Available online at www.sciencedirect.com

SCIENCE DIRECT®

Ecological Modelling 192 (2006) 499–522

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions—possibilities and limitations

Jana M. McPherson[a,*], Walter Jetz[a,b,c,d], David J. Rogers[a]

[a] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
[b] Division of Biological Sciences, University of California San Diego, 9500 Gilman Drive MC 0116, La Jolla, CA 92093-0116, USA
[c] Ecology and Evolutionary Biology Department, Princeton University, Princeton, NJ 08544-1003, USA
[d] Biology Department, University of New Mexico, Albuquerque, NM 87131, USA

Received 12 October 2004; received in revised form 17 July 2005; accepted 3 August 2005
Available online 26 September 2005

Abstract

For effective conservation planning, resource managers around the world require information on species distributions. Available data, however, are often too coarse in spatial resolution to be of practical use. We describe four modelling techniques that harness the information contained in coarse-scale occurrence records to predict species' distributions at the finer resolutions relevant to conservation officers in the field. The techniques are primarily statistical in nature, but can incorporate expert knowledge on species' habitat preferences. We highlight the conceptual strengths and weaknesses of each, and perform a preliminary test of their comparative performance, using empirical data on the distribution of nine bird species in Uganda and satellite-derived environmental indices. Two techniques, the point sampling approach and iterative approach, yield encouraging results. Because the dominance of environmental factors controlling species' distributions shifts with spatial scale, however, cross-scale predictions remain a formidable challenge. Further testing and refinement of the techniques we present here are undoubtedly necessary before they can be applied with any confidence.

Keywords: AUC; Cross-scale prediction; Discriminant analysis; Expert knowledge; Satellite imagery; Spatial scale; Species distribution models

---

1. Introduction

Humanity's growth and pursuit of ever higher living standards have environmental consequences at regional and global scales, including wide-spread habitat fragmentation, the introduction of non-native species and global climate change. To understand and mitigate such

---

[*] Corresponding author. Present address: Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, NS, Canada B3H 4J1. Tel.: +1 902 494 3910.
E-mail address: jana.mcpherson@dal.ca (J.M. McPherson).

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2005.08.007

wide-ranging alterations to the world we live in and depend on, we must assess and predict their ecological implications (Root and Schneider, 1995).

Our ability to understand how ecosystems function and respond to change is complicated, however, by issues of scale, both spatial and temporal, for at least three reasons. Firstly, observed patterns – the primary cue in any analysis of cause and effect – are in some sense simply an artefact of the scale of analysis. As scale changes, patterns change (Levin, 1992; Riitters et al., 1997; Thompson and McGarigal, 2002; Hobbs, 2003; Vaughan and Ormerod, 2003). Secondly, patterns observed at any one scale may be the result of processes operating at another, or indeed a whole suite of scales (Levin, 1992; Huston, 2002). Individual species, and therefore their communities, are often influenced by factors at multiple scales (MacFaden and Capen, 2002; Thompson and McGarigal, 2002; Sergio et al., 2003), and the scale of response can vary from organism to organism (DeGraaf et al., 1998; Orrock et al., 2000; Mitchell et al., 2001; MacFaden and Capen, 2002). Thirdly, the scales at which ecologists, climate modellers, politicians and wildlife managers operate are traditionally quite dissimilar, hindering effective interdisciplinary exchange and communication (Root and Schneider, 1995; Lindenmayer, 2000; Scott et al., 2002; Hobbs, 2003). Consequently, methods that can bridge scale gaps and reliably translate findings at one scale to predictions at another would be highly desirable (Root and Schneider, 1995; Hobbs, 2003).

A particular challenge, indeed a task long deemed impossible, is the generation of fine-grained information from coarse-resolution maps (Hartley et al., 2004). As a growing number of nations strive to document their natural heritage, atlases detailing the distributions of species across countries and continents are becoming increasingly available. Yet the spatial resolution of such atlases tends to be coarse relative to the scales at which conservation strategies must be implemented (Kunin et al., 2000; Hartley et al., 2004). Wildlife management, forestry planning and environmental impact assessments generally operate at scales of several hundred hectares (Böhning-Gaese, 1997; Ferrier et al., 2002; Bustamante and Seoane, 2004). In contrast, taxonomic atlases typically map species distributions at resolutions ranging from 100 km^2 in well-studied regions (e.g. butterflies and birds in the UK: Hill et al., 1999; Lennon et al., 2000) to more than 10,000 km^2 in poorly surveyed parts of the world (e.g. Pan-African databases on vertebrates and plants: Lovett et al., 2000; Balmford et al., 2001; Jetz and Rahbek, 2002).

Can such coarse data be used to map species distributions at the finer scales at which data are needed for conservation planning? Species' distributions have been postulated to exhibit some degree of self-similarity across scales (e.g. Kunin, 1998; Harte et al., 1999). Yet it is thought that the dominant processes influencing species' distributions shift from abiotic factors at coarse scales to biotic interactions at high spatial resolution (Root, 1988; Cumming, 2002; Kadmon et al., 2003). This shift in factors determining distributional patterns is likely to complicate cross-scale predictions.

Perhaps not surprisingly, therefore, attempts to derive fine-grained distribution maps from coarse-grained occurrence records have been rare and results mixed (Lloyd and Palmer, 1998; Collingham et al., 2000; Barbosa et al., 2003; Araújo et al., 2005). The challenge lies in identifying environmental correlates of species' distributions at coarse scales that also capture niche requirements at finer grain. Here we describe four alternative approaches to this problem. Making use of the exceptionally high resolution of the *Bird Atlas of Uganda* (Carswell et al., 2005) and satellite-derived environmental indices, we then test each approach's ability to convert coarse occurrence data (squares of 0.25° longitude/latitude) into high-resolution predictions (squares of 0.01° longitude/latitude). In the process, we also assess whether the performance of empirical models can be improved by the incorporation of expert knowledge on habitat associations.

2. Materials and methods

Models to turn coarse data into fine-scale predictions were constructed for nine bird species in Uganda. The four approaches to cross-scale predictions are described in Section 3. Here, we provide details on the bird distribution data and environmental predictors used by all four approaches. We also describe non-linear discriminant analysis, the statistical algorithm used by three of the four approaches, as well as the process used to evaluate and gauge the performance of models derived with each approach.

2.1. Bird distribution data

Occurrence records for birds were taken from data compiled for the *Bird Atlas of Uganda* (Carswell et al., 2005). This atlas is unique among bird atlas projects in southern and eastern Africa in that species occurrence was referenced to the precise point of observation (with decimal longitude and latitude) rather than a geographic grid with resolutions of 0.25–0.50° (as, for example, in *A Bird Atlas of Kenya* by Lewis and Pomeroy, 1989; *The Atlas of Southern African Birds* by Harrison et al., 1997; and *The Atlas of the Birds of Sul do Save, Southern Mozambique* by Parker, 1999). It thus provides a unique opportunity to test high resolution predictions.

To simulate the low resolution of avian distribution records elsewhere, the Ugandan data were coarsened to 0.25° (1/4°) and intermediate resolutions of 1/8°, 1/16°, 1/32°, 1/64° and 1/100°. At each resolution, a species was designated as present in any grid square containing at least one point locality observations of that species. Because sampling for the *Bird Atlas of Uganda* was not geographically systematic (Tushabe et al., 2000, 2001), however, grid squares lacking observations of the species were not automatically assumed to represent absence. Only if a minimum number of observations of other species had been logged within the square (e.g. 75+ observations at 1/4° resolution), was the absence of a record interpreted as true absence of the species. Squares of ambiguous status were excluded from both model calibration and evaluation.

Given the patchy nature of point locality observations in Uganda, aggregation of data from fine to coarse resolution led to coarse data with a level of observer effort comparable to other bird atlases in the region. Ugandan 1/4° squares with data on average drew on observations from 10.24 point localities. For comparison, the average number of records per 1/4° square in southern Mozambique was 5.11, in Zimbabwe 21.22, in South Africa, Lesotho and Swaziland 56.82 (Harrison et al., 1997; Parker, 1999).

Nine bird species were chosen for analysis, based on the number of quarter-degree squares (QDS) they occupied (Table 1). Three species had very narrow ranges (occupying 3–12 out of a total of 333 QDS), three had intermediate ranges (43–64 QDS) and three were among the most widespread in Uganda (98–136 QDS).

Habitat preferences of these species were extracted from Sibley and Monroe (1990, 1993). Information on each species' altitudinal range was compiled from *The Birds of Africa* (Brown et al., 1982), *The Handbook of the Birds of the World* (del Hoyo et al., 1992) and *The Field Guide to the Birds of East Africa* (Stevenson and Fanshawe, 2001).

2.2. Environmental variables

Environmental correlates were primarily derived from satellite images collected twice daily over a 5-year period (1992–1996) by the U.S. National Oceanic and Atmospheric Administration's Advanced High Resolution Radiometer satellite series. Environmental information obtained from these images included a middle infrared signal, indices of land surface temperature, air temperature, the vapour pressure deficit and the normalised difference vegetation index (NDVI). A further index, cold cloud duration, was derived from 10 years (1989–1998) of European Meteosat imagery. For more information on these indices and their procurement, see Table A1 in Appendix A. All imagery was composited into cloud-free, monthly images and re-sampled from its original spatial resolution of 1 km^2 to resolutions of 1/4°, 1/8°, 1/16°, 1/32°, 1/64° and 1/100°. For each environmental index, we used temporal Fourier analysis, a data reduction technique ideal for summarising seasonal variables (Chatfield, 1996; Rogers et al., 1996), to extract the overall mean, minimum, maximum and variance, plus the amplitude (strength) and phase (timing) of annual, biannual and triannual cycles. Furthermore, altitude, derived from the U.S. Geological Survey's (USGS) global digital elevation model (http://edcdaac.usgs.gov/gtopo30/gtopo30.html), was included among the explanatory variables, yielding a total of 61 candidate predictors (see Table A1 in Appendix A).

Models incorporating expert knowledge on species' habitat preferences also used the USGS Land Cover System map (Version 2; http://edcdaac.usgs.gov/glcc/tabgoode_af.html), re-sampled to 0.01° by nearest neighbour analysis (using IDRISI). USGS land cover categories were paired with species' habitat preferences as shown in Table 2.

Table 1
The nine bird species whose recorded distributions in Uganda were used to test modelling approaches that rely on coarse-resolution data to derive fine-resolution predictions of species' occurrence

Species	Local range	Preferred habitat	Observed habitat
Nahan's Francolin (*Pternistis nahani*, Phasanidae)	3	Humid forest (1000–1400 m)	Evergreen broadleaf forest, *savanna*, *dryland cropland and pasture* (1049–1283 m)
Fox's Weaver (*Ploceus spekeoides*, Passeridae)	6	Swamps (no altitude given)	*Savanna*, *dryland cropland and pasture* (1046–1219 m)
Crested Guineafowl (*Guttera pucherani*, Numididae)	12	Humid forest, edge (0–2200 m)	Evergreen broadleaf forest, *savanna*, *dryland cropland and pasture*, shrubland, cropland-woodland mosaic (533–1524 m)
Lesser Honeyguide (*Indicator minor*, Indicatridae)	43	Riparian woodland, savanna (0–3000 m)	*Savanna*, *dryland cropland and pasture*, evergreen broadleaf forest, cropland-woodland mosaic, shrubland, *grassland*, water bodies, *barren or sparsely vegetated* (719–2433 m)
Northern Puffback (*Dryoscopus gambensis*, Corvidae)	54	Woods (900–2850 m)	Savanna, *dryland cropland and pasture*, *grassland*, cropland-woodland mosaic, evergreen broadleaf forest, shrubland, *urban/built-up*, *water bodies* (636–2321 m)
Yellow-rumped Tinkerbird (*Pogoniulus bilineatus*, Lybiidae)	64	Forest, edge, savanna, scrub (0–3000 m)	Dryland cropland and pasture, cropland-woodland mosaic, savanna, evergreen broadleaf forest, shrubland, grassland, *water bodies*, deciduous broadleaf forest, *barren or sparsely vegetated*, mixed forest, *urban/built up* (533–2809 m)
Bronze Munia (*Spermestes cucullatus*, Passeridae)	98	Savanna, scrub, farmland (0–2200 m)	Dryland cropland and pasture, savanna, cropland-woodland mosaic, *evergreen broadleaf forest, water bodies*, barren or sparsely vegetated, grassland, shrubland, *urban/built-up*, *deciduous broadleaf forest* (494–2433 m)
Green-backed Camaroptera (*Camaroptera brachyura*, Priniidae)	110	Humid forest undergrowth, thickets (0–2200 m)	Savanna, *dryland cropland and pasture*, cropland-woodland mosaic, *water bodies*, evergreen broadleaf forest, *grassland,* shrubland, *barren or sparsely vegetated*, deciduous broadleaf forest, mixed forest, *urban/built-up* (533–2228 m)
Garden Bulbul (*Pycnonotus barbatus*, Pycnonoidae)	136	Woods, forest, secondary growth, towns, usually near water (0–3000 m)	Savanna, *dryland cropland and pasture*, water bodies, cropland-woodland mosaic, shrubland, evergreen broadleaf forest, *barren or sparsely vegetated*, mixed forest, deciduous braodleaf forest (427 - 4536 m)

Species were selected based on their local range size (here the number of quarter-degree squares occupied by the species in Uganda), so as to include species of low, intermediate and high prevalence in Uganda. Also indicated are each species' published habitat preferences, as well as the USGS land cover classes and altitudes observed at point locality occurrence records. Land cover classes are listed in order of the frequency with which they corresponded to species' point localities. Italics indicate observed land cover types and altitudes that fell outside the species' published preferences.

2.3. *Statistical algorithm*

To distinguish the environmental characteristics associated with species' presence and species' absence, several of the modelling approaches described below used non-linear discriminant analysis. Alternative algorithms, such as logistic regression, would have been equally applicable but were not tested in this study.

In discriminant analysis, training data serve to determine the multivariate mean and variance–covariance structure of predictor variables for each of the response variable's states, here species presence and absence. The distribution of predictor variables is assumed to be normal, but their covariance need not be the same for all states in non-linear discriminant analysis (Rogers et al., 1996). The posterior probability of any data point belonging to one response state or another is then calculated based on its position in *n*-dimensional space relative to each state's multivariate mean, where distance between sample point and mean

Table 2
Habitat types mentioned by Sibley and Monroe (1990, 1993) as appropriate for the nine species analysed, and the USGS land cover types thought to represent these habitats in Uganda

Published habitat	Corresponding USGS landcover types that exist in Uganda
Humid forest or humid forest undergrowth	Deciduous broadleaf forest, evergreen broadleaf forest, mixed forest, wooded wetland, cropland-woodland mosaic
Forest or woods	Deciduous broadleaf forest, evergreen broadleaf forest, mixed forest, wooded wetland, cropland-woodland mosaic, shrubland, savanna
Riparian woodlands	Deciduous broadleaf forest, evergreen broadleaf forest, mixed forest, wooded wetlands, cropland-woodland mosaic, shrubland, water bodies
Secondary growth	Deciduous broadleaf forest, evergreen broadleaf forest, mixed forest, wooded wetlands, cropland-woodland mosaic, shrubland
Thicket	Deciduous broadleaf forest, evergreen broadleaf forest, mixed forest, cropland-woodland mosaic, shrubland, savanna
Scrub	Shrubland, savanna
Savanna	Savanna, shrubland, grassland, cropland-grassland mosaic, cropland-woodland mosaic, dryland cropland and pasture
Edge	Crop-land-woodland mosaic, shrubland, mixed shrubland-grassland
Farmland	Dryland cropland and pasture, cropland-grassland mosaic, cropland-woodland mosaic, grassland, barren or sparsely vegetated
Swamp	Water bodies, wooded wetland
Near water	Water bodies
Town	Urban/built-up

is measured as Mahalanobis distance (Green, 1978; Rogers et al., 1996). For presence–absence data, discriminant analysis predicts a species' probability of occurrence.

We implemented non-linear discriminant analysis in custom-written programs (QuickBasic). Predictor variables were selected in forward stepwise fashion based on their ability to increase kappa, a chance-corrected measure of model fit (Cohen, 1960). Stepwise variable selection, although often criticised, was used because (1) its use is widespread and thus represents a reality in distribution modelling and (2) we were not concerned with an ecological interpretation of the models. Our models guarded against unrealistic outcomes by restricting predictions to sites environmentally similar (as indicated by the Mahalanobis distance) to those included in training data.

2.4. Model evaluation

Models were evaluated based on how well they predicted test data, i.e. data not used in model training. In models trained at coarse resolution, test and training data were not strictly independent, because both datasets derived from the point locality observations recorded in the *Bird Atlas of Uganda*. Yet test data were of finer spatial resolution and therefore novel to these models.

Accuracy was measured as the area under the curve (AUC) of receiver operating characteristics (ROC) plots, here calculated non-parametrically using the Wilcoxon statistic (Hanley and McNeil, 1982; Pearce and Ferrier, 2000). AUC facilitates meaningful comparisons between models of species with different range sizes, because it is largely insensitive to species' prevalence (McPherson et al., 2004). It can, however, become erratic at extremely low or high prevalence, so the proportion of presence samples in test data was constrained to fall within 0.35–0.65. Where necessary, this was achieved by sub-sampling records in the more abundant category, either presence or absence.

Following Swets (1988), we considered model accuracy good if $0.9 < \text{AUC} \leq 1.0$, reasonable if $0.7 < \text{AUC} \leq 0.9$, and poor if $0.5 < \text{AUC} \leq 0.7$. As AUC values of 0.5 correspond to random performance, this cut-off served as a worst-case null model for high-resolution predictions.

Comparisons of accuracy achieved by the different modelling approaches were undertaken non-parametrically with Wilcoxon' signed rank tests for matched pairs or Friedman's test for related samples (McClave and Dietrich, 1994).

2.5. Gauging expectations: fine-grained models

Models trained at coarse resolution cannot be expected to predict fine-grained patterns any better than models trained directly at high resolution. To gauge expectations for cross-scale models, we therefore calibrated one model per species with fine-grained data.

These *fine* models were trained at a resolution of 1/64° using non-linear discriminant analysis. Two-thirds of the 1/64° data were used for calibration, the remainder for model evaluation. The AUC achieved served as an ideal-case scenario, indicating the maximum accuracy expected from models trained at coarse resolution.

3. Four approaches to cross-scale predictions

3.1. Direct approach

Most avian distribution data for eastern and southern Africa have a coarser spatial resolution than available remotely sensed environmental data. Any grid square in the region's biodiversity atlases, therefore, contains many component 'pixels' (picture elements), each of which has a unique set of environmental attributes. To train distribution models at the atlas' original resolution (e.g. 0.25°), pixel-level attributes within each atlas square must be amalgamated, for example, by averaging. Model algorithms, such as discriminant analysis or logistic regression, can then identify the statistical relationship between species occurrence and the corresponding averaged environmental conditions.

The direct approach simply projects the statistical relationship thus identified onto individual pixels, applying parameters calibrated at coarse grain to fine-grained environmental attributes (Fig. 1). The method is intuitive and has previously been used by other authors (Collingham et al., 2000; Barbosa et al., 2003; Araújo et al., 2005).

The direct approach was implemented here using non-linear discriminant analysis, a training resolution (atlas square size) of 1/4° and a pixel size for predic-

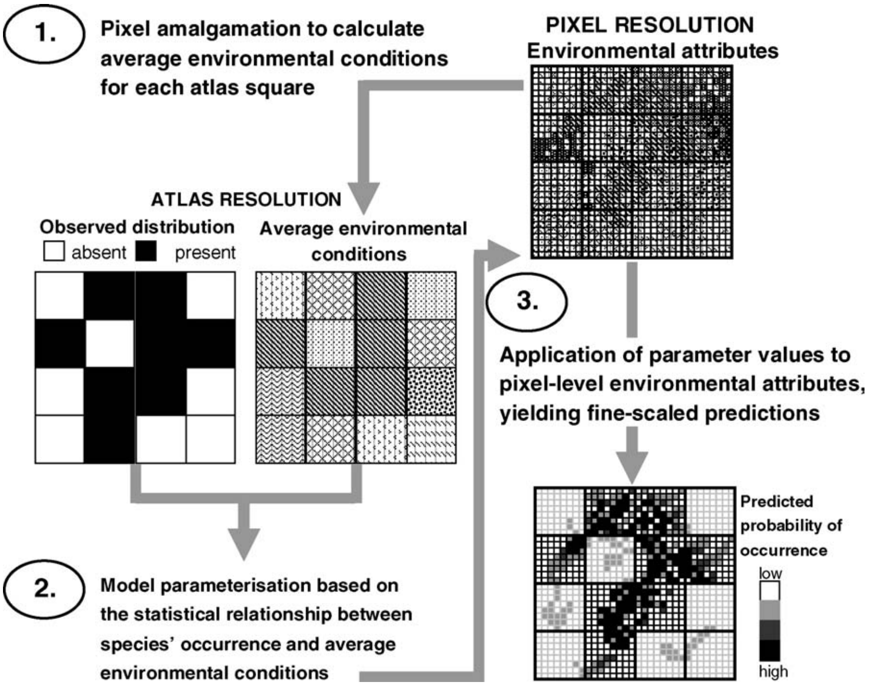


Fig. 1. The direct approach to high resolution predictions. *Step* 1 coarsens fine-grained environmental variables to the spatial resolution of available species' distribution records. *Step* 2 establishes the statistical relationship between these coarse environmental conditions and the species' presence and absence. *Step* 3 utilises parameter values gained in step 2 in conjunction with the original fine-grained environmental variables to produce high-resolution predictions of the species' distribution.

tions of 1/64°. Training sample prevalence (the proportion of presence records in training data) was coerced to 0.5, because this appears to maximise algorithmic performance in discriminant analysis (McPherson et al., 2004). To achieve 0.5 prevalence, the more abundant category (e.g. absence) was sub-sampled randomly to select only as many localities as were available in the less abundant category (e.g. presence).

Unfortunately, the direct approach suffers conceptual drawbacks. The variability among pixels' environmental attributes within a square may exceed the variability among squares' average conditions. Furthermore, the attributes of any one pixel within an atlas square are unlikely to match the square's average for each environmental variable (Huston, 2002). Model parameters thus risk being applied to environmental conditions (variable values and combinations thereof), not encountered during model training. Under such circumstances, predictions may be unrealistic or at least unreliable.

The following three alternative approaches to fine-grained predictions attempt to circumvent this problem.

3.2. *Iterative approach*

The iterative approach is closely related to the direct approach, but attempts to minimise the discrepancy between the environmental values encountered during calibration and prediction. It moves from coarse (atlas square) to fine (pixel) resolution incrementally, doubling the spatial resolution of predictions at each step. Models are first calibrated at the original resolution of atlas squares (e.g. 1/4°). If the model performs satisfactorily at that resolution, its parameters are used to make predictions at the next finer resolution (1/8°). These predictions are classified as predictions of presence or absence using a predicted probability of occurrence of 0.5 as the threshold. The outcome serves as training data for a new model, re-calibrated at double the starting resolution (i.e. at 1/8°), and the process begins anew (Fig. 2).

The underlying rationale is that the average condition in a group of pixels is closer in value to the average condition in half the group than to the attribute of any individual group member. The approach assumes, therefore, that models calibrated at coarse grain make decent predictions at finer grain if the difference between resolutions is not too large.

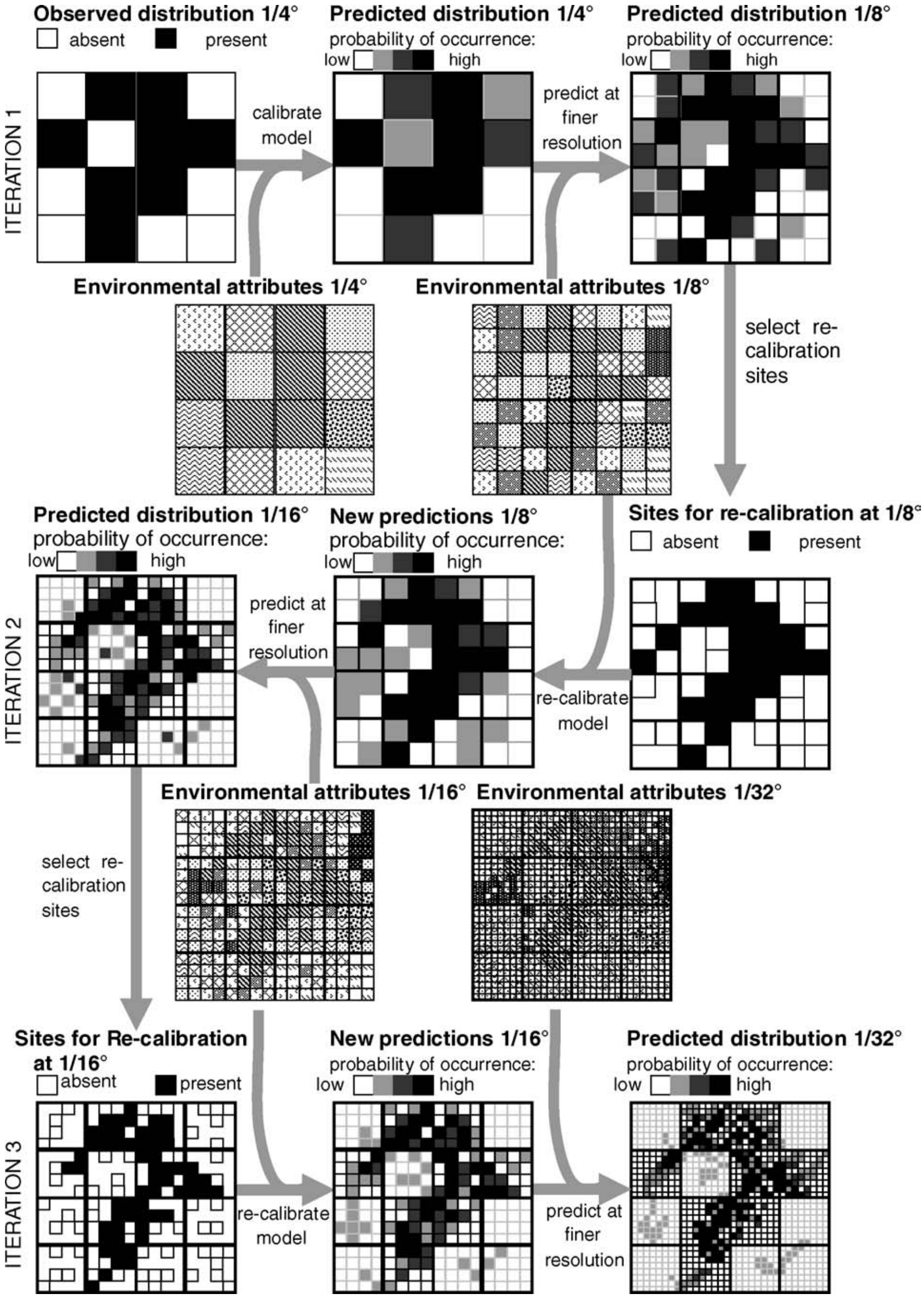
The iterative approach was implemented with a starting resolution of 1/4°, cycling through four rounds to predictions at 1/64°. Non-linear discriminant analysis served as the model algorithm. At all resolutions, training sample prevalence was coerced to 0.5, via sub-sampling. As a safeguard against error propagation from one iteration to the next, presence localities used to calibrate the models in iterations 2–4 had to fall within atlas squares known to harbour the species at the original (1/4°) resolution. No analogous constraint was placed on absence localities, since fine-resolution areas of absence can legitimately occur within coarse-resolution squares of presence. Predictions at each resolution were evaluated with data of equal resolution. They were also evaluated with data at the original, coarse resolution (1/4°) to examine the adequacy of such a test, which might be performed when no fine-resolution data are available (hereafter 'coarse-grain test').

Iterative models were built both purely empirically (*pure* models) and under inclusion of expert knowledge on habitat associations (*habitat* models). In *habitat* models, presence squares selected to re-calibrate models at finer resolutions had to contain an appropriate land cover category. An additional requirement for appropriate altitude did not yield significantly different results, so is not discussed further.

3.3. *Point sampling approach*

The point sampling approach avoids the amalgamation of pixel-level attributes. It randomly chooses, within each atlas square, a fixed number of environmental pixels to calibrate models at high resolution directly (Fig. 3). Lloyd and Palmer (1998) employed this approach in a study of South African bulbuls, albeit sampling only a single (the central) pixel per atlas square. Although intuitive, the approach makes the clearly unrealistic assumption that all component pixels of atlas squares known to harbour a species embody suitable environmental conditions, and those of squares not housing the species unsuitable environmental conditions.

In preliminary analyses the approach seemed insensitive to its starting resolution (1/2°, 1/4° or 1/8°) and the number of pixels sampled per atlas square (10, 50, 100). Analyses reported here used a starting resolution of 1/4° and pixel size of 1/100°. Again, non-linear dis-


ITERATION 1
Observed distribution 1/4°
absent
present
calibrate model
Predicted distribution 1/4°
probability of occurrence:
low
high
predict at finer resolution
Predicted distribution 1/8°
probability of occurrence:
low
high
Environmental attributes 1/4°
Environmental attributes 1/8°
select re-calibration sites
ITERATION 2
Predicted distribution 1/16°
probability of occurrence:
low
high
predict at finer resolution
New predictions 1/8°
probability of occurrence:
low
high
re-calibrate model
Sites for re-calibration at 1/8°
absent
present
Environmental attributes 1/16°
Environmental attributes 1/32°
select re-calibration sites
ITERATION 3
Sites for Re-calibration at 1/16°
absent
present
re-calibrate model
New predictions 1/16°
probability of occurrence:
low
high
predict at finer resolution
Predicted distribution 1/32°
probability of occurrence:
low
high

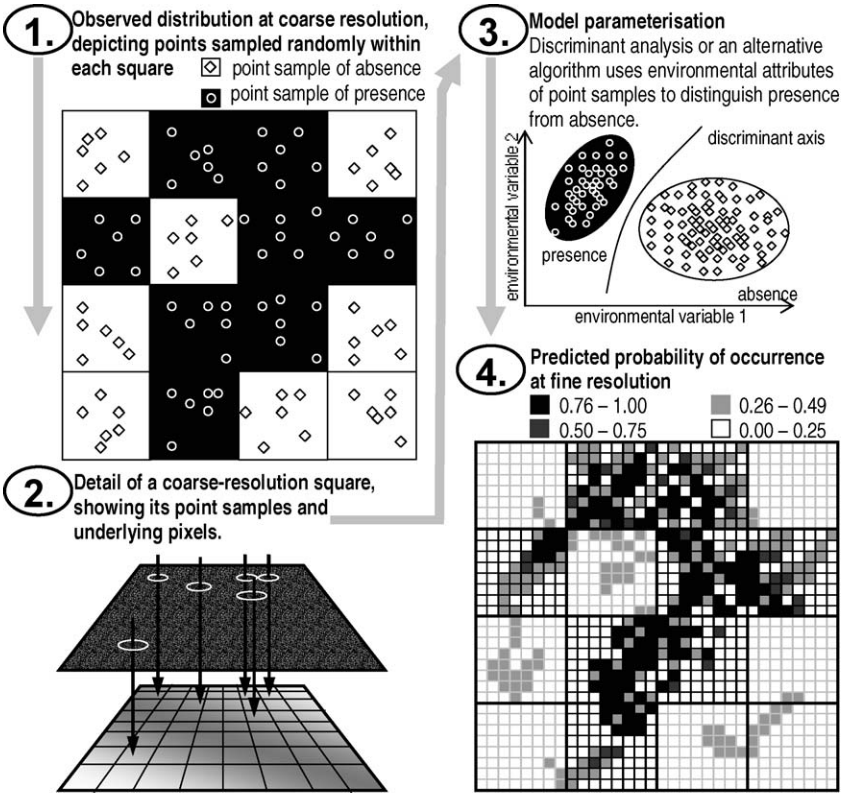


Fig. 3. The point sampling approach to high resolution predictions. *Step* 1 identifies the species' distribution at coarse resolution (black, presence; white, absence) and picks a fixed number of point localities randomly within each square. *Step* 2 matches each point locality to an underlying fine-resolution pixel and records its environmental attributes. *Step* 3 uses these environmental attributes to calibrate model parameters. *Step* 4 applies the calibrated parameters to fine-resolution environmental data to predict the species probability of occurrence at a fine spatial scale.

criminant analysis served as the model algorithm. In *pure* models, 50 pixels per square were sampled at random. In *habitat* models, sampled pixels had to contain a land cover category judged suitable for the species.

Given the randomised nature of sampling, 30 trials were run per species. For comparison with other approaches, both model accuracy and predicted probability of occurrence maps were averaged across the 30 trials per species.

3.4. Clustering approach

The clustering approach is different conceptually in that it uses information only from atlas squares known to harbour the species. It assumes (1) that each occupied square must contain some favourable habitat and (2) that favourable habitat is more homogenous from one occupied square to the next than unfavourable habitat. Accordingly, component pixels of atlas squares har-

Fig. 2. Iterative approach to high resolution predictions. *Iteration* 1 begins by calibrating model parameters based on coarse-resolution records, e.g. 1/4°, of both the species' observed distribution and environmental variables. If the model yields satisfactory predictions at coarse resolution, its parameters are applied to environmental data at double the resolution, i.e. 1/8°, for finer-resolution predictions of the species' probability of occurrence. Among these predictions, *iteration* 2 selects areas of predicted presence and absence for model re-calibration, taking into account sampling prevalence and correspondence with coarse resolution data. Model parameters are then calibrated anew, using predicted occurrence and environmental data at a resolution of 1/8°. Next, the new parameters are applied to environmental data of even higher resolution, e.g. 1/16°, for predictions of the species' probability of occurrence at ever finer resolution. *Iteration* 3 and any subsequent iterations follow the established pattern: selecting sites for re-calibration among predictions of the previous round; renewed parameterisation of the model using environmental data of equal resolution as predictions; application of parameters to finer environmental data.

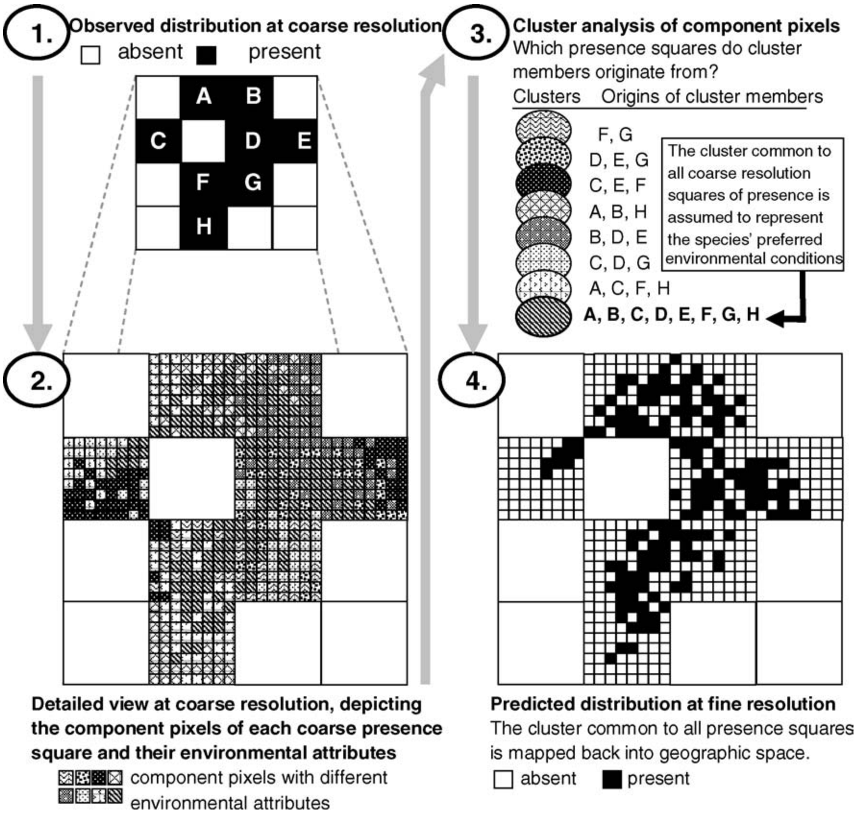


Fig. 4. The clustering approach to high resolution predictions. *Step* 1 identifies the areas of species presence at coarse resolution, squares A–H. *Step* 2 determines the environmental attributes of the finer-resolution pixels constituting the coarse presence squares. *Step* 3 groups the finer resolution pixels into clusters based on their environmental attributes, and then examines cluster membership. Ideally, one cluster contains a number of pixels from each of the original coarse presence squares. *Step* 4 assumes this cluster to represent habitat suitable to the species, and maps it back into geographic space as a prediction of species presence at fine resolution.

bouring the species are jointly subjected to cluster analysis. Resulting clusters whose membership includes pixels from all or most of the original squares, are then mapped back into geographic space as sites of species presence (Fig. 4). To predict species occurrence beyond the boundaries of squares used in analysis, Mahalanobis distances – based on the chosen cluster's multivariate mean and covariance structure – can be calculated to identify environmentally similar locations elsewhere (Green, 1978).

The approach's assumptions seem reasonable. Unless atlas records derive from sightings of purely transient individuals, atlas squares harbouring a species should contain suitable habitat. Habitat favoured by a species should also be less variable than the general environment, unless (1) the species is an extreme generalist, (2) the study area is large enough that habitat preferences exhibit local adaptations, or (3) predictors poorly capture the species' true requirements (e.g. open space regardless of land cover type).

Three methodological impediments, however, affect the approach. First, as in all cluster analyses, it is difficult to determine objectively how many clusters to divide data into. Second, misleading or inconclusive cluster arrangements could result if irrelevant environmental attributes mask differences in indices important to the species. Third, current computing capacity imposes severe limitations on practicability: due to the large sample sizes involved, cluster algorithms in commercially available software proved unable to cope with all but the three most narrowly-ranging species analysed here.

For these three species, the approach was implemented in S-Plus, using *clara*, a medoid-based clustering algorithm designed for large datasets (Insightful, 2001). Several cluster configurations were tested per species, and average silhouette width, an index of cluster separation (Insightful, 2001), was used to identify the best among them. For *pure* models, all component pixels of atlas squares harbouring the species were used in clustering. For *habitat* models, only pixels of suitable land cover were submitted.

4. Results

4.1. Gauging expectations: fine-grained models

Models calibrated at high resolution (1/64°) predicted test data successfully ($AUC \geq 0.7$) in only the three most narrow-ranging species (Table 3). For all remaining species, *fine* models performed no better than random ($AUC \approx 0.5$), with poor predictions for both presence and absence (e.g. compare Fig. 5a and b). Expectations for models calibrated at coarse resolution, therefore, were low for most species.

4.2. Direct approach

The predictive accuracy of direct approach models did not differ significantly from that of *fine* models (Wilcoxon's $Z = -0.77$, $p = 0.44$, $n = 9$; Fig. 6), even though the choice of predictor variables showed little overlap (i.e. the predictor variables chosen by *direct* models were generally different from those chosen in *fine* models; see Table A2 in Appendix A). Only one narrow-ranging species, however, was predicted highly accurately. Predictions were poor for all other species, and grew poorer as species' range size increased (Spearman rank correlation: $r_s = -0.80$, $p < 0.01$, $n = 9$; Table 3).

The approach frequently encountered environmental conditions distinct from those experienced during calibration, where no predictions could be made. It also tended towards ambiguous predictions (predicted probability of occurrence ≈ 0.5) and produced distribution maps of speckled appearance (Fig. 5c).

4.3. Iterative approach (pure models)

High-resolution predictions of iterative models were comparable in overall accuracy to *fine* models (Wilcoxon's $Z = -1.60$, $p = 0.11$, $n = 9$; Fig. 6), again despite little overlap in predictors (Table A2, Appendix A). The approach produced useful models for two narrow-ranging species (Table 3). The correlation between range size and accuracy was, however, not significant ($r_s = -0.3$, $p = 0.43$, $n = 9$) unless the poor model for Nahan's Francolin (the most narrow-ranging species) was excluded ($r_s = -0.86$, $p < 0.01$, $n = 8$).

Table 3
A comparison of the predictive accuracy of models calibrated at fine grain and those calibrated at coarse grain using the direct, point sampling and the iterative approach

Species (range size)	Resolution of calibration data					
	1/64°	1/4°				
	Fine	Direct	Point sampling		Iterative approach	
	Pure	Pure	Pure	Habitat	Pure	Habitat
Nahan's Francolin (3)	**1.00**	0.65	**0.94**	**0.98**	0.46	0.50
Fox's Weaver (6)	**0.75**	**1.00**	**0.98**	0.59	**0.94**	**0.92**
Crested Guineafowl (12)	**0.72**	0.69	**0.83**	**0.83**	**0.84**	**0.90**
Lesser Honeyguide (43)	0.42	0.68	**0.74**	**0.73**	0.59	0.64
Northern Puffback (54)	0.45	0.58	0.63	0.66	0.68	0.66
Yellow-rumped Tinkerbird (64)	0.57	0.54	0.66	0.67	0.67	0.68
Bronze Munia (98)	0.53	0.54	0.57	0.58	0.59	0.58
Green-backed Camaroptera (110)	0.50	0.59	0.55	0.54	0.55	0.59
Garden Bulbul (136)	0.47	0.49	0.52	0.52	0.49	0.49

Accuracy was measured as the area under the curve (AUC). Models were constructed either purely empirically (*pure*) or under inclusion of expert knowledge on habitat associations (*habitat*). Good models ($AUC \geq 0.7$, bolded) were achieved only rarely.

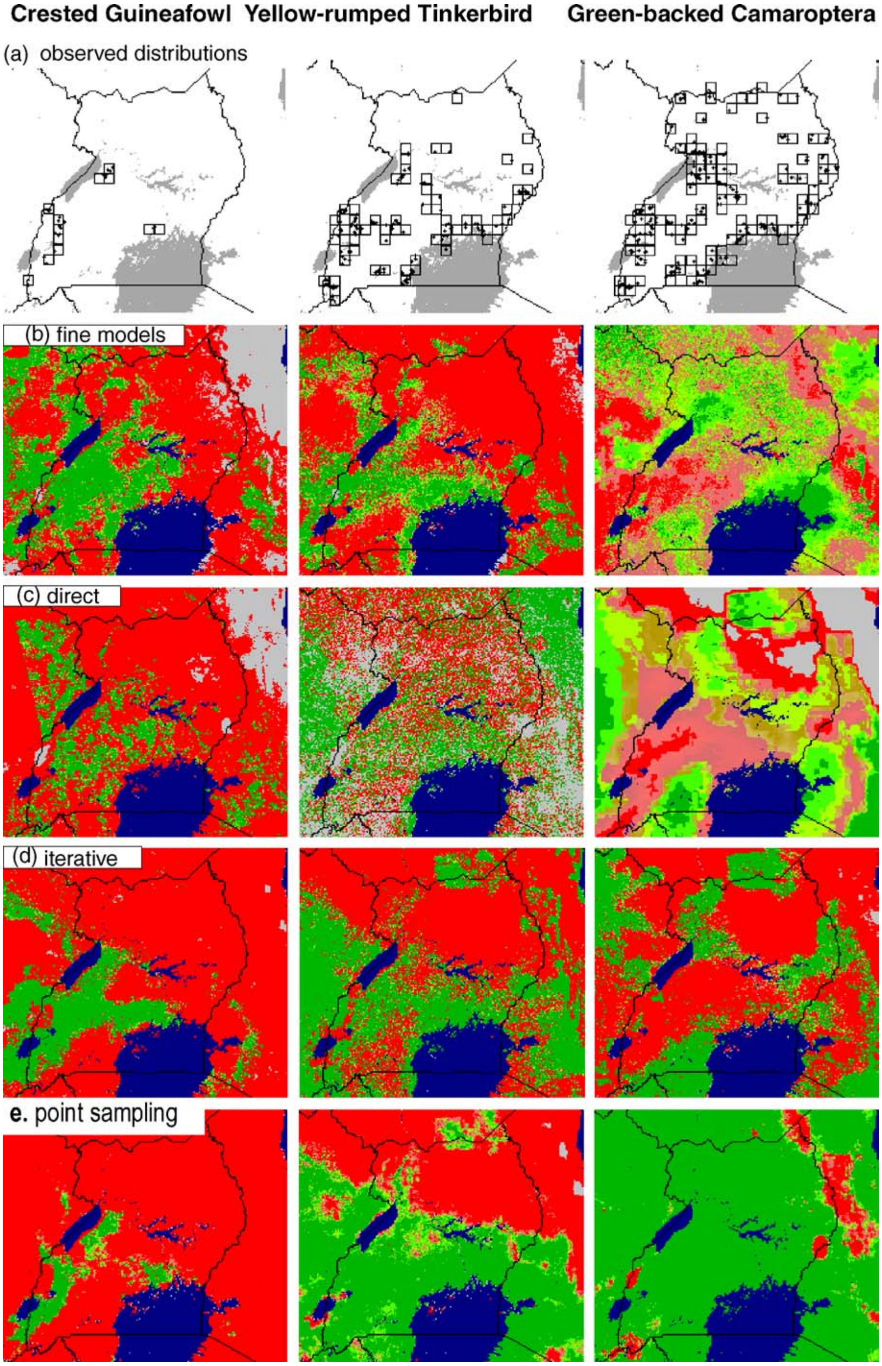


Fig. 5. Comparative performance of different modelling approaches, displaying the observed distribution of three species at 1/4° (empty squares) and 1/100° (black dots) in (a), alongside country boundaries (black lines) and water bodies (grey). High-resolution predictions are shown for *fine* models (b), the direct approach (c), iterative approach (d) and point sampling (e). Predicted probabilities of occurrence range from 0 (red) through 0.5 (yellow) to 1 (green). Areas where models made no predictions, because environmental conditions differed from those encountered during calibration, are shown in grey, water in blue.

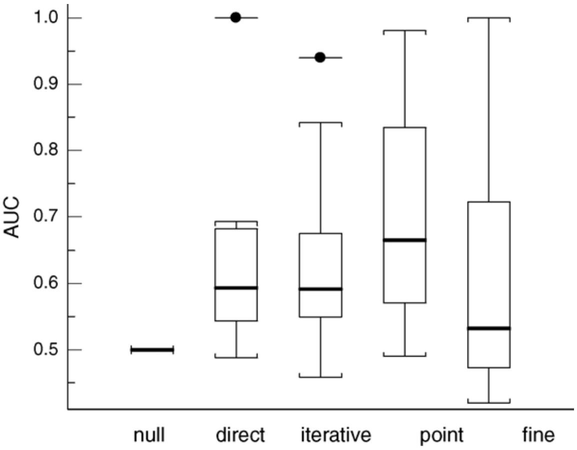


Fig. 6. Comparative accuracy of models trained at coarse resolution via the direct, iterative or point sampling approaches and models trained at fine resolution. Accuracy was measured as the area under ROC curves (AUC). As a point of reference, an AUC of 0.5, indicative of random null models, is shown on the left. Box-plots indicate the median (thick line), inter-quartile range (box) and outliers (filled circles) of AUC values achieved for nine species.

The approach was able to make predictions for most of the study region. Predicted distributions appeared more coherent than in the direct approach and suffered less ambiguity (Fig. 5d). Their general shape was determined early during the iterative process, with only small refinements as spatial resolution increased (e.g. Fig. 7). As a result, changes in accuracy were also limited and there was no significant relationship between the spatial resolution of predictions and the value of AUC calculated for equal-resolution test data (Friedman's test: $\chi^2 = 5.83$, $p = 0.12$, d.f. = 3, $n = 9$). Predictions of the iterative approach were hence no more accurate at 1/16° or 1/32° than at 1/64°.

Coarse-grain tests proved unreliable for model evaluation. Although AUC values achieved in coarse-grain tests were significantly correlated with AUC values obtained for fine resolution test data ($r_s = 0.50, p < 0.01$, $n = 36$), coarse-grain accuracy was only a weak predictor of accuracy at fine resolution ($r^2 = 0.29$, $p < 0.01$, $n = 36$).

4.4. *Point sampling approach (pure models)*

Point sampling models showed only limited overlap with *fine* models in their selection of predictors (Table A2, Appendix A), yet, astonishingly, achieved higher accuracy overall (Wilcoxon's $Z = -2.19$, $p = 0.03$, $n = 9$, Fig. 6). Useful predictions were achieved for four species (Table 3). Accuracy, however, declined sharply with increasing range size ($r_s = -0.97, p < 0.01, n = 9$). Predicted distributions followed occurrence records tightly in narrow-ranging species, but for commoner species the approach tended to predict presence throughout the region (Fig. 5e).

The approach was computationally slow, due to the need for several replications per species, and not significantly better than the direct or iterative approach (Friedman's $\chi^2 = 2.00$, $p = 0.37$, $n = 9$).

4.5. *Clustering approach (pure models)*

The clustering approach, implemented for only the three most narrow-ranging species, proved impracticable. In each species, the best cluster configuration contained two clusters, of which one included pixels from a majority of the original atlas squares. In all three species, however, this cluster proved ill-defined with regards to the study region. Based on Mahalanobis distances, most 1/100° pixels in Uganda were as similar to the cluster's multivariate mean as the original cluster members defining that mean. Essentially, therefore, the clustering approach predicted each of the narrow-ranging species to be omnipresent.

4.6. *Inclusion of expert knowledge*

The incorporation of expert knowledge on species–habitat associations had little impact on model results.

In the iterative approach, *pure* and *habitat* models were similar both visually and statistically (Wilcoxon's $Z = -1.01$, $p = 0.31$, $n = 9$; Table 3).

The same was true for the point sampling approach (Wilcoxon's $Z = -0.18$, $p = 0.86$, $n = 9$), although results for Fox's Weaver stood out. The approach's *habitat* model for this species performed poorly, unlike its *pure* counterpart (Table 3). None of the species' observed point records coincided with land cover categories deemed suitable by the published literature (see Table 1). Some suitable pixels occurred nearby and fell within the species' 1/4° atlas squares. Because suitable pixels were few in number, however, they restricted total sample size in the *habitat* model. The resulting predicted distribution was extremely restricted, omitting a number of known occurrences.

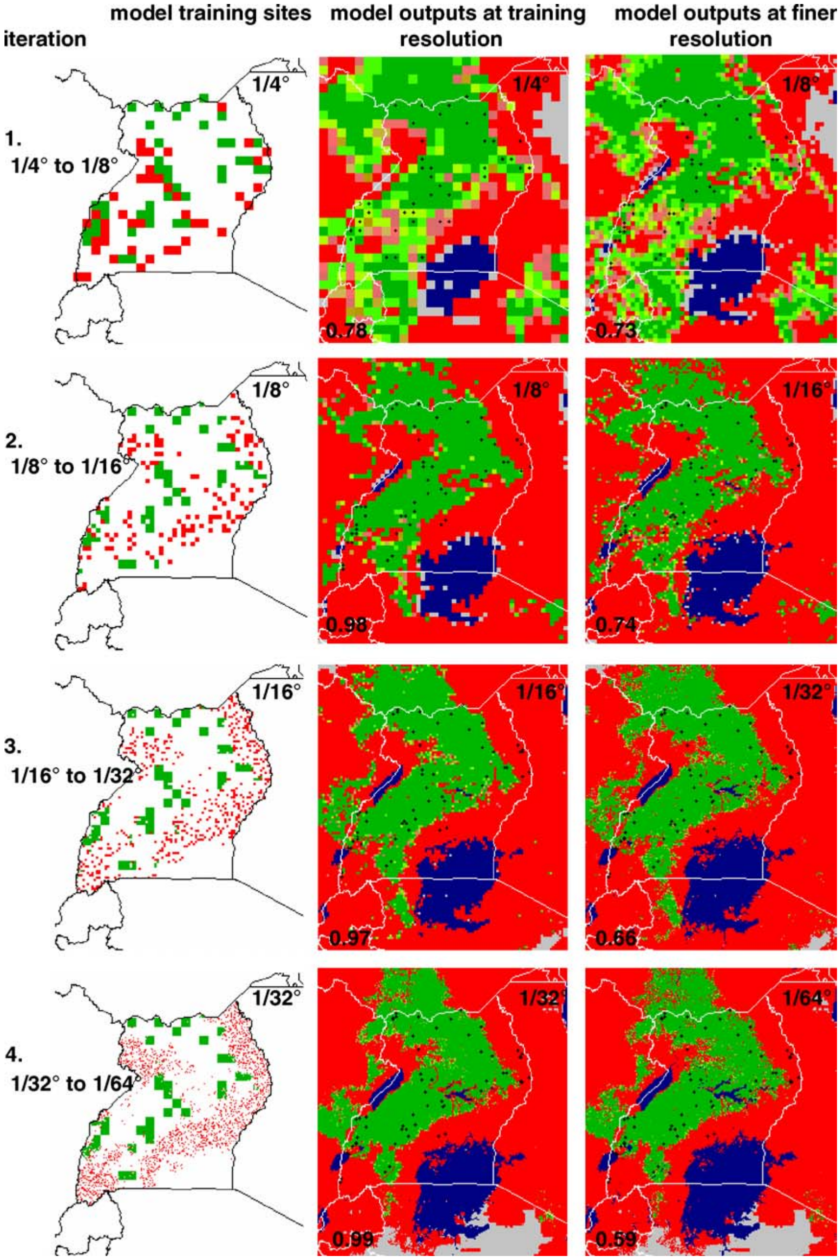


Fig. 7. Iterative predictions of the distribution of Lesser Honeyguide in Uganda. For each iteration (1–4), the left panel indicates sites used for model parameterisation, with green representing presence and red absence. The central panel shows model predictions at the spatial resolution of calibration sites. Predictions at double the resolution appear in the panel on the right. Species' predicted probability of occurrence ranges from 0 (red) through 0.5 (yellow) to 1 (green). No predictions were made (grey) where environmental conditions differed from those encountered at calibration sites. Black dots indicate the species' observed distribution. Water bodies are shown in blue, national borders in white. Model performance is indicated in the bottom left corner of each panel as training accuracy (central panels) or test accuracy (right panels).

In the clustering approach, clusters remained ill-defined even when only pixels of the species' preferred land cover category were taken into consideration.

5. Discussion

5.1. Previous work on cross-scale predictions

Models calibrated at coarse resolution were mostly unsuccessful at predicting the fine-grained distributions of Ugandan bird species. Prima facie, this suggests that the four approaches presented are ineffective tools for cross-scale predictions.

Previous studies provide a mixed opinion on this. Collingham et al. (2000), applying the direct approach to three non-indigenous plant species in Great Britain, found that models parameterised at a resolution of 10 km × 10 km yielded poor predictions at 2 km × 2 km. Using a comparable approach on 12 simulated species' distributions, ecologists at the *UFZ Centre for Environmental Research* in Leipzig, Germany, found that models trained at landscape-level predicted patch-level patterns no better than null models (Carsten Dormann, personal communication, 21 January 2004).

In contrast, Barbosa et al. (2003), who used the direct approach to predict the fine-grained distribution of otters in Spain and Portugal, judged their results satisfactory; assessment of accuracy was, however, qualitative only. A clearer story of success was reported by Lloyd and Palmer (1998): their model of a South African bulbul, calibrated via point sampling at 1/4°, replicated the species' distribution well at 1/15°. Most recently, Araújo et al. (2005), evaluating the direct approach for 81 bird species in Britain, found that, on average, models parameterised at a resolution of 50 km × 50 km predicted distributions at 10 km × 10 km satisfactorily, but noted that accuracy varied widely from species to species.

5.2. The influence of data quality

While cross-scale models for Ugandan bird species performed poorly overall, three of the four approaches were comparable in accuracy to models calibrated at fine resolution. *Fine* models performed no better than random for all but the three most narrow-ranging species. This raises questions over the quality of training data used.

Due to the small number of resident ornithologists, even coverage of the country could not be ensured during data collection for the *Bird Atlas of Uganda* (Tushabe et al., 2000, 2001). Many areas, therefore, received little or no sampling effort. Consequently, records of each species' occurrence may be geographically incomplete, and may not adequately reflect the full range of habitats used. Furthermore, incomplete coverage means that the absence of a record does not reliably indicate absence of the species. Although both training and test data attempted to guard against incorrect absences, the safeguard used (a threshold number of observations per grid square) was far from fail-safe. If observers preferentially recorded specific taxa, such as rare or conspicuous species, while ignoring others that were present but of less interest, improper absences may nonetheless have been included in the analyses.

Biased presence records and incorrect absence data could have led to poor model parameterisation. Incorrect absence data also call into question the results on model accuracy: if test data are riddled with mistakes, it becomes difficult to assess what represents genuine flaws in model predictions versus sampling inaccuracies. Because the likelihood of false absences diminishes as data resolution coarsens (Edwards et al., 1996), models trained at fine resolution may have been worse affected. This might explain the tendency of cross-scale models to outperform *fine* models.

Higher accuracy for narrow-ranging species in *fine* models, and in the direct, iterative and point sampling approach, could reflect better than average data quality due to preferential recording. Particularly Fox's Weaver, an endemic, and Nahan's Francolin, considered endangered (IUCN, 2004), may have attracted special attention. In addition, small test sample sizes (due to small range sizes) could have inflated AUC values for these species. Small test sample size is known to increase the standard error of AUC (McPherson et al., 2004).

Better accuracy for narrow-ranging species, however, could also have ecological reasons. The environmental niche of such species may be more concise, and hence more predictable, than that of wide-roaming species such as the Garden Bulbul. A narrow niche width is thought to facilitate more accurate models

(Mitchell et al., 2001; Hepinstall et al., 2002; Kadmon et al., 2003).

5.3. The choice of predictors

Although cross-scale approaches performed no worse than *fine* models overall, they each had drawbacks, and only one – the point sampling approach – yielded reasonable accuracy for all three species for which *fine* models indicated the potential.

The unsatisfactory performance of cross-scale models could stem, in part, from the use of predictor variables adequate only for the mapping of coarse but not fine patterns in a species' distribution. The satellite-derived predictors used in this study relate mostly to climatic conditions, although the NDVI serves as a surrogate for net primary productivity (Hay, 2000; Kerr and Ostrovsky, 2003) and the middle infrared signal can aid in the discrimination of vegetation types (Hay, 2000; Nagendra, 2001). Abiotic factors such as climate probably determine a species' distribution at coarse resolution. At finer scales, however, biotic attributes become important, including the availability of foraging resources and nest sites, and interactions with predators, competitors, and other individuals of the same species (Root, 1988; Cumming, 2002; Kadmon et al., 2003). Furthermore, the pattern of absence and presence within the broad boundaries of a species' range are also affected by dispersal ability and stochastic events that shape the species' metapopulation dynamics (Pulliam, 2000; Ferrier et al., 2002; Huston, 2002). These processes are unlikely to be captured by remotely-sensed environmental indices.

One might therefore expect models relying on satellite-derived predictors to perform more poorly at fine than coarse resolution. Yet remotely sensed indices such as NDVI have contributed to the accurate prediction of species' distributions at resolutions as fine as 1 km^2 (e.g. Osborne et al., 2001; Suarez-Seoane et al., 2002). Indeed, Seoane et al. (2004) tested whether vegetation-related variables predicted bird distributions better than climatic variables at fine spatial resolution, but found no significant difference. Furthermore, Lloyd and Palmer's (1998) successful cross-scale predictions for the Red-eyed Bulbul in South Africa relied on variables of similar nature as used here (climatic indices and NDVI).

The success of climatic indices in modelling distributions at high resolution does not negate the importance of biotic interactions and stochastic events at fine spatial grains. The dominance of processes shaping species' distributions may shift only gradually over a large range of scales (Hartley et al., 2004). Furthermore, these processes likely are hierarchically structured, such that factors influential at coarse scales shape the processes operating at fine scales (Whittaker et al., 2001). Hierarchical control and a gradual succession of controlling processes would both grant some – albeit diminished – predictive power to abiotic variables at relatively fine spatial grains.

In our analyses, predictors chosen by *fine* models were dissimilar from those chosen by models trained at coarse grain. Yet there was no obvious trend away from purely climatic variables, such as temperature, to more biotic indices, such as NDVI. The limited overlap in predictors may in part be the consequence of inadequate training data (poor model parameterisation), or may reflect collinearity among some predictors.

Overall, it is doubtful that inappropriate predictors were responsible for the poor performance of cross-scale models in Uganda. Inadequate training data are more likely to be at fault.

5.4. The role of expert knowledge

Models were not improved by inclusion of expert knowledge on species habitat associations. Expert knowledge can be incorporated into empirical models in many different ways (Pearce et al., 2001), for example, to screen occurrence records for anomalous observations (Ferrier et al., 2002; Lehmann et al., 2002), formulate habitat suitability indices (Bustamante, 1997; Ferrier et al., 2002), select biologically relevant predictors (Pausas et al., 1995; Puttock et al., 1996; Hill et al., 1999; Mitchell et al., 2001; Johnson et al., 2004), or evaluate and refine predicted distribution maps (Anderson et al., 2003; Engler et al., 2004).

Expert knowledge may have shown little effect in our models because its ability to shape models was weak: it served to select sites for model calibration but did not constrain the choice of predictors. A further impediment to effective influence may have been the sometimes tenuous match between species' published habitat preferences and available land cover information. Moreover, species' perceptions of their environment may deviate considerably from human classifications of habitat (Knight and Morris, 1996; Cablk et al., 2002). Reliance on land cover categories will certainly be problematic where species' preferred habitat types are too fine-grained to be captured in land cover maps and/or difficult to discriminate by remote sensing (Edwards et al., 1996). The remotely sensed land cover map used in this study distinguished only 24 land cover types globally. Alternative land cover categorisations, such as the International Geosphere Biosphere Programme classification scheme or Olson's Global Ecosystem Framework (both available at http://edcdaac.usgs.gov/glcc/globdoc2_0.html), did not, however, yield a better match with species' habitat preferences.

Comparative studies appear rare, but purely knowledge-based and purely empirical models have been found to yield similar results (Bolliger et al., 2000; Petit et al., 2003). Consequently, an integration of the two approaches may generally fail to improve results. In a study encompassing distribution models for 93 Australian vertebrates, incorporation of expert knowledge into otherwise empirical models did not significantly improve performance (Pearce et al., 2001). Results presented here suggest that this holds true even when models attempt to bridge two spatial scales, a situation in which knowledge-based rules may appear to have some advantage.

6. Conclusion

The quick pace of environmental change, our need to understand its ecological implications, and the dissimilar scales at which researchers and decision-makers operate beg for the development of effective methods to translate observations at one spatial scale into reliable predictions at another. In search of a solution, Root and Schneider (1995) proposed a research framework that involves continuous cycling between large and small-scale studies, with results at one scale informing investigations at the other. While in the long term this may be the only route to a thorough understanding of the cause-and-effect relationships shaping our environment, quicker and cheaper solutions may be needed in the meantime.

Statistical approaches to cross-scale predictions, such as those examined here, may have the potential to provide quick answers. In essence, however, they attempt to fashion data where none exist. Their often persuasive map outputs risk creating a false sense of certitude. Conclusive evaluation of the strengths and weaknesses of these approaches will require data of suitable quality. Appropriate empirical data with fine spatial resolution may be difficult to come by, but researchers could resort to simulated data.

Based on the results presented here, and the experience of other authors (Lloyd and Palmer, 1998; Collingham et al., 2000), the point sampling and iterative approaches look most promising. Although data inadequacies bar any firm conclusions, the point sampling approach is likely to work best for species with small ranges whose niche requirements are narrow and geographically clustered. The iterative approach, in contrast, may provide more nuanced predictions for species whose niche is less restricted.

Refinements may be possible in both approaches, for example, by taking into account the species' expected prevalence at fine resolution. Considerable progress has recently been made on cross-scale predictions of species' prevalence (Hartley et al., 2004): Kunin and colleagues (Kunin, 1998; Kunin et al., 2000) and He and Gaston (2000) developed methods to estimate a species' total area of occupancy at fine resolution from coarse-grained distribution maps. Such estimates could be used to manipulate a species' prior probability of occurrence in models involving discriminant analysis or alternative statistical algorithms.

If, however, methods that rely entirely on coarse-grained occurrence records prove unable to accurately predict species' distributions at fine spatial resolution, a compromise may involve coarse-grained models that incorporate fine-scale information obtainable with limited survey effort. Johnson et al. (2004), for example, integrated a model describing patch-level vegetation-use by individuals in one caribou herd with a topography-based model describing the overall distribution of caribou herds in the Canadian Rockies. The integration simply involved multiplying the predicted probability of occurrence of one model with that of the other (Johnson et al., 2004). Although their study did not strictly entail a change in resolution, the approach may be applicable to cross-scale predictions. Coarse-grained occurrence records spanning a large spatial extent could be combined with fine-grained data collected over a small area to yield models that predict, over the large extent, the species' likelihood of occurrence at high spatial resolution, conditional upon broad-scale environmental suitability. If found, reliable methods of cross-scale prediction could be of immense practical use.

Acknowledgements

J.M. McPherson's research was kindly supported by a Scatcherd European Scholarship from the University of Oxford. We thank the authors of the *Bird Atlas of Uganda* for sharing their data. Thanks also to James H. Brown and his lab for hosting Jana during part of this study, and to Carsten Dormann, Duncan McPherson and an anonymous referee for helpful comments on earlier drafts of this paper.

Appendix A

Table A1
The 61 satellite-derived predictor variables available for model parameterisation

Satellite variable	Description	Measure	Name
Cold cloud duration (CCD)	An index of rainfall, based on the number of hours a locality was covered by clouds assumed to bear rain. Rain-bearing clouds are associated with a particular threshold temperature at their tops, which is recorded by channel 2 of the High Resolution Radiometer on board the Meteosat satellite series of the European meteorological satellite programme. On a monthly basis, the index is thought to be accurate to ±38 mm (Hay and Lennon, 1999).	Mean Maximum Minimum Variance Amplitude of annual cycle Phase of annual cycle Amplitude of biannual cycle Phase of biannual cycle Amplitude of triannual cycle Phase of triannual cycle	CCDmean CCDmax CCDmin CCDvar CCDamp1 CCDphase1 CCDamp2 CCDphase2 CCDamp3 CCDphase3
Digital elevation model (DEM)	The United State's Geological Survey provides a digital elevation model with a resolution of 30 arc seconds, here used to determine mean altitude per grid square.	Mean altitude above sea level	DEM
Land surface temperature (LST)	Land surface temperature indices are calculated using a split-window algorithm to minimise the confounding effects of atmospheric attenuation and variable surface emissivity (Goetz et al., 2000). Price's index, used here, is based on channels 4 and 5 of the Advanced Very High Resolution Radiometers (AVHRR) on board the US National Oceanic and Atmospheric Administration's (NOAA) satellites. Both measure emitted thermal infrared. The index is believed to be accurate to ±4 °C in Africa (Hay and Lennon, 1999).	Mean Maximum Minimum Variance Amplitude of annual cycle Phase of annual cycle Amplitude of biannual cycle Phase of biannual cycle Amplitude of triannual cycle Phase of triannual cycle	LSTmean LSTmax LSTmin LSTvar LSTamp1 LSTphase1 LSTamp2 LSTphase2 LSTamp3 LSTphase3
Middle infrared radiation (MIR)	Radiation in the middle infrared is measured by AVHRR channel 3 and relates to temperature, but can also aid in land cover discrimination. While the interaction of middle infrared radiation with targets is not yet well understood, Channels 3 has the advantage of suffering less from atmospheric attenuation than other thermal channels (Hay, 2000)	Mean Maximum Minimum Variance Amplitude of annual cycle Phase of annual cycle Amplitude of biannual cycle Phase of biannual cycle Amplitude of triannual cycle Phase of triannual cycle	MIRmean MIRmax MIRmin MIRvar MIRamp1 MIRphase1 MIRamp2 MIRphase2 MIRamp3 MIRphase3

Table A1 (*Continued*)

Satellite variable	Description	Measure	Name
Normalised difference vegetation index (NDVI)	An index of vegetation biomass derived from AVHRR channels 1 and 2, which measure reflected solar radiation in the visible red and near infrared, respectively. Like other spectral vegetation indices, the NDVI exploits the fact that the photosynthetic pigments of plants absorb visible light in the red wavelengths, while mesophyll tissue (non-photosynthesising plant tissue) reflects near-infrared wavelengths. This leads to a divergence in reflectance between the visible and near-infrared, which distinguishes vegetation from bare soil (Hay, 2000).	Mean	NDVImean
		Maximum	NDVImax
		Minimum	NDVImin
		Variance	NDVIvar
		Amplitude of annual cycle	NDVIamp1
		Phase of annual cycle	NDVIphase1
		Amplitude of biannual cycle	NDVIamp2
		Phase of biannual cycle	NDVIphase2
		Amplitude of triannual cycle	NDVIamp3
		Phase of triannual cycle	NDVIphase3
Air temperature (TAIR)	This index of air temperature assumes that the radiometric (surface) temperature of a fully vegetated canopy is in equilibrium with ambient air temperature, because dense vegetation has a heat capacity similar to its surrounding air. The index therefore regresses LST against NDVI to determine what value LST takes when NDVI is suggestive of full vegetation cover (generally 0.65). At this point, LST should measure canopy surface rather than soil surface temperature. The resulting index is accurate to ±2.98–3.93 °C (Goetz et al., 2000).	Mean	TAIRmean
		Maximum	TAIRmax
		Minimum	TAIRmin
		Variance	TAIRvar
		Amplitude of annual cycle	TAIRamp1
		Phase of annual cycle	TAIRphase1
		Amplitude of biannual cycle	TAIRamp2
		Phase of biannual cycle	TAIRphase2
		Amplitude of triannual cycle	TAIRamp3
		Phase of triannual cycle	TAIRphase3
Vapour pressure deficit (VPD)	An estimate of near-surface atmospheric humidity indicative of the 'drying power' of air (Goetz et al., 2000). It's computation requires an estimate of the near-surface water vapour content of air (U), which is obtained via a split-window algorithm using AVHRR channels 4 and 5. U then serves to determine the dew point temperature at which saturation and condensation would occur. To obtain the VPD, actual vapour pressure (determined with the help of TAIR, see above) is subtracted from saturation vapour pressure. The index is accurate to between ±6.0 mbar and ±10.9 mbar, depending on the satellite imagery's spatial resolution (Goetz et al., 2000; Hay and Lennon, 1999).	Mean	VPDmean
		Maximum	VPDmax
		Minimum	VPDmin
		Variance	VPDvar
		Amplitude of annual cycle	VPDamp1
		Phase of annual cycle	VPDphase1
		Amplitude of biannual cycle	VPDamp2
		Phase of biannual cycle	VPDphase2
		Amplitude of triannual cycle	VPDamp3
		Phase of triannual cycle	VPDphase3

Imagery from the US National Oceanic and Atmospheric Administration's satellites were obtained as 10-day maximum value composites from the US Geological Survey's EROS Data Center (http://edcdaac.usgs.gov/1KM/1kmhomepage.asp). Imagery from the European meteorological satellite programme were obtained courtesy of Fred Snijders through the ARTEMIS program of the United Nations Food and Agricultural Organisation. Monthly compositing was undertaken using ERDAS IMAGINE software (v.8.5). Fourier processing of the resulting monthly data series was accomplished through custom-written programs in QuickBasic. Custom-written programs were also used for re-sampling, to aggregate imagery to coarser resolution by averaging only valid terrestrial pixels.

Table A2

Model details, indicating (a) the number of occurrence records available per species at each spatial resolution and the predictor variables selected in (b) fine, (c) direct, (d) iterative and (e) point sampling models

Species	Nahan's Francolin	Fox's Weaver	Crested Guineafowl	Lesser Honeyguide	Northern Puffback	Yellow-rumped Tinkerbird	Bronze Munia	Green-backed Camaroptera	Garden Bulbul
(a) Number of occurrences at each spatial resolution									
1/4	3	6	11	41	51	61	94	105	130
1/8	3	6	20	50	62	99	133	163	215
1/16	4	6	22	54	67	127	157	204	282
1/32	7	6	25	57	73	142	167	236	340
1/64	9	6	26	60	76	154	175	246	362
1/100	11	6	27	60	76	156	175	256	390
(b) Fine model predictor variables (trained at 1/64° to predict at 1/64°)									
	1. VPDmean	1. VPDmean	1. NDVImean	1. VPDmin	1. CCDphase1	1. TAIRmean	1. MIRmean	1. CCDamp3	1. MIRamp1
			2. CCDvar	2. MIRamp2	2. MIRamp3	2. NDVImean	2. MIRphase1	2. MIRmax	2. MIRphase1
			3. NDVIphase2	3. MIRmax	3. TARIamp1	3. VPDphase1	3. TAIRphase1	3. CCDamp2	3. VPDphase2
				4. LSTamp2	4. CCDamp2	4. CCDvar	4. CCDamp3		4. VPDamp3
				5. MIRphase2	5. NDVImax	5. DEM	5. CCDamp1		5. MIRphase3
				6. LSTphase2	6. MIRamp2	6. LSTmax	6. LSTphase2		6. MIRphase2
				7. MIRphase1	7. LSTphase2	7. LSTphase3	7. TAIRphase2		7. LSTphase3
				8. NDVIphase1	8. LSTphase3	8. NDVIvar	8. NDVImean		8. MIRmean
				9. LSTmax	9. VPDmean	9. CCDmean	9. MIRmin		9. NDVImax
					10. MIRphase3	10. MIRamp3	10. MIRvar		10. TAIRamp1
EPV	6.0	4.0	5.7	4.4	5.0	10.1	11.0	33.7	8.7
(c) Direct approach predictor variables (trained at 1/4° to predict at 1/64°)									
	1. MIRphase1	1. VPDmin	1. MIRmean	1. CCDamp3	1. CCDphase1	1. NDVImin	1. MIRphase1	1. CCDphase2	1. LSTvar
			2. TAIRvar	2. NDVImean	2. LSTamp2	2. LSTamp2	2. MIRphase2	2. VPDva-	
				3. NDVImin	3. NDVIamp3	3. DEM		3. MIRmax	
						4. MIRamp2		4. LSTphase3	
						5. MIRamp3		5. TAIRphase2	
						6. MIRphase1			
EPV	3.0	6.0	5.5	13.7	15.0	6.3	4.2	5.0	5.0

Table A2 (*Continued*)

Species	Nahan’s Francolin	Fox’s Weaver	Crested Guineafowl	Lesser Honeyguide	Northern Puffback	Yellow-rumped Tinkerbird	Bronze Munia	Green-backed Camaroptera	Garden Bulbul
(d) Iterative approach predictor variables (final iteration: trained at 1/32° based on predictions form previous iterations to predict at 1/64°)									
	1. MIRphasei	1. VPDmin	1. VPDmean	1. CCDmax	1. CCDphase1	1. NDVImin	2. CCDamp2	1. VPDphase2	1. LSTvar
			2. CCDamp2	2. CCDamp3	2. NDVImin	2. MIRamp2	2. LSTphase3	2. CCDphase2	
			3. VPDphase3	3. CCDamp1	3. DEM	3. TAIRmean	3. MIRphase1	3. CCDamp3	
			4. CCDphase1	4. NDVIamp1		4. VPDmax	4. MIRphase2	4. NDVIphase1	
			5. TAIRphase2	5. NDVImean		5. MIRamp3	5. VPDmean	5. VPDvar	
			6. MIRvar	6. CCDphase3		6. MIRmean	6. TAIRphase2	6. TAIRphase1	
			7. CCDamp3	7. MIRamp2		7. LSTamp2	7. MIRmax	7. CCDamp1	
			8. LSTamp2	8. MIRamp1		8. NDVIamp3	8. TAIRmax	8. MIRamp3	
			9. LSTphase3	9. CCDphase2			9. TAIRamp3	9. NDVIvar	
				10. VPDphase1			10. VPDvar	10. MIRphase1	
EPV	153.0	375.0	76.4	187.1	676.7	281.0	280.6	237.6	1445.0
(e) Point sampling approach predictor variables (predictors most frequently picked across 30 models per species trained by randomly sampling fifty 1/100° points within each 1/4° square to predict at 1/100°)									
	1. NDVImin	1. VPDphae3	1. NDVImin	1. CCDamp3	1. NDVIphase1	1. CCDvar	1. VPDamp2	1. DEM	1. MIRmean
	2. VPDmin	2. VPDmin	2. TAIRphase2	2. VPDvar	2. CCDphase2	2. DEM	2. TAIRvar	2. CCDmax	2. CCDmean
	3. TAIRphase2	3. TAIRvar	3. CCDmax	3. VPDphase2	3. VPDamp2	3. NDVIamp1	3. CCDamp2	3. CCDamp2	3. CCDphase1
	4. CCDamp2	4. CCDphase1	4. TAIRphase1	4. TAIRmean	4. VPDvar	4. VPDmin	4. NDVIamp1	4. TARIvar	4. CCDamp2
	5. CCDamp3	5. CCDamp1	5. CCDphase2	5. CCDvar	5. TAIRphase1	5. NDVImin	5. TAIRphase3	5. VPDphase2	5. CCDphase3
	6. CCDvar	6. VPDmean	6. CCDamp3	6. VPDamp2	6. CCDmax	6. VPDphase2	6. DEM	6. CCDphase3	6. CCDmax
	7. DEM	7. VPDamp3	7. CCDamp1	7. VPDmin	7. NDVImin	7. VPDvar	7. LSTmax	7. MIRmean	7. TAIRphase1
	8. CCDphase1	8. VPDvar	8. TAIRamp2	8. CCDphase3	8. TAIRmax	8. MIRamp1	8. VPDamp2	8. NDVImax	8. VPDphase2
	9. CCDphase3	9. TAIRamp3	9. VPDphase1	9. NDVImax	9. CCDvar	9. VPDphase3	9. NDVImin	9. VPDphase3	9. NDVIvar
	10. MIRphase1	10. CCDamp2	10. VPDphase2	10. DEM	10. DEM	10. NDVIphase2	10. VPDvar	10. VPDmin	10. MIRamp1
EPV	15.0	30.0	60.0	215.0	228.6	188.7	100.0	40.0	20.0

Variable names follow Table A1. Events per variable (EPV) indicate the number of presences (or absences if these were less numerous) used to train models, divided by the number of predictors selected.

References

Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol. Modell. 162, 211–232.

Araújo, M.B., Thuiller, W., Williams, P.H., Reginster, I., 2005. Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. Glob. Ecol. Biogeogr. 14, 17–30.

Balmford, A., Moore, J.L., Brooks, T., Burgess, N., Hansen, L.A., Williams, P., Rahbek, C., 2001. Conservation conflicts across Africa. Science 291, 2616–2619.

Barbosa, A.M., Real, R., Olivero, J., Vargas, J.M., 2003. Otter (*Lutra lutra*) distribution modeling at two resolution scales suited to conservation planning in the Iberian Peninsula. Biol. Conserv. 114, 377–387.

Böhning-Gaese, K., 1997. Determinants of avian species richness at different spatial scales. J. Biogeogr. 24, 49–60.

Bolliger, J., Kienast, F., Bugmann, H., 2000. Comparing models for tree distributions: concept, structures, and behavior. Ecol. Modell. 134, 89–102.

Brown, L.H., Urban, E.K., Newman, K. (Eds.), 1982. The Birds of Africa, vol. 1. Academic Press, London.

Bustamante, J., 1997. Predictive models for lesser kestrel *Falco naumanni* distribution, abundance and extinction in southern Spain. Biol. Conserv. 80, 153–160.

Bustamante, J., Seoane, J., 2004. Predicting the distribution of four species of raptors (Aves: Accipitridae) in southern Spain: statistical models work better than existing maps. J. Biogeogr. 31, 295–306.

Cablk, M., White, D., Kiester, A.R., 2002. Assessment of spatial autocorrelation in empirical models in ecology. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences—Issues of Accuracy and Scale. Island Press, Washington, pp. 429–440.

Carswell, M., Pomeroy, D.E., Reynolds, J., Tushabe, H., 2005. Bird atlas of Uganda. British Ornithologists' Club, London, 553 pp.

Chatfield, C., 1996. The Analysis of Time Series: An Introduction, fifth ed. London, Chapman & Hall.

Cohen, J.A., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46.

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. J. Appl. Ecol. 37, 13–27.

Cumming, G.S., 2002. Comparing climate and vegetation as limiting factors for species ranges of African ticks. Ecology 83, 255–268.

DeGraaf, R.M., Hestbeck, J.B., Yamasaki, M., 1998. Associations between breeding bird abundance and stand structure in the White Mountains, New Hampshire and Maine, USA. For. Ecol. Manage. 103, 217–233.

del Hoyo, J., Elliott, A., Sargatal, J. (Eds.), 1992. Handbook of the Birds of the World, vols. 1–9, Lynx Editions. Barcelona.

Edwards, T.C., Deshler, E.T., Foster, D., Moisen, G.G., 1996. Adequacy of wildlife habitat relation models for estimating spatial distributions of terrestrial vertebrates. Conserv. Biol. 10, 263–270.

Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J. Appl. Ecol. 41, 263–274.

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. Biodivers. Conserv. 11, 2275–2307.

Goetz, S.J., Prince, S.D., Small, J., 2000. Advances in satellite remote sensing of environmental variables for epidemiological applications. In: Hay, S.I., Randolph, S.E., Rogers, D.J. (Eds.), Remote Sensing and Geographical Information Systems in Epidemiology, vol. 47. Academic Press, London, UK, pp. 289–307.

Green, P.E., 1978. Analyzing Multivariate Data. Dryden Press, Hinsdale, Ill.

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.

Harrison, J.A., Allan, D.G., Underhill, L.G., Herremans, M., Tree, A.J., Parker, V., Brown, C.J. (Eds.), 1997. The Atlas of Southern African Birds. Birdlife South Africa, Johannesburg.

Harte, J., Kinzig, A., Green, J., 1999. Self-similarity in the distribution and abundance of species. Science 284, 334–336.

Hartley, S., Kunin, W.E., Lennon, J.J., Pocock, M.J.O., 2004. Coherence and discontinuity in the scaling of species' distribution patterns. Proc. R. Soc. Lond. Ser. B Biol. Sci. 271, 81–88.

Hay, S.I., 2000. An overview of remote sensing and geodesy for epidemiology and public health applications. In: Hay, S.I., Randolph, S.E., Rogers, D.J. (Eds.), Remote Sensing and Geographical Information Systems in Epidemiology, vol. 47. Academic Press, London, U.K., pp. 1–35.

Hay, S.I., Lennon, J.J., 1999. Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. Trop. Med. Int. Health 4, 58–71.

He, F.L., Gaston, K.J., 2000. Estimating species abundance from occurrence. Am. Nat. 156, 553–559.

Hepinstall, J.A., Krohn, W.B., Sader, S.A., 2002. Effects of niche width on the performance and agreement of avian habitat models. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences—Issues of Accuracy and Scale. Island Press, Washington, pp. 593–606.

Hill, J.K., Thomas, C.D., Huntley, B., 1999. Climate and habitat availability determine 20th century changes in a butterfly's range margin. Proc. R. Soc. Lond. Ser. B Biol. Sci. 266, 1197–1206.

Hobbs, N.T., 2003. Challenges and opportunities in integrating ecological knowledge across scales. For. Ecol. Manage. 181, 223–238.

Huston, M.A., 2002. Critical issues for improving predictions. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences—Issues of Accuracy and Scale. Island Press, Washington, pp. 7–21.

Insightful^{TM}, 2001. S-Plus 6 for Windows Guide to Statistics, vol. 2. Insightful Corporation, Seattle, WA.

IUCN, 2004. 2004 IUCN Red List of Threatened Species. IUCN, Gland, Switzerland. <http://www.iucnredlist.org>. Last accessed: 02/03/2005.

Jetz, W., Rahbek, C., 2002. Geographic range size and determinants of avian species richness. Science 297, 1548–1551.

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. J. Appl. Ecol. 41, 238–251.

Kadmon, R., Farber, O., Danin, A., 2003. A systematic analysis of factors affecting the performance of climatic envelope models. Ecol. Appl. 13, 853–867.

Kerr, J.T., Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. Trends Ecol. Evol. 18, 299–305.

Knight, T.W., Morris, D.W., 1996. How many habitats do landscapes contain? Ecology 77, 1756–1764.

Kunin, W.E., 1998. Extrapolating species abundance across spatial scales. Science 281, 1513–1515.

Kunin, W.E., Hartley, S., Lennon, J.J., 2000. Scaling down: on the challenge of estimating abundance from occurrence patterns. Am. Nat. 156, 560–566.

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

Lennon, J.J., Greenwood, J.J.D., Turner, J.R.G., 2000. Bird diversity and environmental gradients in Britain: a test of the species-energy hypothesis. J. Anim. Ecol. 69, 581–598.

Levin, S.A., 1992. The problem of pattern and scale in ecology. Ecology 73, 1943–1967.

Lewis, A., Pomeroy, D.E., 1989. A Bird Atlas of Kenya. Balkema, Rotterdam, Netherlands.

Lindenmayer, D.B., 2000. Factors at multiple scales affecting distribution patterns and their implications for animal conservation—Leadbeater's Possum as a case study. Biodivers. Conserv. 9, 15–35.

Lloyd, P., Palmer, A.R., 1998. Abiotic factors as predictors of distribution in southern African Bulbuls. Auk 115, 404–411.

Lovett, J.C., Rudd, S., Taplin, J., Frimodt-Moller, C., 2000. Patterns of plant diversity in Africa south of the Sahara and their implications for conservation management. Biodivers. Conserv. 9, 37–46.

MacFaden, S.W., Capen, D.E., 2002. Avian habitat relationships at multiple scales in a New England forest. For. Sci. 48, 243–253.

McClave, J.T., Dietrich, F.H., 1994. Statistics, sixth ed. Maxwell Macmillan International, New York.

McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? J. Appl. Ecol. 41, 811–823.

Mitchell, M.S., Lancia, R.A., Gerwin, J.A., 2001. Using landscape-level data to predict the distribution of birds on a managed forest: effects of scale. Ecol. Appl. 11, 1692–1708.

Nagendra, H., 2001. Using remote sensing to assess biodiversity. Int. J. Remote Sens. 22, 2377–2400.

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. Ecol. Appl. 10, 1356–1366.

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. J. Appl. Ecol. 38, 458–471.

Parker, V., 1999. The Atlas of Birds of Sul do Save, Southern Mozambique. Avian Demography Unit and the Endangered Wildlife Trust, Cape Town and Johannesburg, South Africa.

Pausas, J.G., Braithwaite, L.W., Austin, M.P., 1995. Modeling habitat quality for arboreal marsupials in the south coastal forests of New-South-Wales, Australia. For. Ecol. Manage. 78, 39–49.

Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Modell. 133, 225–245.

Pearce, J.L., Cherry, K., Drielsma, M., Ferrier, S., Whish, G., 2001. Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. J. Appl. Ecol. 38, 412–424.

Petit, S., Haysom, K., Pywell, R., Warman, L., Allen, D., Booth, R., Firbank, L., 2003. Habitat-based models for predicting the occurrence of ground-beetles in arable landscapes: two alternative approaches. Agric. Ecosyst. Environ. 95, 19–28.

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecol. Lett. 3, 349–361.

Puttock, G.D., Shakotko, P., Rasaputra, J.G., 1996. An empirical habitat model for moose, *Alces alces*, in Algonquin Park, Ontario. For. Ecol. Manage. 81, 169–178.

Riitters, K.H., O'Neill, R.V., Jones, K.B., 1997. Assessing habitat suitability at multiple scales: a landscape-level approach. Biol. Conserv. 81, 191–202.

Rogers, D.J., Hay, S.I., Packer, M.J., 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. Ann. Trop. Med. Parasitol. 90, 225–241.

Root, T., 1988. Energy constraints on avian distributions and abundances. Ecology 69, 330–339.

Root, T.L., Schneider, S.H., 1995. Ecology and climate—research strategies and implications. Science 269, 334–341.

Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B., 2002. Introduction. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences—Issues of Accuracy and Scale. Island Press, Washington, pp. 1–5.

Seoane, J., Bustamante, J., Diaz-Delgado, R., 2004. Competing roles for landscape, vegetation, topography and climate in predictive models of bird distribution. Ecol. Modell. 171, 209–222.

Sergio, F., Pedrini, P., Marchesi, L., 2003. Adaptive selection of foraging and nesting habitat by black kites (*Milvus migrans*) and its implications for conservation: a multi-scale approach. Biol. Conserv. 112, 351–362.

Sibley, C.G., Monroe, B.L., 1990. Distribution and Taxonomy of the Birds of the World. Yale University Press, New Haven, USA.

Sibley, C.G., Monroe, B.L., 1993. A Supplement to the Distribution and Taxonomy of Birds of the World. Yale University Press, New Haven, USA.

Stevenson, T., Fanshawe, J., 2001. Field Guide to the Birds of East Africa : Kenya, Tanzania, Uganda, Rwanda, Burundi. T&AD Poyser, London.

Suarez-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. J. Appl. Ecol. 39, 755–771.

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

Thompson, C.M., McGarigal, K., 2002. The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). Landsc. Ecol. 17, 569–586.

Tushabe, H., Pomeroy, D., Reynolds, J., 2000. Predicting bird species occurrences for conservation planning in Uganda. Ostrich 71, 307–309.

Tushabe, H., Reynolds, J., Pomeroy, D., 2001. Innovative aspects of the Bird Atlas of Uganda. Ostrich, 183–188.

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. Conserv. Biol. 17, 1601–1611.

Whittaker, R.J., Willis, K.J., Field, R., 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. J. Biogeogr. 28, 453–470.