case, both approaches are of interest. During subsequent model development, the analysis may benefit from different methods for these two roles (Mac Nally 2000). At the point of fieldwork design, the subsequent aim of the study should be considered when the choice is made of which environmental variables to record.

Many environmental variables correlate with species' distribution patterns. All such variables could theoretically be employed for predictive modeling, but to improve understanding of an organism-environment relationship, variables whose correlations represent actual biological relationships are required—so-called direct variables (Table 2). Working on Himalayan river birds, Manel et al. (1999) showed that distribution models based on elevation and slope had, in many instances, predictive abilities similar to those of models based on detailed habitat data. For explanatory applications, however, such models are of little value, necessitating further investigation into the correlations between elevation and slope and direct variables. Variables such as elevation and slope are therefore considered indirect (Table 2).

Indirect variables can be useful in predictive modeling, most notably for the increases in overall survey efficiency

Table 1. Major problems experienced in the use of distribution models and ways in which shortcomings in the training data may contribute.

<i>Problem</i>	<i>Possible data-related causes</i>
Poor predictive ability (either reclassifying training data or application to new data)	failure to record important predictor variables, failure to take account of interacting species, predictor variables recorded at spatial scales where they correlate weakly with distribution, poor coverage of environmental space during sampling—leading to inaccurate characterization of species/environment relationships, training data unrepresentative of the wider environment, statistical weaknesses in the training data (e.g., too few sampling sites for the number of predictor variables analyzed), spatial autocorrelation overlooked—ignoring a potentially important predictor and reducing the reliability of modeling (e.g., increased risk of including spurious variables)
Consistent over- or underoptimism in predictions with new data (presence/absence models)	study organism recorded at an atypically high or low proportion of sites in the training data compared with wider environment
Apparent responses to environmental conditions inconsistent with biological understanding	relationships between organism and predictor variables poorly sampled and so inaccurately modeled, failure to record critical environmental variables, failure to take account of interacting species, training data poorly suited to model development (e.g., too few sampling sites for number of variables analyzed), spatial autocorrelation overlooked—local processes generating autocorrelation (e.g., dispersal) may modify correlations between organisms and environmental variables
Organism–environment relationships difficult to interpret	use of environmental variables that have indirect relationships with the distribution of the organism rather than directly affecting it, necessitating additional level of interpretation of organism–environment relationships

they provide (Table 2). The ability to use map-derived variables to predict the distribution of river birds in the Himalayas, rather than having to visit each site on the ground in difficult terrain, would be highly beneficial (Manel et al. 1999). Nevertheless, direct predictors are favored in most cases for their reliability and generality (Austin

& Meyers 1996). This stems from their biological relationships with the organism, which should remain relatively constant compared with indirect variables, whose correlations with direct variables by which their relationships to the organism actually occur are location-specific (Table 2; Austin et al. 1984). Studies encompassing large

Table 2. Choice between indirect and direct predictor variables in the design of field studies.^a

	<i>Direct predictors^b</i>	<i>Indirect predictors</i>
Definition	variables with biological relationships with the study species	variables whose correlations with the study species represent correlations with a series of intermediate, direct factors rather than a direct relationship
Examples	climate, nesting sites, soil nutrients (plants), interacting species, site isolation	elevation, topography, geographical position, geology, soil nutrients (animals)
Strengths	structure of the model more easily interpreted in biologically meaningful terms, potentially greater ecological value direct biological relationships should generalize better to new areas and be more effective for use in climate/habitat change scenarios than indirect ones provides more valuable information for conservation management	data sets often widely available a priori in maps or geographic information system low cost, ease of collection can be effective predictors (e.g., elevation and topography in mountainous areas) encompass a range of correlated variables so should: reduce multicollinearity problems, increase stability of model development as fewer candidate variables are considered, result in more parsimonious models, save resources by recording fewer variables
Weaknesses	variables often require greater effort to record, especially for biotic variables several variables may need to be recorded where a single indirect predictor (e.g., elevation) would suffice, data sets may need to be estimated for large spatial extents, often with indirect variables, reducing their overall accuracy (e.g., estimation of climatic data based on elevation)	correlation with direct variables tends to be location specific, making models particularly weak in extrapolation (e.g., latitudinal change in elevation/climate correlations) limited value for interpretation - biological meaning has to be read into the relationships, adding a further level of interpretation, thereby increasing uncertainty

^aSources: Austin 1980; Austin et al. 1984; Franklin 1995; Austin & Meyers 1996; Guisan & Zimmermann 2000.

^bTo simplify the generalization of this framework across all distribution modeling, this is a combination of Austin's (1980) direct and resource gradients.

spatial extents, perhaps looking at the complete biogeographical range of an organism, should place particular importance on the use of direct predictors, where feasible. Similarly, correlations between indirect and direct variables may alter over time following major environmental change, in which case direct predictors are again preferred.

The distributions of interacting species should also be considered as possible variables in distribution modeling. Competing species, generalist predators, parasites, and pathogens can exclude a species from a region, either locally or at broad spatial scales (Begon et al. 1996). As such, they may account for a proportion of the unexplained variation in the distribution pattern that remains once a model based purely on the physical environment is fitted (Table 1). The extra proportion such interacting species explain represents a difference between the fundamental and realized niches. The former considers simply the range of environmental conditions the organism can tolerate, and the latter also accounts for the regions from which the organism is excluded by other species (Hutchinson 1957). Field data always pertain to the realized niche (Austin et al. 1990; Westman 1991), and failure to account for this may weaken the analytical process, confound interpretation, and generate false-positive (commission, Type I) prediction errors.

In many cases, data describing the distributions of interacting species could be recorded with little extra effort. Similar organisms may be recorded in the same sampling technique, such as potential avian competitors in a standard point count or transect. Where competitors or predators have been included in distribution models, valuable increases in predictive ability and especially explanatory value have been observed. Özesmi and Özesmi (1999) showed major improvements in modeling success if competing wetland breeding birds were considered in their model. Similarly, distance to the nearest Golden Eagle (*Aquila chrysaetos*) nesting cliff proved a useful predictor for Peregrine (*Falco peregrinus*) selection of nesting cliffs (Gainzarain et al. 2000). Lowe and Bolger (2002) showed how piscivorous predators were valuable predictors of salamander abundance. Rushton et al. (2000) showed a similar effect with a nonindigenous predator, the American mink (*Mustela vison*), on the distribution of the water vole (*Arvicola terrestris*) in the United Kingdom. There also is good potential to develop these simplistic approaches to plant distribution modeling in instances where distribution patterns are strongly affected by small numbers of easily recorded species. For example, the populations of certain mistletoe species in New Zealand are being affected by reduced populations of their avian pollinators (Robertson et al. 1999), whereas introduced herbivores and livestock can affect floristic composition in a range of habitats (e.g., Veblen et al. 1992; Nugent et al. 2001).

Accurately Capturing Organism-Environment Relationships

Multivariate wildlife-habitat models define the distribution of a species within environmental space—a multivariate concept covering all possible combinations of environmental conditions, rather than individual environmental gradients (Margules & Stein 1989). To accurately capture this distribution, samples need to be taken across the complete region of environmental space within which the organism occurs, including sites that define the boundaries of the organism's niche. Failure to do so undermines the analytical process and could lead to serious errors in both explanatory work and predictive applications (Table 1).

The effects of incomplete sampling are clearly illustrated when organism-environment relationships are visualized in terms of species-response curves to individual gradients, as is widely done in plant-distribution modeling. If, for example, a species shows a simple Gaussian response to an environmental gradient, incomplete sampling that overlooked the upper end of the gradient may result in a sigmoid curve being fitted (Fig. 1a). Erroneous predictions of occurrence (Type I or commission errors) would then be made in applying the model to sites high up on the gradient. This type of sampling-generated problem can be seen in practice where models are built from a limited geographic region, especially the edge of a species' range, and then applied more widely. Similarly, missing both ends of a gradient may provide the erroneous impression that the organism shows no response to that environmental factor, when, again, complete sampling would reveal a broadly Gaussian response (Fig. 1b). Austin et al. (1984) provide an empirical example of sampling effects, obtaining different response curves for a *Eucalyptus* species depending on the completeness with which a rainfall gradient was sampled. Further discussion is provided by Mohler (1983) and Westman (1991). The implications for sampling design are profound: through different combinations of oversampling and undersampling regions of environmental space, any response curve could theoretically be obtained (Yee & Mitchell 1991).

Environmental stratification is the only systematic approach to sampling environmental space (Margules & Austin 1994). Major environmental gradients are identified, such as elevation and geology, and each is divided into a series of strata at regular intervals. Stratification then attempts to place samples within every possible combination of the strata from these environmental gradients. For example, with five elevation and four geological categories, there are 20 potential strata to sample. This number may become very large for stratification by several gradients, but fortunately only a fraction usually occur in the environment (e.g., 215 of 504 possible strata in a forested region of New South Wales [Austin & Heyligers

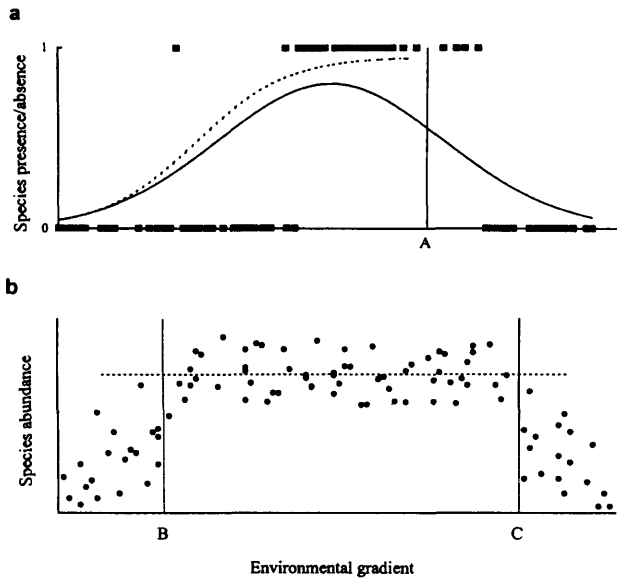


Figure 1. The potential impact of incomplete sampling of environmental gradients on subsequent interpretation. In these hypothetical examples, the occurrence or abundance of a species is recorded along a single environmental gradient. In (a) sampling only as far as point A (missing the top end of the gradient) could lead to a sigmoid response curve being fitted (dashed line) in place of the Gaussian response curve (solid line). In (b) sampling only between points B and C would lead to the erroneous conclusion that the environmental factor has no relationship with the organism (dashed line).

1989] or 15 of 20 broad climatic strata in southern Australian forest [Neave et al. 1996]). Sampling sites should be allocated equally between all strata present, with several replicates in each, to provide robust, wide coverage of environmental space for later analysis (Guisan & Zimmermann 2000). Hirzel and Guisan (2002) demonstrate how, for a fixed sample size, stratifying sites in this way improves model performance compared with randomized sampling.

Successful environmental stratification is conditional on identifying the complete expanse of environmental space found within the study region to ensure comprehensive sampling. Every region in which a predictive model may be applied must be identified, and the geographical limit of explanatory conclusions clarified. Collingham et al. (2000) demonstrated the consequences of failing to account for the region of model application. They constructed a predictive model for alien riparian weeds at a regional scale (263 km²) and applied it nationally to England and Wales. Performance was poor, especially compared with a model built at the national scale and applied regionally (Collingham et al. 2000).

The basic sampling protocol may be further customized to the individual needs of the study, conserving precious financial resources, especially when large spatial scales are involved. An increased level of subjectivity may be introduced, but Austin and Heyligers (1989) argue that this should not be a problem, provided that sampling rules are stated explicitly so that any biases are evident. Two ideas in particular are gaining recognition. The first is gradsect sampling, in which sampling is limited to discrete areas selected to contain maximum environmental heterogeneity to minimize traveling between sites. Empirical work collecting biological inventories demonstrated similar performance between gradsect sampling and full environmental stratification across the area (Wessels et al. 1998). The second is the use of rules to limit the resources expended in sampling each site. An example is specifying a maximum distance from the road network at which a site will still be sampled (e.g., Austin & Heyligers 1989). Examples of environmental stratification methods and gradsect sampling are given by Gillison and Brewer (1985), Austin and Heyligers (1989), Neave et al. (1996), and Wessels et al. (1998).

Sampling Scale for Analyzing Organism-Environment Relationships

The correlations observed between organisms and their environments are frequently dependent on the spatial scale of observation and may vary between scales in a complex manner (e.g., Wiens et al. 1987; Orians & Wittenberger 1991; Carroll et al. 1999; Collingham et al. 2000; Orrock et al. 2000; Johnson et al. 2001). This is a manifestation of the way in which the apparent importance of such factors as different environmental variables, behavioral decisions, interspecific interactions, and demographic factors change with the scale of observation (Wiens 1989). Consequently, models employing a single study resolution provide only a limited "snapshot" of organism-environment relationships. The generality of these relationships to other scales is unknown, and it may not be possible to find strong organism-environment relationships at the spatial scale of the sampling, creating weak predictive models (Table 1). The solution is to employ a multiscale sampling approach wherever possible, unless there are good reasons for restricting modeling to a single scale (e.g., testing out a standardized habitat survey for birds; Buckton & Ormerod 1997). In such cases, caveats to the model's use regarding spatial scale should always be made clear.

Multiscale distribution modeling is still in its relative infancy, often requiring greater quantities of data and new, more complex analytical approaches. Many studies now incorporate simple information to describe the landscape context of the sampling sites, in addition to simple habitat quality at those sites. This may include the distance

to neighboring patches of similar habitat or site isolation (e.g., Hinsley et al. 1995; Lowe & Bolger 2002) or the density of similar habitat types in the area (e.g., Vander Haegen et al. 2000). An alternative approach is to build models at two distinct scales, such as for quadrats of 1 m² and 100 m² for rare plant species in the Appalachians (Wiser et al. 1998).

More systematic approaches to multiscale investigations utilize scale hierarchies, incorporating four or more spatial scales (e.g., Carroll et al. 1999; Söderström & Pärt 2000; Apps et al. 2001; Mackey & Lindenmayer 2001). This provides a convenient conceptual framework for multiscale modeling and is readily compatible with multiscale management frameworks often employed at regional or national levels (e.g., Saunders et al. 1993; Wikramanayake et al. 1998; Zacharias & Roff 2000). This should maximize the relevance of models in decision-making, enhancing their overall conservation value. Sampling at several spatial scales increases the chance of identifying the scale at which distribution is most effectively predicted, increasing the potential efficacy of the model (Wiens 1989).

The choice of scales to examine in a hierarchy requires careful thought, unless it is predetermined by management objectives. Convenient scales for research, such as standard grid squares on a map, are entirely arbitrary. Assuming little prior knowledge of the species, study design may benefit from identifying an objective, logical classification of the environment into a series of distinct spatial scales, at each of which the organism-environment relationships could be analyzed (Poizat & Pont 1996). Tools for spatial analysis provide objective ways of identifying the scales at which environmental patterns change, from which it is inferred that the major factors shaping the environment change (Wiens 1989). Krummel et al. (1987) provide an example of this, applying fractal analysis to a fragmented woodland landscape. Regions of self-similarity across spatial scales were observed, suggesting similar controlling factors but with a marked change when woodlands reached around 60–73 ha. This appeared to be the scale at which the fragmentation process was occurring and would be an obvious choice for a spatial scale in a modeling study examining the distribution of woodland birds in fragmented landscapes.

Improving Analytical Success

Well-planned experiments consider the subsequent data analysis during their conception, and this applies equally to distribution modeling, even though it does not conform to a rigid experimental structure. Adequate analytical power at the model-development stage is crucial to successful model development yet has scarcely been mentioned in the ecological literature, in contrast to the

medical literature (e.g., Harrell et al. 1984, 1996; Peduzzi et al. 1996). The crucial stage in model development is the selection of which environmental variables are to be retained in the final model and, to a lesser extent, the estimation of their weightings and coefficients. Although most work has concerned regression models, the recommendations should still provide useful starting points for other distribution modeling techniques.

The analytical power and reliability of regression models is largely a function of the number of sites sampled relative to the number of candidate variables considered (Peduzzi et al. 1996). For presence/absence models, the number of sampling sites is restricted to that fraction representing the rarer outcome event (presence or absence). As the number of observations per variable increases, the analytical power and the reliability of the analysis increases (Steyerberg et al. 1999). As the ratio decreases, the risk of spurious variables being included grows, and the chance of overlooking “important” variables increases. A range of studies has concentrated on establishing a minimum number of sampling sites per candidate variable to provide adequate statistical power. Most suggest a minimum advisable ratio of around 10:1 (e.g., Harrell et al. 1984; Peduzzi et al. 1996), although some suggest as little as half that (e.g., Freedman et al. 1992).

Conservation biologists face an acute challenge regarding analytical power. Sampling resources limit the overall sample size, and environmental complexity demands the use of extensive numbers of variables. A large proportion of studies has used ratios far short of 10:1, requiring further caution in their interpretation. The use of biological knowledge to select candidate predictors for recording, and during the model-selection process, should help to some degree, as could the use of indirect variables because they act as proxies for a series of other variables (Table 2). Data-reduction techniques such as principal components analysis may also be valuable, especially for prediction, because they can dramatically reduce the dimensionality of the data set and overcome colinearity between variables (Guisan et al. 1998; Wiser et al. 1998; Manel et al. 1999; Cumming 2000; Lowe & Bolger 2002). Their major disadvantage is in complicating interpretation.

For distribution models utilizing presence/absence data, the prevalence of the organism within the data set—the proportion of sites at which it is observed—has two important effects on the subsequent data analysis. The first concerns analytical power, which, as stated above, is limited by the frequency of the rarer outcome event (presence or absence). Conservation biology's primary concern with rare species, whose prevalence in training data may be well below 20%, leaves it exposed to serious analytical weaknesses. The second implication of prevalence is the potential for bias if the prevalence observed in the training data is atypically high or low (Pearce &

Ferrier 2000). Many presence/absence modeling algorithms are biased toward the larger outcome group (presence or absence) during their calibration (Fielding 1999). If the prevalence is markedly different in subsequent applications of the model, perhaps in a different geographic region, predictions could consistently over- or underestimate the probability of occurrence (Pearce & Ferrier 2000).

There is no panacea for the problems of species prevalence, and rare organisms will continue to cause problems by necessitating the collection of large amounts of data in the hope of obtaining a reasonable level of analytical power. This effort may be aided by biasing sampling toward rare strata expected to support the organism or including sites already known to hold the species (e.g., traditional nesting sites for raptors; Fielding & Haworth 1995). This must be done sparingly if the prevalence is not to be inflated to the point where bias is introduced to the model. Environmental stratification divided equally between strata will itself bias sampling toward rarer strata and so may increase the prevalence. The consequences of these artificial increases in prevalence could include overoptimistic assessments of the species' status and resources wasted targeting conservation efforts toward unsuitable sites. When working with very rare or endangered species, however, these false positive (Type I, commission) errors may be considered less serious than overlooking any of the limited number of occupied sites.

Consistent biases in the model may only come to light after widespread application across new areas and between years (due to population fluctuations). Even then, they will not affect a model that is merely required to rank sites according to their relative probability of occupancy or suitability for the species (Pearce & Ferrier 2000). However, if absolute occupancy or suitability predictions are required, the model could be easily recalibrated at a later date to improve its accuracy. This is perhaps to be expected if a distribution model is to be widely used as a conservation tool.

Tackling Spatial Autocorrelation

Awareness of spatial autocorrelation and the need to address it in fieldwork and data analysis has increased dramatically in recent years (Legendre 1993; Koenig 1999; Carroll & Pearson 2000). From the standpoint of field study design, before distribution data are available to characterize autocorrelation, a decision is required to either attempt to exclude autocorrelation from the training data or to use it as a candidate predictor variable. Attempting to model a distribution showing spatial autocorrelation, without taking account of it, can affect both predictive and explanatory analyses (Table 1).

Two main ideas underpin the desire to exclude autocorrelation. The first is to retain simple descriptions of organism-environment relationships at discrete sites for management purposes. Practitioners may find limited value in a model that draws much of its information from landscape context and surrounding sites—factors likely to be beyond their management. The second idea is to increase the reliability of the modeling process because site independence is a basic requirement for valid statistical testing in model selection (Legendre 1993; Guisan & Theurillat 2000). Positive spatial autocorrelation, for example, tends to increase the probability of spurious predictor variables being included in the model (see Legendre 1993).

Attempts to remove autocorrelation generally involve the imposition of a minimum distance between samples in response to an assumed patchy distribution. Guisan and Theurillat (2000), for example, employed a 250-m spacing between samples in distribution modeling for alpine vegetation. Such a distance is necessarily speculative, unless distribution data are available in advance.

Attempting to filter out autocorrelation carries several disadvantages, however. It involves deliberately overlooking a potentially valuable variable for both prediction and explanation. Autocorrelation between neighboring sites, at distances less than the rigid intersite spacing, may still affect species occurrence at the sampling sites, perhaps through local territorial behavior, in a way that cannot be explained with autocorrelation removed. This could create "noise" in the data set, weakening the analysis. From a sampling viewpoint, fixing a rigid intersite spacing may make it difficult to sample environmental space fully, especially where extensive heterogeneity exists over short distances. Finally, in terms of subsequent model application, there may be problems if there is a desire to apply the model to sets of contiguous sites.

Including spatial autocorrelation within distribution modeling provides a series of potential benefits. A greater understanding of the processes controlling distribution may be achieved in which sites are no longer considered in isolation but in the landscape context. At local scales in particular, the accuracy of predictions may be improved (Augustin et al. 1996; Araújo & Williams 2000). Strong autocorrelation effects may also allow more parsimonious models to be built, because the quantity of environmental information required may fall (Bivand 1984). All of these could have real benefits in conservation biology.

Field study design may be little affected by a decision to model autocorrelation should it be detected in the data set; the major differences are reserved for the model-development stage. Nevertheless, it is worth considering how best to locate sampling sites with respect to future autocorrelation considerations. Primarily, the concern should be to obtain a large sample of sites over a wide range of intersite distances so that the analytical

Table 3. Summary of major recommendations for improved field-study design and the potential benefits.

<i>Recommendation</i>	<i>Potential advantages</i>
Use of variables showing a direct relationship with the organism wherever possible	improved predictive ability, especially over large geographical extents or in predicting responses to environmental change, greater ease of interpretation
Consideration of interacting species	improved predictive ability, greater biological validity (modeling of realized niche), greater explanatory power and ease of interpretation
Identification of complete geographical region of interest, prior to sampling	improved predictive ability with new data because model does not need to extrapolate to new conditions, explanatory conclusions more widely applicable
Environmental stratification, with equal division of samples between strata	improved predictive ability, more accurate explanatory analysis
Multiscale approach to sampling	improved predictive ability, greater explanatory understanding, greater relevance to conservation planning
Aim to sample at least 10 sites for every environmental variable considered	more reliable model development and explanatory analysis, improved predictive ability
Greater awareness of effects of prevalence in presence/absence modeling	
Aim to model spatial autocorrelation, where present; test to ensure adequate statistical power for autocorrelation analyses in design of sampling scheme	facilitated detection, characterization and subsequent modeling of autocorrelation, improved understanding of mechanisms generating distribution pattern, greater predictive accuracy
Partition resources to collect independent evaluation data; environmental stratification used in the collection process	essential to test distribution models, thereby increasing scientific rigor of observational analyses, essential to have idea of model generality and predictive ability, environmental stratification ensures that evaluation data will provide a thorough test

process is powerful and can accurately quantify the distance over which autocorrelation is observed. There are no absolute rules for sample sizes, but the reliability of the analysis increases with sample size (Sokal & Oden 1978). The irregular location of sites within environmental stratification is advantageous in this respect because regularly spaced sites may fall in or out of phase with environmental patterns, rendering their analytical power weak (Fortin et al. 1989). It is a straightforward piece of geometry to calculate the distances between all possible pairings of the sampling sites once the coordinates are known, and this allows a simple confirmation of good coverage across a range of intersite distances. Where there are obvious weaknesses in the coverage, attempts should be made to adjust site positions to remedy the situation.

Once the data are collected, simple tests will indicate the presence of autocorrelation. Correlograms, showing how autocorrelation statistics such as Moran's *I* vary with intersite distance, are the most valuable tools. Details of such tests are given by Sokal and Oden (1978), Cliff and Ord (1981), and Legendre and Fortin (1989), and examples of simple autocorrelation testing on training data have been given by Beard et al. (1999), Villard et al. (1999), Higgins et al. (1999), Fisher et al. (2002), and Lobo and Martin-Piera (2002). If autocorrelation patterns are evident, the environmental variables should then be compared to see whether the patterns could be accounted for solely by autocorrelation in the environment. If so,

or if no autocorrelation is evident, spatial autocorrelation need not be considered further (Smith 1994).

Collecting Suitable Data with Which to Evaluate the Model

Thorough evaluation is an integral part of developing wildlife-habitat models; consequently, resources should always be allocated during fieldwork to collect data for this purpose. The correlative nature of distribution modeling demands that such evaluation with independent data be carried out to increase the scientific rigor of the work and confidence in the conclusions (Manel et al. 1999; Gaston & Blackburn 1999). Only through proper evaluation can the modeling problems in Table 1 be correctly identified. In the past, the standard of model evaluation was poor, with little evaluation based on new data (Manel et al. 2001), although the situation is improving with independent test data being used in an increasing number of studies. Such data, collected from a geographically discrete region, are the only valid test of a distribution model (Chatfield 1995). Prediction success with the training data has little value, whereas resampling techniques, such as bootstrapping and jack-knifing, may be affected by spatial autocorrelation between sites (Manel et al. 1999). Longer-range autocorrelation could also affect the training and test regions themselves, rendering evaluation data only

pseudoindependent. This risk is slight, however, compared with resampling approaches, in which sites in close proximity can be used for training and testing. A simple autocorrelation test can confirm the independence of an evaluation set from the training data.

Aside from geographical independence, evaluation data should be collected so as to provide the model with a thorough test. Although a random selection of sites allows basic evaluation, a much more effective approach is to systematically collect sites representing the full range of environmental space, mirroring collection of the training data. This gains particular importance in light of the fact that the evaluation set will generally contain fewer sites than the training set. Samples collected inside and outside the niche identified by the distribution model are important, testing both the overall predictive accuracy and the reliability and logic of the model in widespread application. The coverage of the evaluation data can be roughly assessed relative to the training data by ordinating the training and test data with principal components analysis and examining the ordination biplots produced. Regions of environmental space sampled in the training data, but missing in the evaluation set, become clear on simple biplots because the evaluation sites fail to completely overlap the training sites. Where gaps in coverage exist, there should be caveats to model application until such time as the model is tested in those conditions too.

Conclusions

Distribution modeling is a complex undertaking with specialized data requirements. Problems with the training data may be pervasive, contributing to a series of observed modeling problems (Table 1). By addressing these problems directly (Table 3), we hope that valuable improvements in the quality of training data can be achieved. Ultimately, improved predictive ability and better explanatory analysis should result, and greater confidence could be placed in both. Given the burgeoning use of distribution modeling in conservation biology, any such improvements stand to be very valuable.

Acknowledgments

We thank three anonymous reviewers and the editors for their helpful and insightful comments on the manuscript. Financial support for our work on distribution modeling is provided by the U.K. Environment Agency.

Literature Cited

- Apps, C. D., B. N. McLellan, T. A. Kinley, and J. P. Flaa. 2001. Scale-dependent habitat selection by mountain caribou, Columbia Mountains, British Columbia. *Journal of Wildlife Management* 65:65-77.
- Araújo, M. B., and P. H. Williams. 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* 96:331-345.
- Augustin, N. H., M. A. Muggleston, and S. T. Buckland. 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* 33:339-347.
- Austin, M. P. 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42:11-21.
- Austin, M. P., and P. C. Heyligers. 1989. Vegetation survey design for conservation: gradsect sampling of forests in north-eastern New South Wales. *Biological Conservation* 50:13-32.
- Austin, M. P., and J. A. Meyers. 1996. Current approaches to modeling the environmental niche of eucalypts: implications for management of forest biodiversity. *Forest Ecology and Management* 85:95-106.
- Austin, M. P., R. B. Cunningham, and P. M. Fleming. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55:11-27.
- Austin, M. P., A. O. Nicholls, and C. R. Margules. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* 60:161-177.
- Beard, K. H., N. Hengartner, and D. K. Skelly. 1999. Effectiveness of predicting breeding bird distributions using probabilistic models. *Conservation Biology* 13:1108-1116.
- Begon, M., J. L. Harper, and C. R. Townsend. 1996. *Ecology: individuals, populations and communities*. 3rd edition. Blackwell Science, Oxford, United Kingdom.
- Bivand, R. S. 1984. Regression modeling with spatial dependence: an application of some class selection and estimation methods. *Geographical Analysis* 16:25-37.
- Buckton, S. T., and S. J. Ormerod. 1997. Use of a new standardized habitat survey for assessing the habitat preferences and distribution of upland river birds. *Bird Study* 44:327-337.
- Carroll, S. S., and D. L. Pearson. 2000. Detecting and modeling spatial and temporal dependence in conservation biology. *Conservation Biology* 14:1893-1897.
- Carroll, C., W. J. Zielinski, and R. F. Noss. 1999. Using presence/absence data to build and test spatial habitat models for the Fisher in the Klamath Region, U.S.A. *Conservation Biology* 13:1344-1359.
- Chatfield, C. 1995. Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A*. 158:419-466.
- Clevenger, A. P., J. Wierzchowski, B. Chruszcz, and K. Gunson. 2002. GIS-generated, expert-based models for identifying habitat linkages and planning mitigation passages. *Conservation Biology* 16:503-514.
- Cliff, A. D., and J. K. Ord. 1981. *Spatial processes: models and applications*. Pion, London.
- Collingham, Y. C., R. A. Wadsworth, B. Huntley, and P. E. Hulme. 2000. Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology* 37(supplement 1):13-27.
- Cumming, G. S. 2000. Using between-model comparisons to fine-tune linear models of species range. *Journal of Biogeography* 27:441-455.
- Fielding, A. H. 1999. An introduction to machine learning methods. Pages 1-35 in A. H. Fielding, editor. *Machine learning methods for ecological applications*. Kluwer Academic Publishers, Boston.
- Fielding, A. H., and P. F. Haworth. 1995. Testing the generality of bird-habitat models. *Conservation Biology* 9:1466-1481.
- Fisher, R. N., A. V. Suarez, and T. J. Case. 2002. Spatial patterns in the abundance of the coastal horned lizard. *Conservation Biology* 16:205-215.
- Fortin, M.-J., P. Drapeau, and P. Legendre. 1989. Spatial autocorrelation and sampling design in plant ecology. *Vegetatio* 83:209-222.
- Franco, A. M. A., J. C. Brito, and J. Almeida. 2000. Modelling habitat selection of Common Cranes *Grus grus* wintering in Portugal using multiple logistic regression. *Ibis* 142:351-358.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19:474-499.

- Freedman, L. S., D. Pee, and D. N. Midthune. 1992. The problem of underestimating the residual error variance in forward stepwise regression. *The Statistician* 41:405-412.
- Gainzarain, J. A., R. Arambarria, and A. F. Rodríguez. 2000. Breeding density, habitat selection and reproductive rates of the Peregrine Falcon *Falco peregrinus* in Álava (northern Spain). *Bird Study* 47:225-231.
- Gaston, K. J., and T. M. Blackburn. 1999. A critique for macroecology. *Oikos* 84:353-368.
- Gillison, A. N., and K. R. W. Brewer. 1985. The use of gradient directed transects or gradsects in natural resource surveys. *Journal of Environmental Management* 20:103-127.
- Guisan, A., and J.-P. Theurillat. 2000. Assessing alpine plant vulnerability to climate change: a modeling perspective. *Integrated Assessment* 1:307-320.
- Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.
- Guisan, A., J.-P. Theurillat, and F. Kienast. 1998. Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science* 9:65-74.
- Harrell, F. E., K. L. Lee, R. M. Califf, D. B. Pryor, and R. A. Rosati. 1984. Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine* 3:143-152.
- Harrell, F. E., K. L. Lee, and D. B. Mark. 1996. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 15:361-387.
- Higgins, S. I., D. M. Richardson, R. M. Cowling, and T. H. Trinder-Smith. 1999. Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. *Conservation Biology* 13:303-313.
- Hinsley, S. A., P. E. Bellamy, I. Newton, and T. H. Sparks. 1995. Habitat and landscape factors influencing the presence of individual breeding bird species in woodland fragments. *Journal of Avian Biology* 26:94-104.
- Hirzel, A., and A. Guisan. 2002. Which is the optimal strategy for habitat suitability modelling. *Ecological Modelling* 157:331-341.
- Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology* 22:415-427.
- Johnson, C. J., K. L. Parker, and D. C. Heard. 2001. Foraging across a variable landscape: behavioural decisions made by woodland caribou at multiple spatial scales. *Oecologia* 127:590-602.
- Koenig, W. D. 1999. Spatial autocorrelation of ecological phenomena. *Trends in Ecology & Evolution* 14:22-26.
- Krummel, J. R., R. H. Gardner, G. Sugihara, R. V. O'Neill, and P. R. Coleman. 1987. Landscape patterns in a disturbed environment. *Oikos* 48:321-324.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659-1673.
- Legendre, P., and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80:107-138.
- Lobo, J. M., and F. Martin-Piera. 2002. Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables. *Conservation Biology* 16:158-173.
- Lowe, W. H., and D. T. Bolger. 2002. Local and landscape-scale predictors of salamander abundance in New Hampshire headwater streams. *Conservation Biology* 16:183-193.
- Mackey, B. G., and D. B. Lindenmayer. 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography* 28:1147-1166.
- Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: the distinction between—and reconciliation of—'predictive' and 'explanatory' models. *Biodiversity and Conservation* 9:655-671.
- Manel, S., J. M. Dias, S. T. Buckton, and S. J. Ormerod. 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology* 36:734-747.
- Manel, S., S. T. Buckton, and S. J. Ormerod. 2000. Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan rivers. *Journal of Applied Ecology* 37:756-770.
- Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38:921-931.
- Margules, C. R., and M. P. Austin. 1994. Biological models for monitoring species decline: the construction and use of data bases. *Philosophical Transactions of the Royal Society of London Series B* 344:69-75.
- Margules, C. R., and J. L. Stein. 1989. Patterns in the distributions of species and the selection of nature reserves: an example from the *Eucalyptus* forests in south-eastern New South Wales. *Biological Conservation* 50:219-238.
- Mladenoff, D. J., T. A. Sickley, and A. P. Wydeven. 1999. Predicting gray wolf landscape recolonization: logistic regression models vs. new field data. *Ecological Applications* 9:37-44.
- Mohler, C. L. 1983. Effect of sampling pattern on estimation of species distributions along gradients. *Vegetatio* 54:97-102.
- Neave, H. M., T. W. Norton, and H. A. Nix. 1996. Biological inventory for conservation evaluation. I. Design of a field-survey for diurnal, terrestrial birds in southern Australia. *Forest Ecology and Management* 85:107-122.
- Nugent, G., W. Fraser, and P. Sweetapple. 2001. Top down or bottom up? Comparing the impacts of introduced arboreal possums and 'terrestrial' ruminants on native forests in New Zealand. *Biological Conservation* 99:65-79.
- Orians, G. H., and J. F. Wittenberger. 1991. Spatial and temporal scales in habitat selection. *The American Naturalist* 137(supplement):29-49.
- Orrock, J. L., J. F. Pagels, W. J. McShea, and E. K. Harper. 2000. Predicting presence and abundance of a small mammal species: the effect of scale and resolution. *Ecological Applications* 10:1356-1366.
- Özesmi, S. L., and U. Özesmi. 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling* 116:15-31.
- Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133:225-245.
- Peduzzi, P., J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49:1373-1379.
- Poizat, G., and D. Pont. 1996. Multi-scale approach to species-habitat relationships: juvenile fish in a large river. *Freshwater Biology* 36:611-622.
- Ripley, B. D. 1994. Neural networks and related methods for classification. *Journal of the Royal Statistical Society Series B* 56:409-456.
- Robertson, A. W., D. Kelly, J. J. Ladley, and A. D. Sparrow. 1999. Effects of pollinator loss on endemic New Zealand mistletoes (Loranthaceae). *Conservation Biology* 13:499-508.
- Rushton, S. P., G. W. Barreto, R. M. Cormack, D. W. Macdonald, and R. Fuller. 2000. Modelling the effects of mink and habitat fragmentation on the water vole. *Journal of Applied Ecology* 37:475-490.
- Saunders, D. A., R. J. Hobbs, and G. W. Arnold. 1993. The Kellerberrin project on fragmented landscapes: a review of current information. *Biological Conservation* 64:185-192.
- Smith, P. A. 1994. Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters* 4:47-61.
- Söderström, B., and T. Pärt. 2000. Influence of landscape scale on farmland birds breeding in semi-natural pastures. *Conservation Biology* 14:522-533.
- Sokal, R. R., and N. L. Oden. 1978. Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society* 10:199-228.
- Steyerberg, E. W., M. J. C. Eijkemans, and J. D. F. Habbema. 1999. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology* 52:935-942.

- Vander Haegen, W. M., F. C. Dobler, and D. J. Pierce. 2000. Shrubsteppe bird response to habitat and landscape variables in eastern Washington, USA. *Conservation Biology* 14:1145-1160.
- Veblen, T. T., M. Mermoz, C. Martin, and T. Kitzberger. 1992. Ecological impacts of introduced animals in Nahuel-Huapi-National-Park, Argentina. *Conservation Biology* 6:71-83.
- Villard, M. A., M. K. Trzcinski, and G. Merriam. 1999. Fragmentation effects on forest birds: relative influence of woodland cover and configuration on landscape occupancy. *Conservation Biology* 13:774-783.
- Wessels, K. J., A. S. Van Jaarsveld, J. D. Grimbeek, and M. J. Van Der Linde. 1998. An evaluation of the gradsect biological survey method. *Biodiversity and Conservation* 7:1093-1121.
- Westman, W. E. 1991. Measuring realized niche spaces: climatic response of chaparral and coastal sage scrub. *Ecology* 72:1678-1684.
- Wiens, J. A. 1989. Spatial scaling in ecology. *Functional Ecology* 3:385-397.
- Wiens, J. A., J. T. Rotenberry, and B. Van Horne. 1987. Habitat occupancy patterns of North American shrubsteppe birds: the effects of spatial scale. *Oikos* 48:132-147.
- Wikramanayake, E. D., E. Dinerstein, J. G. Robinson, U. Karanth, A. Rabinowitz, D. Olson, T. Mathew, P. Hedao, M. Conner, G. Hemley, and D. Bolze. 1998. An ecology-based method for defining priorities for large mammal conservation: the tiger as a case study. *Conservation Biology* 12:865-878.
- Wiser, S. K., R. K. Peet, and P. S. White. 1998. Prediction of rare-plant occurrence: a southern Appalachian example. *Ecological Application* 8:909-920.
- Wolf, A., C. K. Nielson, T. Weber, and T. J. Gibbs-Kieninger. 2002. Statewide modeling of bobcat, *Lynx rufus*, habitat in Illinois, USA. *Biological Conservation* 104:191-198.
- Yee, T. W., and N. D. Mitchell. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2:587-602.
- Zacharias, M. A., and J. C. Roff. 2000. A hierarchical approach to conserving marine biodiversity. *Conservation Biology* 14:1327-1334.

