Diversity and Distributions, (Diversity Distrib.) (2014) **20**, 797–812

BIODIVERSITY RESEARCH

Uncertainty, priors, autocorrelation and disparate data in downscaling of species distributions

Petr Keil[1,2]*, Adam M. Wilson[1] and Walter Jetz[1]

[1]*Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 06520, USA,* [2]*Center for Theoretical Study, Charles University and the Academy of Sciences of the Czech Republic, Jilská 1, 11000 Praha 1, Czech Republic*

*Correspondence: Petr Keil, Center for Theoretical Study, Charles University and the Academy of Sciences of the Czech Republic, Jilská 1, 110 00 Praha 1, Czech Republic.
E-mail: pkeil@seznam.cz

ABSTRACT

Aim We provide a step-by-step demonstration of how a map of species' detections at continental extent can be downscaled to a fine-grain map of occurrence probabilities using a two-scale hierarchical Bayesian modelling (HBM). The method requires fine-grain environmental data, but it does not require fine-grain data on species detections. We demonstrate how it can incorporate spatial autocorrelation (SAC) and informative priors based on known habitat preferences, and how it can provide maps of uncertainty about the downscaled predictions.

Location USA.

Methods We used range map and point record data on the distribution of American three-toed woodpecker (*Picoides dorsalis*, Baird 1858) to produce a reliable coarse-grain (160 km × 160 km) map of the species' presences and absences. We developed an HBM combining coarse-grain information with fine-grain (20 km × 20 km) environmental data to predict probabilities of occurrence at the fine grain together with 95% prediction intervals. The model incorporated SAC in the form of conditional autoregressive (CAR) random effects. It also incorporated prior knowledge on habitat preferences in the form of prior distribution of parameters. We evaluated the predictions using 751 well-surveyed fine-grain cells.

Results Our HBM produced reliable fine-grain probabilities of occurrence that matched the detections and non-detections in the 751 validation cells well (Nagelkerke's R^2 = 0.69, AUC = 0.93). By mapping the uncertainty in the downscaled predictions, we identified areas of low uncertainty and high occurrence probability, as well as large areas of high prediction uncertainty. Mapping the autocorrelation term enabled to identify areas of likely spurious observations.

Main conclusions We demonstrate how hierarchical downscaling enables estimation of species distributions at grains finer than the grain of the original data. The approach can integrate various types of information on distribution and biology in a single statistical framework, and it enables propagating and mapping prediction uncertainty. Yet there are also computational challenges for large datasets.

Keywords

Cross-scale, dispersal limitation, maps of uncertainty, multigrain, niche modelling, small area estimation, species distribution modelling.

INTRODUCTION

Knowledge about geographical distribution of species is fundamental for both basic and applied ecology. Ideally, species' distributions are captured in maps with as much detail as possible (i.e. at high resolution, or fine grain). In reality, it is impossible to survey for all the species at all locations in the area of interest. Most large-scale data on species distributions do not come from rigorous or complete sampling, and they typically contain high rates of false non-detections (e.g. in

© 2014 John Wiley & Sons Ltd

DOI: 10.1111/ddi.12199
http://wileyonlinelibrary.com/journal/ddi

A Journal of Conservation Biogeography
Diversity and Distributions

case of point records; Fig. 1) or false detections (such as expert-drawn range maps; Fig. 1) and other kinds of bias (Graham & Hijmans, 2006; McPherson & Jetz, 2007; Beale & Lennon, 2012; Jetz *et al.*, 2012).

The common practice to work with biased point record data is to use presence-only or presence-'pseudoabsence' species-distribution modelling (SDM; Fig. 1; Phillips *et al.*, 2009; Peterson *et al.*, 2011). Although these are useful tools to create preliminary maps of habitat suitability for a species, it is not always straightforward to interpret these maps as they do not show actual probabilities of occurrence (Yackulic *et al.*, 2012). For example, the interpretation of the MaxEnt's habitat suitability index is a debated issue (Phillips, 2008), and the results are sensitive to the 'pseudoabsence' selection algorithm (Lobo *et al.*, 2010; Barbet-Massin *et al.*, 2012). Thus, it is often necessary to convert the suitability maps into binary presences–absences using an arbitrary threshold (Liu *et al.*, 2005), and it is difficult to map uncertainty about the predictions (Rocchini *et al.*, 2011). This constrains their application beyond exploratory analysis (Yackulic *et al.*, 2012).

In contrast, with reliable data on presences and absences, or with knowledge about detection probability of a species, SDMs become a powerful statistical tools that model actual probabilities of occurrence (or 'occupancy' in some circles; Guisan *et al.*, 2002; MacKenzie *et al.*, 2002, 2006; Peterson *et al.*, 2011; Kéry *et al.*, 2013) and uncertainty about the predictions (using prediction intervals; Geisser, 1993). An additional advantage of such probabilities is that they do not need to be arbitrarily thresholded into binary presences–absences for further ecological applications (Storch *et al.*, 2003; Šizling & Storch, 2004; MacKenzie *et al.*, 2006). Finally, parametric presence–absence SDMs can be used for formal statistical inference, hypotheses testing and estimation of uncertainty about model parameters (using confidence or credible intervals).

Some data on species distributions that cover large continuous geographical areas are reasonably accurate for

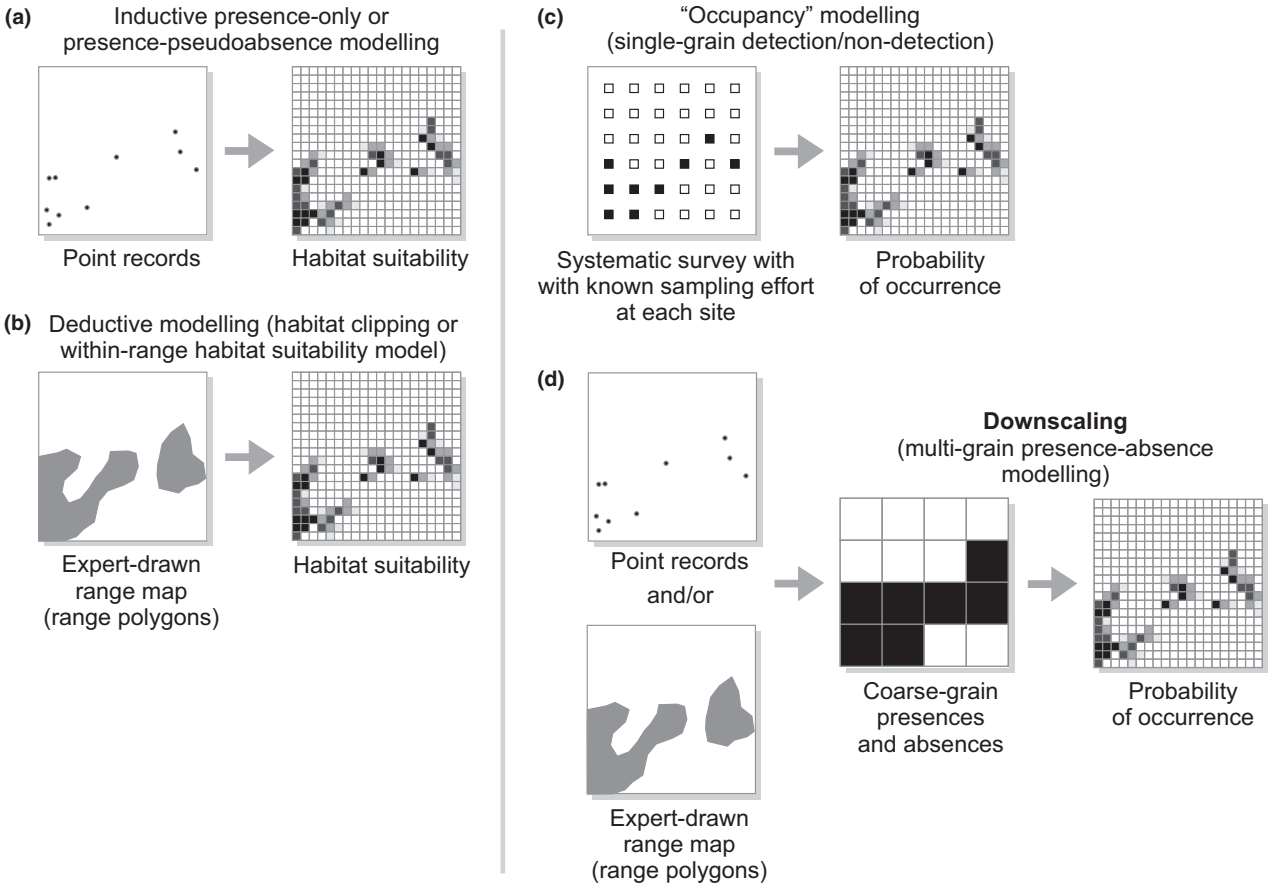

Figure 1 Four approaches for deriving high-resolution gridded distribution maps using data on species' distributions. (a) Point records data are fitted into a fine-grain grid and subjected to inductive presence-only or presence–pseudoabsence SDM. (b) Expert-drawn range maps are fitted into a fine-grain grid, and some parts of the range are 'clipped out' according to the known species' habitat requirements (Jetz *et al.*, 2007; Rondinini *et al.*, 2011). (c) Systematically collected data on species occurrences are subjected to statistical presence–absence modelling (also known as 'occupancy modelling;' MacKenzie *et al.*, 2006). (d) Downscaling approach proposed in this article: scale-free point records or range maps are fitted into a coarse-grain grid and then subjected to downscaling. We stress that the downscaling approach only works when fine-grain data on environment are available.

describing likely presences and absences of species at coarse grains (see the next section). These include expert range maps of continental to global extents (Ceballos & Ehrlich, 2006; Jetz *et al.*, 2007; IUCN, 2012) and broad-scale atlas efforts in selected parts of the world (Hagemeijer & Blair, 1997; Harrison *et al.*, 1997; Lahti & Lampinen, 1999). However, their grain is too coarse for many applied purposes and too coarse to be straightforwardly used in models of fine-grain processes governing species distributions. With the increasing availability of high-resolution (fine grain) environmental data (Jetz *et al.*, 2012), ecologists have started to explore the potential to combine these data with coarse-grain species' distributions to produce fine-grain maps of occurrence (Araújo *et al.*, 2005; McPherson *et al.*, 2006; Niamir *et al.*, 2011; Rondinini *et al.*, 2011; Bombi & D'Amen, 2012).

This is what we hereafter call *downscaling* of species-distribution models (Fig. 1). Similar to the conventional SDM approaches mentioned above, downscaling aims to produce fine-grain maps of species distributions. However, instead of predicting distributions in other (unsurveyed) locations at the same resolution as the resolution of the original distributional data, it predicts them at (unsurveyed) resolutions that are finer than the resolution of the original data. The advantage of downscaling is that it can exploit the coarse-grain data which in many cases represent the only knowledge on species' distributions.

The present paper follows up on previous work (Keil *et al.*, 2013) where we introduced the use of hierarchical Bayesian modelling (HBM) for downscaling of species-distribution models. Here, we (1) re-introduce a general statistical framework to downscale species' probability of occurrence ('occupancy') to finer grains using HBM. (2) We show how to incorporate spatial autocorrelation (SAC) and prior knowledge on habitat preferences into the downscaling framework. (3) We illustrate how this approach allows the mapping and interpretation of uncertainty about the downscaled predictions. (4) We illustrate our methods using a case study of distribution of American three-toed woodpecker (*Picoides dorsalis*, Baird 1858) in the United States. Our paper presents a statistical framework for modelling of species distributions across multiple scales that can integrate different kinds of knowledge on species distributions and biology.

COARSE-GRAIN PRESENCES AND ABSENCES

We purport that all primary human knowledge on species' distributions is in some form based on point observations in the field. In practice, this information is converted into, for example, geo-referenced point records with known or unknown spatial uncertainty (Guralnick *et al.*, 2007), gridded census data (Hurlbert & White, 2005; Hurlbert & Jetz, 2007) and/or converted by experts into two-dimensional abstractions called range maps.

As geometrical abstractions, point records and range maps do not have resolution or grain (similarly to a point mass in physics). They both gain a specific spatial resolution only after they are fitted (or aggregated) into a grid (Fig. 1), which is often practical. However, if point records or expert range maps are fitted into a grid that is too fine, range maps will generate unacceptable rates of false detection (errors of commission; Liu *et al.*, 2011), while point records will generate false non-detections (errors of omission; Liu *et al.*, 2011); see also Graham & Hijmans (2006) and Jetz *et al.* (2012). Also, many regional distributional atlases have already been compiled and published solely at a coarse grain (Hagemeijer & Blair, 1997; Harrison *et al.*, 1997; Lahti & Lampinen, 1999).

Hurlbert and Jetz (2007) showed that, in South Africa and Australia, record-based atlases and range maps give congruent (i.e. reliable) estimates of species richness from resolutions of at least 100 km × 100 km or 200 km × 200 km. This was later confirmed by Hawkins *et al.* (2008) using European datasets. La Sorte and Hawkins (2007) performed further theoretical exploration of the issue although they do not give a clear recommendation on reliability of particular grains. Ever since, the recommendation is that macroecological studies on groups such as vertebrates that use expert range maps to explore large-scale patterns of diversity should be conducted using coarse grains (roughly 100 km × 100 km or coarser).

The recommendation for range maps ultimately stems from the simple logic that the coarser the resolution, the lower the rate of false detections (Hurlbert & Jetz, 2007) – by coarsening the grain, we can be more confident that a species could be observed somewhere in a grid cell, but we are less confident about the exact point locations of the observations within the cell. Inverse logic should hold for point record data: the coarser the resolution, the lower the number of false non-detections (see Appendix S2 for demonstration of this in Supporting Information). Moreover, aggregating the point data to grids can potentially correct for georeferencing errors which can easily reach 10–20 km for older museum specimens and tropical regions – the rationale is that, instead of saying that a species was observed exactly at a given point, we state that the species was observed somewhere in a larger grid cell.

However, we warn that a comprehensive theory of spatial scaling of false negatives and false positives is yet absent (see Appendix S2). There is a large body of literature that deals with local-scale detectability of species, false absence rates and how these vary as a function of detection method and the environment (MacKenzie *et al.*, 2002, 2006; Tyre *et al.*, 2003; Royle & Dorazio, 2008; Webster *et al.*, 2008; Kéry *et al.*, 2013). Attempts appear to acknowledge the issue of false non-detections at large-scale models (Karanth *et al.*, 2009). However, there is still no literature on how these issues vary as a function of grain resolution. An additional complication is the rate of misidentifications (Miller *et al.*, 2011), which we assume to be either absent or negligible when compared to the other observational errors, which may not necessarily be true (Miller *et al.*, 2011). Again, spatial scaling of the misidentification bias in biodiversity data is an unknown territory.

In this study, we aggregated point observations in order to generate a coarse-grain dataset to illustrate downscaling. An alternative modelling approach would be to integrate the various data types (grid, point and polygon) within the model rather than performing this *ad-hoc* aggregation prior to model fitting. However, our primary objective with this study is to illustrate downscaling from one spatial grain to a finer grain (McInerny & Purves, 2011; Keil *et al.*, 2013) using a hierarchical model (Clark & Gelfand, 2006). This facilitates probabilistic inference about unobserved (latent) fine-grain components of the models, given that other fine-grain variables are known (i.e. the fine-grain data on environment in our case). Such models enable downscaling species probability of occurrence to grains finer than the grain of the original distributional data (Fig. 1).

THE RATIONALE OF HIERARCHICAL DOWNSCALING

Here, we show how the coarse-grain detections and non-detections can be downscaled to finer resolutions and used to study species–environment associations at finer grain that is limited only by the grain of the environmental data. To perform the downscaling, we first need to formalize the link between fine-grain environment, unknown fine-grain species occurrences and known coarse-grain species occurrences (introduced by Keil *et al.*, 2013).

Let P_i equals the probability that the ith ($i \in 1{:}I$) coarse-grain grid cell is occupied by the species of interest, and p_{ij} equals the probability that the species occupies the jth fine-grain grid cell ($j \in 1{:}n$), where I is the number of coarse-grain cells, and n is the number of fine-grain cells in coarse-grain cell i. Each coarse-grain grid cell is composed of n fine-grain grid cells, and P_i can be defined as one minus the joint probability of absence ($1\text{-}p_{ij}$) in each interior fine-grain grid cell j:

$$P_i = 1 - \prod_{j=1}^{n} (1 - p_{ij}) \tag{1}$$

Note that we are treating the fine-grain absences as independent events, and thus, the joint probability of fine-grain absences is equal to the product of individual probabilities. Equation 1 gives the link between probabilities of occurrence at two grains (resolutions), but it can be generalized to model the relationship between any number of grains. Let us now model the relationship between (unknown) p_{ij} and a vector of environmental variables (X_{ij}) for each location using a function $f()$:

$$p_{ij} = f(X_{ij}) \tag{2}$$

The $f()$ can be any commonly used function from sigmoidal to more complicated unimodal or multimodal functions. GAM and GLM (using logit or probit link functions) offer a first-choice set of such functions, although other link functions may be useful for species with unbalanced presences and absences such as the extreme value link function (Wang & Dey, 2010) or the symmetric power link function (Jiang *et al.*, 2014). Equation 2 can also be extended by adding a coarse-grain spatial random effect ρ_i (or even a fine-scale ρ_{ij}) to incorporate SAC into the model (Latimer *et al.*, 2006, and the next section):

$$p_{ij} = f(X_{ij}, \rho_i) \tag{3}$$

We can now combine equations 1–3 to link the observed coarse-grain presences–absences (O_i) with the fine-grain environment (X_{ij}). O_i can be treated as an outcome of a Bernoulli trial with probability P_i of observing the species in the ith coarse-grain grid cell:

$$O_i \sim \text{Bernoulli}\left(1 - \prod_{j=1}^{n} (1 - f(X_{ij}, \rho_i))\right) \tag{4}$$

Equation 4 reveals that even though we may not have the fine-grain detection/non-detection data, we can still estimate their relationship with fine-grain environment. In principle, this could be accomplished using maximum likelihood estimation, or Markov chain Monte Carlo (MCMC) techniques to sample from posterior distributions of parameters of $f()$. The estimated parameters of $f()$ can be then used to predict p_{ij} using equation 3. Alternatively, the posterior distribution of each p_{ij} can be monitored during the MCMC sampling.

SPATIAL AUTOCORRELATION

Spatial autocorrelation, that is, the higher similarity of closer locations, is a common phenomenon in ecology (Lichstein *et al.*, 2002; Latimer *et al.*, 2006; Dormann, 2007). In species distributions, SAC implies that presence or absence at one grid cell is not independent from presence or absence in a nearby cell (Latimer *et al.*, 2006). At large scales, SAC in the probability of occurrence of a species can emerge for at least two general reasons: (1) dispersal processes which include conspecific attraction and territoriality, or (2) a spatially autocorrelated niche component of the species that is not accounted for in the available environmental data (Dormann, 2007). During the last decade, numerous spatially explicit methods to account for SAC emerged and are summarized in a review by Dormann (2007).

Incorporating SAC into a downscaling model has several practical advantages. The posterior values of spatial component can be mapped to provide new insights into spatial processes that were not accounted for by the environmental covariates (Borcard *et al.*, 2004). Furthermore, by separating the contribution of the environment and the spatial effects to the expected probability of presence, it is possible to estimate aspects of both the potential and the realized niches. The spatial effects bring the model closer to predicting the actual probabilities of occurrence p (the realized distribution) by accounting for regions with more (or fewer) observed occurrences than would be expected given the potential distribution. For example, if there are environmentally suitable fine-grain habitats within an unoccupied coarse-grain grid cell, the spatial component will allow the p_{ij} in that cell to be low.

Alternatively, probabilities predicted by 'spatially ignorant' presence–absence statistical models (e.g. non-spatial logistic regression) are by definition biased towards the more prevalent category – either presence or absence (McPherson & Jetz, 2007). In case of a relatively rare species distributed over an area of tens of thousands of grid cells, the absences at fine grain would have such overwhelming effect that no matter how strong the association with any environmental variable, the predicted p_{ij} would always be extremely low. A spatial component ρ allows the p_{ij} values in a region to be low (even if environmentally suitable) if there are few observed presences. This effectively balances the presence/absence ratio, while avoiding subsampling of the data (as done by McPherson & Jetz, 2007). Finally, the way we model SAC (through ρ) can account for the effect of major dispersal barriers (mountain ranges, deserts) while assuming that within each occupied coarse-grain cell *i*, the fine-grain suitable habitats always lay within dispersal radius of the species.

PRIOR KNOWLEDGE ON HABITAT PREFERENCES

There is an immense and growing body of natural history knowledge pertaining to suitable habitats (environmental conditions) in which species are typically found. Such data can be provided by distributional atlases, naturalist handbooks, databases (IUCN, 2012) or directly provisioned expert knowledge. In cases with limited fine-scale occurrence data, such information may be useful for improving predictions. A crude way of using prior information on habitat preference is to use 'deductive modelling' or habitat-based 'clipping' (Fig. 1) of expert-drawn range maps to partly eliminate false presences at fine grain (for broad-scale applications see Jetz *et al.*, 2007; Rondinini *et al.*, 2011). The approach is equivalent to fitting a logistic function with extremely high or low prior $\beta \rightarrow \pm\infty$ (Fig. 2a). However, this is straightforward only when we can divide habitats into strictly suitable or unsuitable category. In reality, prior knowledge often has a semi-quantitative continuous character (e.g. 'the species prefers mountains to lowlands'), and there is a great deal of associated uncertainty. The Bayesian framework can incorporate such prior knowledge (with associated uncertainty) into the model and propagate the uncertainty through to the results (Clark & Gelfand, 2006; McCarthy, 2007), although prior specification is a challenging (and often under-appreciated) element of model development (Winkler, 1967).

Presence–absence models do not directly estimate the outcomes (presence or absence) but typically logit- or probit-transformed probability of observing the response variable (i.e. outcome of a Bernoulli trial). This complicates the process as it is not straightforward how to convert, for example, an expert's vague verbal description of a suitable habitat ('the species prefers mountains to lowlands') into a probability distribution of a model parameter on a logit scale. Gelman *et al.* (2008) provide some guidelines for prior elicitation in this type of model, recommending specification of the prior distributions directly on the logit scale. In Fig. 2, we show that this can be done by first plotting several hypothetical responses using various parameter values (Fig. 2a), which can help to estimate the prior distribution (Fig. 2b) (Winkler, 1967), or at least its sensible upper and lower bounds.

Adding this information has the potential to affect the results (otherwise there would be no point), and so models that use informative priors have been criticized for introducing subjectivity (Lele & Dennis, 2009). A real danger is that small datasets and informative priors can lead to posteriors that are driven by the priors rather than the data (Van Dongen, 2006). However, in situations where the goal is to use all available information to make the best possible predictions for policy or management, it can be beneficial to explicitly incorporate prior knowledge (Fienberg, 2011). Furthermore, in some situations, the use of even very broad ('uninformative') priors can bias the results (Van Dongen, 2006); thus, the sensitivity of the posterior distributions to the priors (whether informative or uninformative) should be assessed (Van Dongen, 2006).

UNCERTAINTY IN DOWNSCALED PREDICTIONS

Currently, the approach to express uncertainty in SDM predictions is rooted in the paradigm of ensemble modelling (Araújo & New, 2007) where model predictions from different modelling techniques are put together and prediction uncertainty is measured as the discrepancy in the predictions produced by different techniques (Hartley *et al.*, 2006; Buisson *et al.*, 2010; Luoto *et al.*, 2010). This uncertainty can then be mapped alongside of the mean of the predictions. Interestingly, the concept of estimating prediction uncertainty within a single modelling technique is almost absent in current SDM literature (with a few notable exceptions, for example Royle *et al.*, 2002; Webster *et al.*, 2008; Ibáñez *et al.*, 2009; Chakraborty *et al.*, 2011; Kéry *et al.*, 2013).

Rocchini *et al.* (2011) call for 'maps of ignorance' which would provide a quantitative *input* into an SDM so that it takes into account our uncertainty about species-distributional data (for example arising from the presence-only character of the point records). Beale and Lennon (2012) express similar sentiment calling for correct quantification of uncertainty in SDM *output*. They also suggest that rather than attempt to eliminate uncertainty completely (something impossible by definition), it is better to quantify uncertainty correctly. The papers review potential sources of uncertainty but do not offer an exact solution that would propagate uncertainty through a model into the predictions, although Beale and Lennon (2012) briefly mention the potential of hierarchical models to do so.

The uncertainty in our case study results mostly from errors in the distributional data, model specification, variance of prior parameter distributions, model fitting and from particular number of grid cells (sample size) at each spatial resolution. One of the advantages of Bayesian framework is

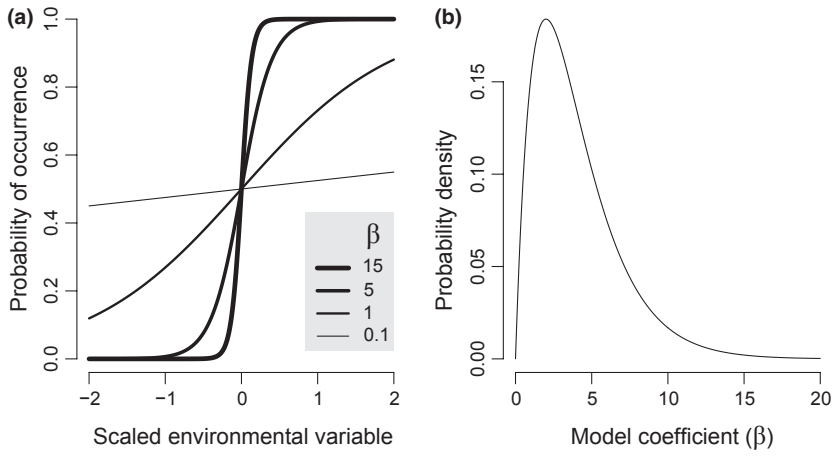

Figure 2 Illustration of how different values of β (see equation 5) affect the probability of occurrence. This can be used to specify informative prior distributions of model parameters. Panel (a) shows how the 'slope' (β) of logistic function changes its shape from a step-like function (β = 15) to a mild increase (β = 0.1). Panel (b) shows the probability density function used as informative priors on β_1 and β_4 in Downscaling model 1. The function is a gamma distribution with rate parameter 2 and scale of 0.5.

that it naturally quantifies and propagates all of the uncertainties simultaneously. As a result of the modelling procedure, we not only have the full posterior distribution (and hence uncertainty) of parameter estimates, we can also estimate prediction intervals PI (Geisser, 1993) of the estimated probabilities of occurrence at each grid cell. The spans of PIs (the magnitude of the uncertainty) can then be mapped alongside the actual predicted probabilities.

We note that in presence–absence SDMs, the data are usually modelled by Bernoulli distribution whose parameter (p: the probability of success) is bounded between 0 and 1. Hence, the closer is the mean p to 0 or 1, the narrower is the scope for the possible uncertainty around p. We suggest that one way to (partly) avoid the strong effect of the boundary on p is to describe the posterior distributions of p by quantiles instead of moments. We also suggest that this is a potentially fruitful direction of future research.

CASE STUDY: AMERICAN THREE-TOED WOODPECKER

Case study dataset

To demonstrate our approach, we selected the American three-toed woodpecker (*Picoides dorsalis*, Baird 1858) – a habitat specialist that depends on occasionally disturbed (by burns or beetle infestation) coniferous forests in high elevations (del Hoyo *et al.*, 2002; Imbeau & Desrochers, 2002; Wiggins, 2004; Zarnetske, 2006; Gagné *et al.*, 2007). The conservation importance of the species is unclear due to its low abundance in mostly temporary habitats and uncertainties in temporal trends of abundance (Wiggins, 2004).

Species-distribution data

We extracted all available point records (detections) of the species within mainland United States from the freely available eBird reference dataset 3.0 (Sullivan *et al.*, 2009; Munson *et al.*, 2011) which listed a total of 2261 detected woodpecker presences in 4.6 million sampling events. The sampling events consist of lists of species that were detected at a particular site. For each sampling event, the time, date, duration, surveyed area and number of observers are recorded. We also digitized three available expert-drawn range maps of the woodpecker in the USA. These came from (1) NatureServe (Ridgley *et al.*, 2007), (2) del Hoyo *et al.* (2002) and (3) Wiggins (2004). We decided to use only the range map of Wiggins (2004) which in relation to available point data offered the best detail and fit with the known habitat requirements. We put together information from both the range map by Wiggins (2004) and eBird point records to create a 160 km × 160 km grid of presences and absences (Fig. 3). If the range map overlapped a grid cell, it added an extra observation in that grid cell. The woodpecker was considered to be present in each 160 × 160 km grid cell if there were at least two observations in that cell. Hence, if only the range map indicated presence, but there were no actual point observations in the cell, the species was considered absent (we are aware that this would be highly problematic for species with too small or very spatially biased sample). Also, if there was only one point observation in the grid cell and no overlap with the range map, the species was considered absent (see Fig. 3a for the resulting coarse-grained map). We chose the 160 km × 160 km grain as it is, for North American birds, coarse enough to give congruent patterns of species distributions (as well as more reliable presences and absences) for both point records and range maps (Hurlbert & Jetz, 2007; Hawkins *et al.*, 2008).

Validation data

The eBird reference dataset offered high-quality fine-grain data for validation of our downscaled predictions (Fig. 3b). We first selected 'well-sampled' fine-grain 20 km × 20 km grid cells with more than 50 eBird sampling events (Sullivan *et al.*, 2009; Munson *et al.*, 2011). From these, we selected 375 grid cells in which the woodpecker was observed and a randomly selected set of 188 high-altitude grid cells (above 1000 m) and 188 low-altitude grid cells (below 1000 m) in

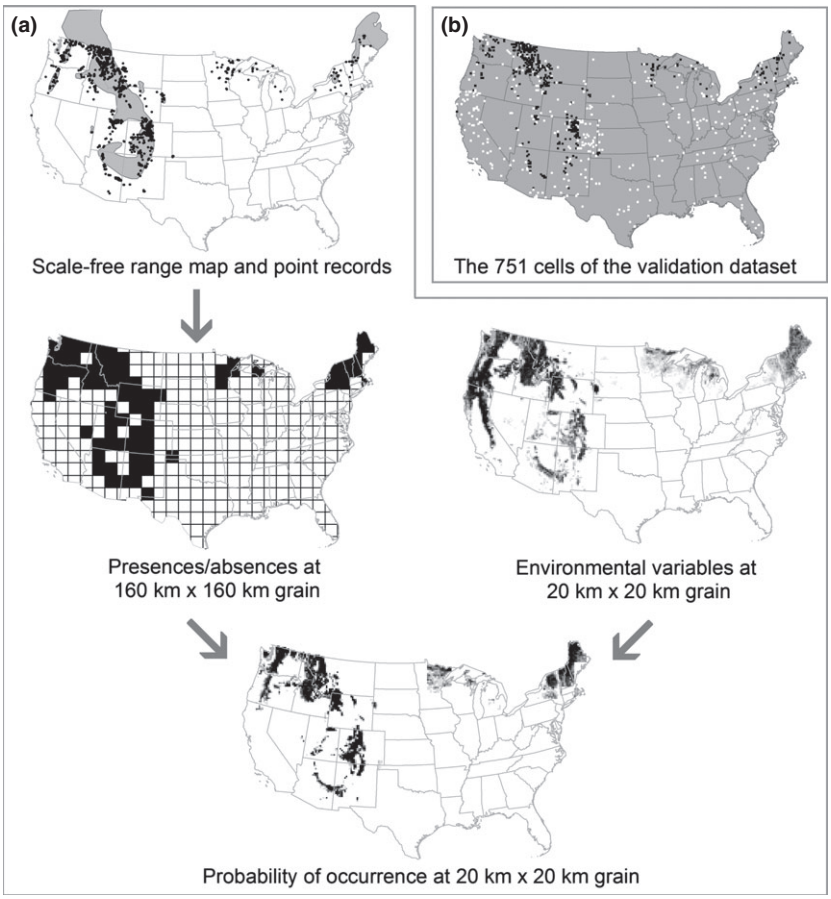

Figure 3 Graphic representation of the goals of our case study. (a) Point records (black dots) and expert-drawn range maps (grey polygon) for the American three-toed woodpecker (*Picoides dorsalis*) are fitted into a grid of 160 km × 160 km to represent reliable presences and absences. These presence–absence data are linked to fine-grain environmental conditions by a hierarchical model which treats the unobserved fine-grain probability of occurrence of the species as latent variable. The (downscaled) fine-grain probabilities of occurrence are estimated by MCMC sampling, monitored and plotted on a map. (b) To validate the downscaled fine-grain probabilities, we compare them with the observed fine-grain occurrences in the 751 well-surveyed cells of the validation dataset. Black points indicate detections, white are non-detections.

which the woodpecker was never observed (see the map in Fig. 3b). The 1000 m threshold was used to provide a sufficient number of high-altitude cells; random sampling from the whole USA would give almost none of these. We judged the performance of the downscaling models (see below) by how well their predictions matched the observed fine-grain occurrences in the total of 751 cells of the fine-grain validation dataset (Fig. 3b).

Environmental variables

We used four environmental variables that were selected to represent potential drivers of the woodpecker's geographical distribution in the 20 km × 20 km fine-grain grid. First, the bird prefers higher elevations from 1300 m up to 3350 m in West and from 360 to 1250 m in East United States (del Hoyo *et al.*, 2002) and coniferous forests. Hence, we calculated mean altitude (ALT) in each fine-grain grid cell using 2.5 arc-min SRTM dataset (http://www2.jpl.nasa.gov/srtm/). Second, the bird lives in coniferous forests, preferentially but not exclusively with spruce present (del Hoyo *et al.*, 2002; Imbeau & Desrochers, 2002; Wiggins, 2004; Zarnetske, 2006; Gagné *et al.*, 2007).

We calculated (F) square-root-transformed total area of potentially suitable coniferous forest types (categories Douglas-fir, White-red-jack pine, Spruce-fir, Ponderosa pine, Western white pine, Lodgepole pine and Fir-spruce) in each grid cell using the USDA Forest Service and USGS AVHRR-derived forest cover dataset available at http://nationalatlas.gov/atlasftp.html. Third, species distributions at large scales are generally considered to be correlated with climate (Peterson *et al.*, 2011), and we selected two climatic variables to represent precipitation and temperature regimes (also a proxy of fire risk) that were the least correlated with altitude and coniferous forest area. These were mean annual temperature (T) and precipitation (PD) in the driest month, derived from 2.5 arc-min WorldClim dataset (Hijmans *et al.*, 2005). All four environmental variables were cantered to zero mean and standardized to variance of 1. Pearson's correlations between the scaled variables were −0.52 (ALT, T), −0.51 (ALT, PD), 0.36 (ALT, F), −0.44 (T, F) and −0.075 (PD, F).

Case study models and methods

Our specific goal was to predict (and validate) the probability of occurrence of the focal species in a 'fine' grid of 20 km × 20 km (20,240 cells) using the data on detections and non-detections in the 'coarse' 160 km × 160 km grid (372 cells) and environmental variables at the 'fine' 20 km × 20 km grain. We use the terms 'fine' and 'coarse' only for convenience; the method could be applied at any two spatial grains.

We constructed five models (Table 1) predicting probability of occurrence of the American three-toed woodpecker at the fine 20 km × 20 km grain (Fig. 5), where the first two are reference models and the last three are the downscaling models:

Fine-grain model was the simplest model in the set and was intended as a reference model. It did not involve any downscaling, and it was fitted entirely at the fine-grain using the validation dataset of the 751 fine-grain grid cells (see the previous section). The model is a classical logistic regression:

$$o_k \sim \text{Bernoulli}\left(\left(1 + e^{-(\beta_0+\beta_1 \text{ALT}_k+\beta_2 \text{T}_k+\beta_3 \text{PD}_k+\beta_4 \text{F}_k)}\right)^{-1}\right) \quad (5)$$

where o_k is the observed fine-grain presence (detection) or absence (non-detection) in the kth grid cell ($k \in$ 1: 751) of the validation dataset.

Full 2-scale model was the most complex 'full-feature' model in the set, and it was also a reference model meant to be compared with the downscaling models below. The model integrated spatial effects at the coarse grain, and it used logistic function to link fine-grain probabilities of occurrence p_{ij} to fine-grain environmental data.

The model operates simultaneously at two spatial grains at which the observed presences and absences are linked to probabilities of occurrence. First, the observed presence or absence in each coarse-grain grid cell (O_i) is linked to the interior fine-grain grid cell probability of occurrence (p_{ij}) (see equation 1 for the meaning of i and j):

$$O_i \sim \text{Bernoulli}\left(1 - \prod_{j=1}^{n} (1 - p_{ij})\right) \quad (6)$$

At the same time, the observed presence and absence in each well-surveyed fine-grain grid cell of the reference dataset (o_{ij}) is linked to the fine-grain probability of occurrence (p_{ij}):

$$o_{ij} \sim \text{Bernoulli}(p_{ij}) \quad (7)$$

The fine-grain probability of occurrence is estimated using a logistic regression with the fine-grain environmental variables (see also equation 3):

$$p_{ij} = f(X_{ij}, \rho_i) = (1 + e^{-(\beta_0+\beta_1 \text{ALT}_{ij}+\beta_2 \text{T}_{ij}+\beta_3 \text{PD}_{ij}+\beta_4 \text{F}_{ij}+\rho_i)})^{-1} \quad (8)$$

where ρ_i is the coarse-grain spatial random effect that accounts for SAC and O_i is the observed presence or absence of the species at the 160 km × 160 km grain. For its intuitive nature and widespread use (Lichstein *et al.*, 2002; Latimer *et al.*, 2006; McPherson & Jetz, 2007), we chose the intrinsic conditional autoregressive (CAR) process (Besag *et al.*, 1991) using the car.normal distribution in the GeoBUGS module in OpenBUGS (Lunn *et al.*, 2009) to model the coarse-grain spatial random component ρ as follows:

$$\rho_i | \rho_{l,\, l \in \delta_i} \sim N\left(\bar{\rho}_i, \frac{w^2}{m_i}\right) \quad (9)$$

where δ_i is the set of neighbours of grid cell i, $\bar{\rho}_i$ is the mean of the spatial random effects of these neighbours, and m_i is the number of neighbours. The parameter w^2 is a variance term that specifies the magnitude of spatial variation. We

Table 1 Medians of fitted parameters (95% credible intervals in the brackets) of the four models predicting probability of occurrence of American three-toed woodpecker (*Picoides dorsalis*, Baird 1858) at the 20 km × 20 km grain. Performance metrics include Nagelkerke's R^2 (measures how well models fit the data) and AUC (measures how well models discriminate the data). Both metrics were calculated using the validation dataset

	Full 2-scale model	Fine-grain model	Downscaling model 1	Downscaling model 2	Downscaling model 3
Fine-grain data	Yes	Yes	No	No	No
Coarse-grain data	Yes	No	Yes	Yes	Yes
Autocorrelation (CAR)	Yes	No	Yes	Yes	No
Informative prior on β_1 and β_4	No	No	Yes	No	No
β_0	**−10.37** (−12.36, −8.85)	**−2.17** (−2.57, −1.8)	**−17.89** (−54.11, −9.89)	**−48.22** (−112.38, −19.15)	**−6.96** (−7.63, −6.43)
β_1 (ALT)	−0.14 (−1.95, 1.62)	0.12 (−0.12, 0.35)	**2.51** (0.53, 14.62)	6.98 (−1.36, 6.98)	**0.32** (0.06, 0.62)
β_2 (T)	**−2.2** (−4.46, 0)	**−1.89** (−2.35, −1.45)	**−5.42** (−18.7, −0.94)	**−13.27** (−33.41, −3.23)	**−1.22** (−1.71, −0.72)
β_3 (PD)	**−1.21** (−2.84, 0.34)	0.0065 (−0.24, 0.36)	0.86 (−3, 7.1)	4.73 (−3.05, 15.49)	**0.56** (0.18, 0.9)
β_4 (F)	**1.12** (0.64, 1.63)	**0.58** (0.4, 0.78)	**2.65** (0.99, 8.67)	**6.81** (1.2, 18.47)	**0.96** (0.7, 1.24)
Nagelkerke's R^2	0.768	0.602	0.69	0.665	0.58
AUC	0.958	0.902	0.934	0.930	0.895

Abbreviations next to βs stand for ALT – Altitude; T – mean annual temperature; PD – precipitation in driest month; F – area of coniferous forest. Parameters in bold indicate that 95% credible intervals do not overlap zero.

chose to incorporate SAC only at the coarse grain because of the computational limitations (2GB limit of the 32-bit OpenBUGS v3.2.1; OpenBUGS Foundation). Also, we were trying to capture coarse-scale processes with the CAR such as dispersal limitation, rather than fine-scale variability.

Downscaling model 1 is almost identical to the Full 2-scale model. The only difference is that it does not use any observed presence and absence (o_{ij}) at the fine-grain cells. It is fitted using only the O_i and X_{ij} data. To summarize this model formally, it consists of the following equations (the notation is identical to the Full 2-scale model):

$$O_i \sim \text{Bernoulli}\left(1 - \prod_{j=1}^{n} (1 - p_{ij})\right)$$

$$p_{ij} = f(X_{ij}, \rho_i) = \left(1 + e^{-(\beta_0 + \beta_1 \text{ALT}_{ij} + \beta_2 \text{T}_{ij} + \beta_3 \text{PD}_{ij} + \beta_4 \text{F}_{ij} + \rho_i)}\right)^{-1}$$

$$\rho_i \mid \rho_{l,l \in \delta_i} \sim N\left(\bar{\rho}_i, \frac{w^2}{m_i}\right)$$

Another distinct feature is that the model used informative priors (Fig. 2). We set informative prior distributions only on coefficients β_1 (the known positive response of the species to increasing altitude; del Hoyo *et al.*, 2002; Gagné *et al.*, 2007; Imbeau & Desrochers, 2002; Wiggins, 2004) and β_4 (the known positive response to the increasing area of coniferous forests; del Hoyo *et al.*, 2002). In both cases, we did not expect a sharp step-like response of the species to environment (i.e. 'the species almost certainly and suddenly appears after crossing a certain threshold value along the gradient of increasing forest area or elevation'; $\beta > 15$), nor did we expect a weak gradual response (i.e. 'the species' probability of occurrence increases very slowly along the whole gradient of increasing forest area or elevation'; $\beta > 0.1$). Hence, we used gamma distribution with rate parameter 2 and scale of 0.5 (Fig. 2b). For β_0, β_2 and β_3, we used normally distributed non-informative priors with zero mean and variance of 100.

Downscaling model 2 was formally identical to the Downscaling model 1, but it used non-informative priors on all coefficients $\beta_{0\ldots}\beta_4$ (normally distributed with mean 0 and variance of 100).

Downscaling model 3 differed from the previous models by not using the coarse-grain spatial random component ρ and no prior information. Its complete definition is (the notation is identical to the Full 2-scale model):

$$O_i \sim \text{Bernoulli}\left(1 - \prod_{j=1}^{n} (1 - p_{ij})\right)$$

$$p_{ij} = \left(1 + e^{-(\beta_0 + \beta_1 \text{ALT}_{ij} + \beta_2 \text{T}_{ij} + \beta_3 \text{PD}_{ij} + \beta_4 \text{F}_{ij})}\right)^{-1}$$

For all models, we estimated posterior distributions of β, ρ and p using OpenBUGS (Lunn *et al.*, 2009). We ran four MCMC chains, each of 200,000 iterations in total, discarded the first 150,000 iterations as burn-in and thinned the remaining 50,000 by saving every 50th iteration. We visually checked the resulting chains (of β) for convergence; we did not perform any additional formal convergence diagnostics. BUGS and R codes and data that were used to fit the models are provided in Appendix S1. Finally, we used the fine-grain validation dataset to measure discrimination capacity of all of the models (by AUC; Liu *et al.*, 2011) and their goodness of fit (by Nagelkerke's pseudo R^2; Nagelkerke, 1991).

Case study results

Model coefficients and spatial patterns of the predictions from all five models were qualitatively similar (Table 1, Figs 4 and 5), but they differed in magnitude and uncertainty.

Full 2-scale model and fine-grain model

Both of the reference models gave the narrowest posterior distributions of coefficients (Table 1, Fig. 4). The Full 2-scale model outperformed all of the models by having the highest AUC and Nagelkerke's R^2 (Table 1, Fig. 5). Both models also produced large number of grid cells with high occurrence probability ($P > 0.5$) and low uncertainty (span of the 95% pred. intervals < 0.5) (Appendix S2). The overall prediction uncertainty was much lower in the two reference models than in the downscaling models (Appendix S2). Notably, the fine-grain model with no spatial random effects predicted high p_{ij} values in areas where the species has never been observed (e.g. Sierra Nevada, CA; Fig. 5).

Downscaling models

Downscaling models 1 and 2 (with spatial random effects) gave much broader posterior distributions of parameters (greater uncertainty) and larger magnitude of the effects of environmental variables compared with the previous two models (Fig. 4). The Downscaling model 2 with non-informative priors led to broadest posteriors (Fig. 4) and more extreme values of the coefficients (Fig. 4), as expected (informative priors tend to 'squeeze' the posteriors).

With AUC values of 0.93 and 0.92, the Downscaling models 1 and 2 were able to successfully discriminate presences and absences in the validation dataset (Table 1, Fig. 5). They also fit the data reasonably well as shown by the Nagelkerke's pseudo R^2 of 0.69 and 0.665 for the Downscaling models 1 and 2, respectively (Table 1, Fig. 5). Importantly, using informative priors in model 1 somewhat improved the predictions (higher AUC and Nagelkerke's pseudo R^2). Downscaling model 3 that used non-informative priors and no SAC performed worst (AUC = 0.895, Nag. R^2 = 0.58; Table 1, Fig. 5). Surprisingly, although the fine-grain Reference model would generally be considered to perform well (AUC = 0.902, Nag. R^2 = 0.602), it still performed worse than the first two downscaling models.

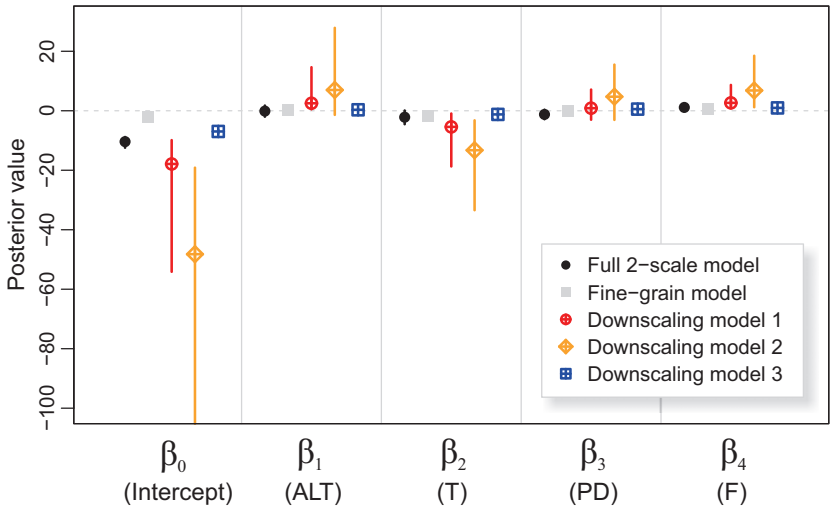

Figure 4 Median values of model parameters and their 95% credible intervals for Downscaling model 1 (informative priors, spatial random effects), Downscaling model 2 (non-informative priors, spatial random effects), Downscaling model 3 (non-informative priors, no spatial random effect) and Reference model (non-informative priors, no spatial random effects, fitted using the 751 well-surveyed cells), and the Reference model. Note that informative prior distributions were set only for β_1 and β_4 in Model 1 (see also Fig. 2). For exact values, see Table 1. Abbreviations below βs stand for ALT – altitude; T – mean annual temperature; PD – precipitation in driest month; F – area of coniferous forest.

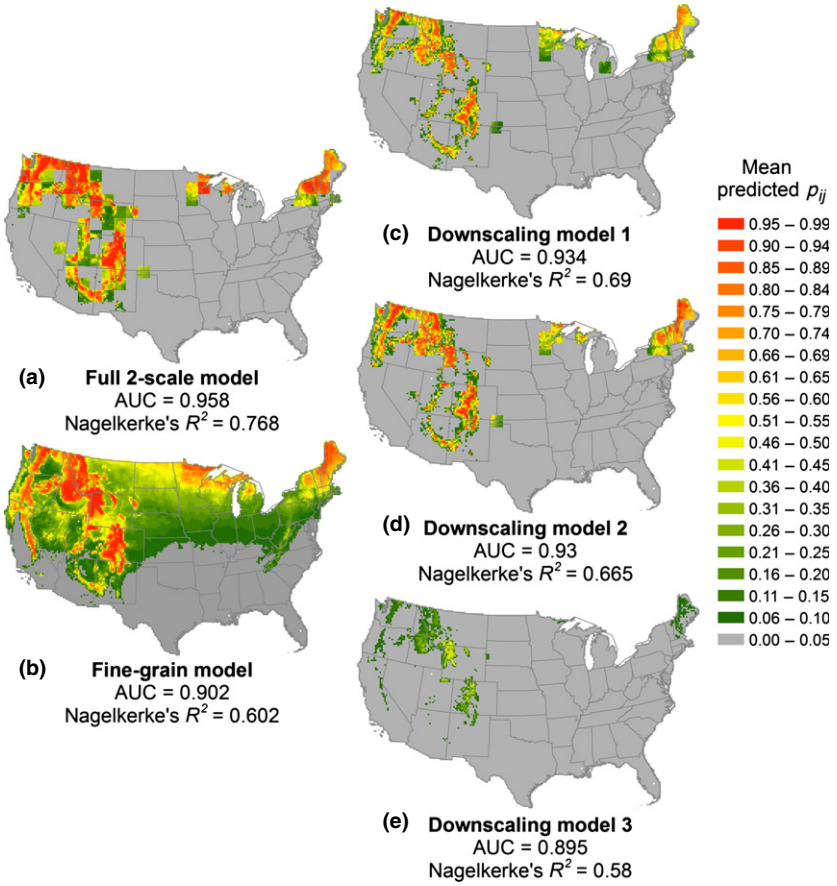

Figure 5 Mean predicted values of fine-grain (20 km × 20 km) probability of occurrence (p_{ij}) of the American three-toed woodpecker, as predicted by the five models (a–e). AUC and Nagelkerke's R^2 were calculated using the predicted p_{ij} values and the observed presences–absences in the 751 well-surveyed grid cells at the 20 km × 20 km grain of the validation dataset (Fig. 3b).

As it performed best, we further elaborate the predictions and their uncertainty of the Downscaling model 1 (see Appendix S2 for details on the other models) which incorporated both SAC and prior information (Fig. 6). We were able to identify vast areas with low probability of occurrence ($p_{ij} < 0.5$) and low uncertainty (the grey areas in Fig. 6e) as well as small number of areas where the presence of the species is likely ($p_{ij} > 0.5$) and more certain (the green areas in

Fig. 6e). However, we also identified considerable areas of very uncertain model predictions (the red and blue areas in Fig. 6e). Special attention should be paid to areas with extreme values of the spatial random effects ρ (Fig. 6d). Many of the low probability and high uncertainty areas (the red areas in in Fig. 6e) coincide with high values of the spatial random component (e.g. the coarse-grain grid cell overlapping borders of Texas, New Mexico, Colorado and Kansas) (Fig. 6d,e). Similarly, there are areas of seemingly suitable habitats that are, however, quite certainly not occupied by the species – these coincide with the extremely low values of spatial random effects (Fig. 6d).

DISCUSSION

We previously demonstrated the hierarchical downscaling approach in Keil *et al.* (2013) using a multispecies dataset with relatively small spatial extent from San Diego county, California, and downscaling from 15 km × 15 km to 5 km × 5 km. Here, we show that the hierarchical approach can be used to downscale maps of species distribution at much larger (near-continental) extents and over larger spans of grains (from 160 km × 160 km down to 20 km × 20 km). We also show that explicitly invoking SAC and prior information on species' habitat preferences can be beneficial for such downscaling efforts.

Benefits of spatially explicit modelling

As is often the case (Dormann, 2007), the incorporation of SAC (here as CAR random effects) improved performance of our models. However, there has been another benefit of the CAR random effects: it revealed specific areas of potentially dubious data quality or areas where our environmental predictors performed poorly. For example, note the relatively large negative spatial effects (ρ) in the south-western corner of the region (California and Nevada, Fig. 6d). This complements the non-spatial models which predict some probability of occurrence in California and Nevada, although the species was never observed there. This indicates that there were fewer occurrences documented in that region than would be expected given the environmental data. Patterns like this could be due to several factors, including (1) less sampling effort in that region leading to relatively more frequent false negatives in the data, (2) important environmental factors limiting the species that were not included in the model, (3) limited species' dispersal into the region or (4) biological interactions (e.g. interspecific competition) limiting the species' range. In this case, a closely related Black-backed Woodpecker (*Picoides arcticus*, Swainson, 1832) inhabits similar habitats and occupies much of the Sierra Nevada and nearby regions, potentially displacing *Picoides dorsalis* (Bock & Bock, 1974).

Conversely, there are several coarse grid cells in our case study data with presences that are unexpected given the environment. The model accounts for observations in those areas with relatively large positive spatial effects (see Fig. 6d). Those observations could be a result of misidentification, intraspecific variation in habitat preferences, infrequent occurrence (e.g. in sink locations on range edges) or an accidental visit of the organism. Incorporating spatial effects into the model allows the (spatially autocorrelated) possibility that a species could be absent from a region with a suitable environment or present in unsuitable environment. In the maps, these anomalous regions appear 'blocky' due to the scale at which we estimated the spatial effects. It would be possible, and perhaps desirable, to use a spatially continuous spatial model, such as a predictive process model (Banerjee *et al.*, 2008), to account for autocorrelation in place of the simple CAR used here. By design, spatial effects capture spatial variability that is unexplained by the rest of the model, and thus, they cannot explain the cause of the patterns in the posterior estimates of ρ. However, the mapped spatial random components are useful to generate additional hypotheses, consider alternative environmental data or identify dubious distributional data.

Benefits of prior knowledge

Without a fine-grain validation dataset, any downscaling effort will be a blind walk in an unknown territory where every additional piece of information can guide the model towards more accurate predictions, especially in case of species that are extremely rare or common at the coarse-grain (Keil *et al.*, 2013). In spite of its associated controversy, usage of prior information on model structure (Hooten & Wikle, 2008) and model parameters (Dupuis & Joachim, 2006; Pagel & Schurr, 2012; Zipkin *et al.*, 2012; Gopalaswamy *et al.*, 2013) is starting to find its way to occupancy- and species-distribution modelling. An extreme case of large-scale use of extremely strong priors (although not explicitly described as such) is the 'range-clipping' approach (Jetz *et al.*, 2007; Rondinini *et al.*, 2011; Fig. 1b). Here, we have outlined and tested a simple and conceptually consistent way to specify priors for parameters of logistic regression – even our rather conservative priors on only two model parameters led to decreased uncertainty in parameter estimates and to improved model performance.

Benefits of quantifying uncertainty

As noted earlier, maps of prediction uncertainty are still rare in the field of SDM (Rocchini *et al.*, 2011; Beale & Lennon, 2012). Studies that explicitly work with statistical uncertainty *sensu stricto* are also rare (Royle *et al.*, 2002; Webster *et al.*, 2008; Ibáñez *et al.*, 2009; Chakraborty *et al.*, 2011; Kéry *et al.*, 2013). We have shown that mapping of the uncertainty around single-model predictions can considerably change their interpretations. At first glance, the maps of mean or median predictions (Fig. 5) give seemingly straightforward and even 'pretty' description of where to expect the species (we note that this is how the vast majority of SDM results are presented). However, a closer look at the prediction

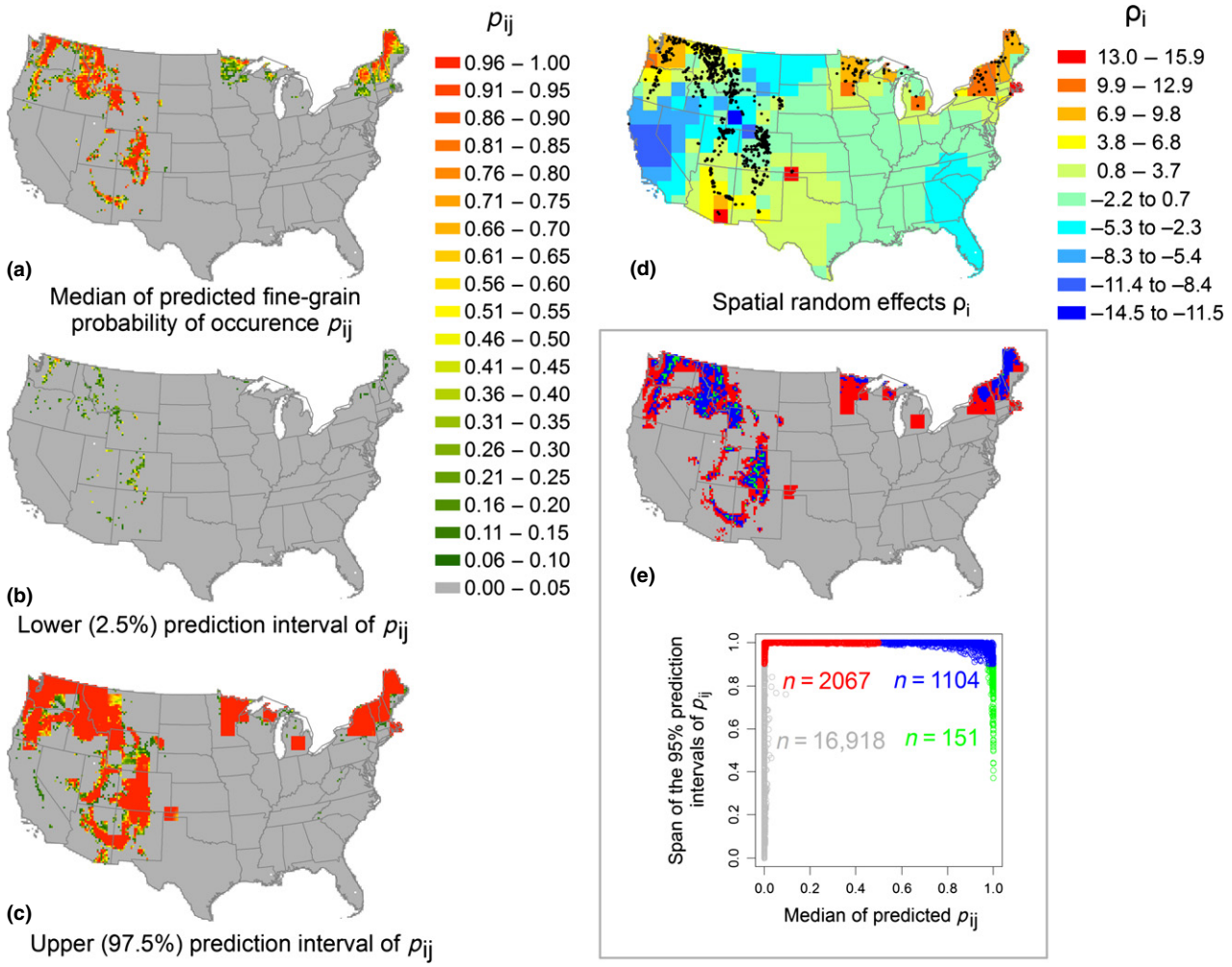

Figure 6 Detailed elaboration of the predictions (at the 20 km × 20 km grain) of the Downscaling model 1 that incorporated informative priors on habitat preferences and spatial autocorrelation. Medians (a) and 95% prediction intervals (b,c) of the probability of the woodpecker occurrence. For example, a 2.5 quantile (b) of 0.51 (yellow) indicates that there is only a 2.5% probability that the true probability of presence is less than 0.51. Conversely, a 97.5% quantile (c) of 0.05 (grey) indicates that there is a 97.5% probability that the true probability of presence is <0.05 (virtually certain to be absent). (d) Median values of the spatial random effects. (e) Fine-grain grid cells were classified into four categories, according to the predicted probability of the woodpecker's occurrence and uncertainty around this prediction. Appendix S2 provides equivalent maps for Downscaling models 2, 3 and the Reference model.

uncertainties (Fig. 6e) reveals that only a small fraction of the fine-grain grid cells give confident predictions of high occurrence probability.

One of the novelties of our approach is that our estimates of uncertainty incorporate the cross-scale relationships inferred by the model. So we have an explicit representation of where we can trust the model (e.g. the grey and green areas in Fig. 6e) and where we should be more wary (the blue and red areas in the same figure). The cells with the span of 95% CI of p_{ij} larger than 0.9 indicate areas with large uncertainty as to whether or not *P. dorsalis* occupies the cell, which represents the state of knowledge given the data.

We argue that quantifying (and communicating) the prediction uncertainties is vital to interpreting and/or using model results for decision-making. For example, consider two locations with median posterior of $p_{ij} > 0.95$ (indicating high probability of occurrence). Now imagine that the 95% CI of one location ranges from nearly 0 to almost 1 (a 'blue' pixel from Fig. 6e), while the other from 0.6 to nearly 1 (a 'green' pixel from Fig. 6e). We should be less confident (and less willing to act upon) on the information in the first location. The quantification of uncertainty provides an explicit mechanism to incorporate model uncertainty into decisions, such as conservation plans.

Furthermore, one could include additional sources of uncertainty such as in the underlying environmental data itself as explained in Wilson and Silander (2014). In contrast, many of the commonly used techniques (Elith *et al.*, 2006; Peterson *et al.*, 2011) are not capable of estimating uncertainty and result in predictions that offer only limited opportunity for inference (Chakraborty *et al.*, 2011; Yackulic *et al.*, 2012). Finally, quantification of uncertainty is especially

valuable for downscaling because, in many cases, no fine-grain validation data are available. Quantifying the prediction uncertainties will help us decide how much confidence to have in model predictions.

Warning: technical limits

So far it might have seemed that our approach has only benefits and advantages. There is, however, a severe limitation. Current MCMC samplers can have high computational demands and cannot be simply parallelized within individual MCMC chains. Moreover, the very nature of the modelling makes any analysis at large scales more difficult to code, debug and by orders of magnitude more time-consuming than the single-scale SDM techniques implemented in established software packages. A simple additional calculation or model modification can take days, weeks or even months to run in OpenBUGS or JAGS. We previously tested our downscaling approach on a simple and small dataset and corresponding models (Keil *et al.*, 2013) which did not pose such a challenge. Here, we increased the extent of the data and included the spatial random components – and proved it possible and beneficial. However, there was a cost of steep learning curve of the Bayesian methods, extensive data and model preparations, long waiting times and unpredictable software crashes (see Appendix S1 for more details).

These limitations must be addressed if the approach is to be used for large sets of species, for automatized model-selection purposes and cross-validation and for even larger extents than we used (e.g. global). We are optimistic that this will soon be possible. The emerging Hamiltonian Monte Carlo samplers (Hoffman & Gelman, 2011) and within-chain parallelizing procedures (Chakraborty *et al.*, 2010) may provide speedup. Finally, likelihood optimization algorithms (Bolker, 2008) may serve as a fast alternative to MCMC, although their capabilities to fit complex spatial models propagate uncertainty and use informative priors are limited.

Coda

Our current understanding of the global distribution of biodiversity is based upon observations of species occurrences at a variety of spatial grains and with different limitations ranging from presence-only point observations to more absence-focused expert-drawn range maps that are reliable at coarse scales of hundreds of km. The availability of fine-resolution data on our physical environment (e.g. topography and climate) has led to a biodiversity 'scale gap' in the data availability at different spatial resolutions that constrains conservation and management (Jetz *et al.*, 2012). Although potentially computationally challenging, the framework presented here helps to overcome the scale gap by integrating heterogeneous species' distributional data with fine-scale environmental data to enable statistical modelling of biodiversity at fine spatial resolution. Looking forward, we see great promise in this type of approach, including modelling species occurrences using scale-free non-homogeneous Poisson process models that can then be aggregated to any desired grain (Chakraborty *et al.*, 2011).

ACKNOWLEDGEMENTS

We are especially grateful to Marc Kéry and Oliver Schweiger for the thorough review. The research leading to these results has received funding from People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA Grant agreement no 302868. This work was also supported by NSF Grants DBI0960550 and DEB 1026764, NASA Biodiversity Grant NNX11AP72G to WJ and Yale YCEI fellowship funding to AMW, and by the Yale Program in Spatial Biodiversity Science and Conservation.

REFERENCES

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.

Araújo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2005) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, **14**, 17–30.

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

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, **3**, 327–338.

Beale, C.M. & Lennon, J.J. (2012) Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 247–258.

Besag, J., York, J. & Mollié, A. (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1–20.

Bock, C.E. & Bock, J.H. (1974) On the geographical ecology and evolution of the three-toed woodpeckers, *Picoides tridactylus* and *P. arcticus*. *American Midland Naturalist*, **92**, 397–405.

Bolker, B.M. (2008) *Ecological models and data in R*. Princeton University Press, Princeton.

Bombi, P. & D'Amen, M. (2012) Scaling down distribution maps from atlas data: a test of different approaches with virtual species. *Journal of Biogeography*, **39**, 640–651.

Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826–1832.

Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.

Ceballos, G. & Ehrlich, P.R. (2006) Global mammal distributions, biodiversity hotspots, and conservation. *Proceedings of the National Academy of Sciences USA*, **103**, 19374–19379.

Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M. & Silander, J.M. (2010) Modeling large scale species abundance with latent spatial processes. *The Annals of Applied Statistics*, **4**, 1403–1429.

Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M. & Silander, J.A. (2011) Point pattern modeling for degraded presence-only data over large regions. *Journal of the Royal Statistical Society. Series C (Applications)*, **60**, 757–776.

Clark, J.S. & Gelfand, A.E. (2006) *Hierarchical modelling for the environmental sciences: statistical methods and applications*. Oxford University Press, Oxford.

Dormann, C.F. (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129–138.

Dupuis, J.A. & Joachim, J. (2006) Bayesian estimation of species richness from quadrat sampling data in the presence of prior information. *Biometrics*, **62**, 706–712.

Elith, J., Graham, C.H., Anderson, P.R. *et al.* (2006) Novel methods improve prediction of species? Distributions from occurrence data *Ecography*, **29**, 129–151.

Fienberg, S.E. (2011) Bayesian models and methods in public policy and government Settings. *Statistical Science*, **26**, 212–226.

Gagné, C., Imbeau, L. & Drapeau, P. (2007) Anthropogenic edges: their influence on the American three-toed woodpecker (*Picoides dorsalis*) foraging behaviour in managed boreal forests of Quebec. *Forest Ecology and Management*, **252**, 191–200.

Geisser, S. (1993) *Predictive inference: an introduction*. Chapman & Hall, New York.

Gelman, A., Jakulin, A., Pittau, M.G. & Su, Y.-S. (2008) A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, **2**, 1360–1383.

Gopalaswamy, A.M., Royle, J.A., Delampady, M., Nichols, J.D., Karanth, K.U. & Macdonald, D.W. (2013) Density estimation in tiger populations: combining information for strong inference. *Ecology*, **93**, 1741–1751.

Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, **15**, 578–587.

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

Guralnick, R.P., Hill, A.W. & Lane, M. (2007) Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters*, **10**, 663–672.

Hagemeijer, W.J.M. & Blair, M.J. (1997) *The EBCC Atlas of European breeding birds*. European Bird Consus Council and T. & A. D. Poyser, Berkhamsted.

Harrison, J.A., Allan, D.G., Underhill, L.G., Herremans, M., Tree, A.J., Parker, V. & Brown, C.J. (1997) *The atlas of Southern African birds*. BirdLife South Africa, Johannesburg.

Hartley, S., Harris, R. & Lester, P.J. (2006) Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters*, **9**, 1068–1079.

Hawkins, B.A., Rueda, M. & Rodriguez, M.A. (2008) What do range maps and surveys tell us about diversity patterns? *Folia Geobotanica*, **43**, 345–355.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hoffman, M.D. & Gelman, A. (2011) The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. arXiv, 1111.4264.

Hooten, M.B. & Wikle, C.K. (2008) A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics*, **15**, 59–70.

del Hoyo, J., Elliot, A. & Sargatal, J. (2002) *Handbook of the birds of the world. Vol. 7. Jacamars to Woodpeckers*. Lynx Edicions, Barcelona.

Hurlbert, A.H. & Jetz, W. (2007) Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences USA*, **104**, 13384–13389.

Hurlbert, A.H. & White, E.P. (2005) Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecology Letters*, **8**, 319–327.

Ibáñez, I., Silander, J.A., Wilson, A.M., LaFleur, N., Tanaka, N. & Tsuyama, I. (2009) Multivariate forecasts of potential distributions of invasive plant species. *Ecological Applications*, **19**, 359–375.

Imbeau, L. & Desrochers, A. (2002) Foraging ecology and use of drumming trees by three-toed woodpeckers. *The Journal of Wildlife Management*, **66**, 222–231.

IUCN (2012) *The IUCN red list of threatened species*. IUCN, Cambridge.

Jetz, W., Wilcove, D.S. & Dobson, A.P. (2007) Projected impacts of climate and land-use change on the global diversity of birds. *PLoS Biology*, **5**, 1211–1219.

Jetz, W., McPherson, J.M. & Guralnick, R.P. (2012) Integrating biodiversity distribution knowledge: toward a global Map of Life. *Trends in Ecology & Evolution*, **27**, 151–159.

Jiang, X., Dey, D.K., Prunier, R., Wilson, A.M. & Holsinger, K.E. (2014) A new class of flexible link functions with application to species co-occurrence in Cape floristic region. *The Annals of Applied Statistics*, **7**, 1837–2457.

Karanth, K.K., Nichols, J.D., Hines, J.E., Karanth, K.U. & Christensen, N.L. (2009) Patterns and determinants of mammal species occurrence in India. *Journal of Applied Ecology*, **46**, 1189–1200.

Keil, P., Belmaker, J., Wilson, A.M., Unitt, P. & Jetz, W. (2013) Downscaling of species distribution models: a hierarchical approach. *Methods in Ecology and Evolution*, **4**, 82–94.

Kéry, M., Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2013) Analysing and mapping species range dynamics using occupancy models. *Journal of Biogeography*, **40**, 1463–1474.

La Sorte, F.A. & Hawkins, B.A. (2007) Range maps and species richness patterns: errors of commission and estimates of uncertainty. *Ecography*, **30**, 649–662.

Lahti, T. & Lampinen, R. (1999) From dot maps to bitmaps: Atlas Florae Europae goes digital. *Acta Botanica Fennica*, **162**, 5–9.

Latimer, A.M., Wu, S., Gelfand, A.E. & Silander, J.A. (2006) Building statistical models to analyze species distributions. *Ecological Applications*, **16**, 33–50.

Lele, S.R. & Dennis, B. (2009) Bayesian methods for hierarchical models: are ecologists making a Faustian bargain? *Ecological Applications*, **19**, 581–584.

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.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.

Liu, C., White, M. & Newell, G. (2011) Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography*, **34**, 232–243.

Lobo, J.M., Jimenez-Valverde, A. & Hortal, J. (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, **33**, 103–114.

Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009) The BUGS project: evolution, critique, and future directions. *Statistics in Medicine*, **28**, 3049–3067.

Luoto, M., Marmion, M. & Hjort, J. (2010) Assessing spatial uncertainty in predictive geomorphological mapping: a multi-modelling approach. *Computers & Geosciences*, **36**, 355–361.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

MacKenzie, D.I., Nichols, D.J., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy estimation and modeling*. Academic Press, Burlington.

McCarthy, M.A. (2007) *Bayesian methods for ecology*. Cambridge University Press, Cambridge.

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

McPherson, J.M. & Jetz, W. (2007) Type and spatial structure of distribution data and the perceived determinants of geographical gradients in ecology: the species richness of African birds. *Global Ecology and Biogeography*, **16**, 657–667.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2006) Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions: possibilities and limitations. *Ecological Modelling*, **192**, 499–522.

Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L. & Weir, L.A. (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422–1428.

Munson, M.A., Webb, K., Sheldon, D., Fink, D., Hoachaka, W.M., Iliff, M.J., Riedewald, M., Sorokina, D., Sullivan, B.L., Wood, C.L. & Kelling, S. (2011) *The eBird reference dataset*, Version 3.0, Cornell Lab of Ornithology and National Audubon Society, Ithaca.

Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691–692.

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Muñoz, A.R. & Real, R. (2011) Finessing atlas data for species distribution models. *Diversity and Distributions*, **17**, 1173–1185.

Pagel, J. & Schurr, F.M. (2012) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, **21**, 293–304.

Peterson, A.T., Pearson, R.G., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological niches and geographic distributions*. Princeton University Press, Princeton.

Phillips, S.J. (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography*, **31**, 272–278.

Phillips, S.J., Dudik, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Ridgley, R.S., Allnutt, T.F., Brooks, T., McNicol, D.K., Mehlman, D.W., Young, B.E. & Zook, J.R. (2007) *Digital distribution maps of the birds of the western hemisphere 3.0*, NatureServe, Arlington.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, **35**, 211–226.

Rondinini, C., Marco, M.D., Chiozza, F., Santulli, G., Baisero, D., Visconti, P., Hoffmann, M., Schipper, J., Stuart, S.N., Tognelli, M.F., Amori, G., Falcucci, A., Maiorano, L. & Boitani, L. (2011) Global habitat suitability models of terrestrial mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2633–2641.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical modelling and inference in ecology*. Academic Press, Burlington.

Royle, J.A., Link, W.A. & Sauer, J.R. (2002) Statistical mapping of count survey data. *Predicting species occurrences* (ed. by J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 625–628. Island Press, Washington.

Šizling, A.L. & Storch, D. (2004) Power-law species-area relationships and self-similar species distributions within finite areas. *Ecology Letters*, **7**, 60–68.

Storch, D., Šizling, A.L. & Gaston, K.J. (2003) Geometry of the species-area relationship in central European birds: testing the mechanism. *Journal of Animal Ecology*, **72**, 509–519.

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. & Kelling, S. (2009) eBird: a citizen-based bird observation network in the biological sciences. *Biological Conservation*, **142**, 2282–2292.

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.

Van Dongen, S. (2006) Prior specification in Bayesian statistics: three cautionary tales. *Journal of Theoretical Biology*, **242**, 90–100.

Wang, X. & Dey, D.K. (2010) Generalized extreme value regression for binary response data: an application to B2B electronic payments system adoption. *The Annals of Applied Statistics*, **4**, 2000–2023.

Webster, R.A., Pollock, K.H. & Simons, T.R. (2008) Bayesian spatial modelling of data from avian point count surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, **13**, 121–139.

Wiggins, D. (2004) *American three-toed woodpecker (*Picoides dorsalis*): a technical conservation assessment*. USDA Forest Service, Rocky Mountain Region.

Wilson, A.M. & Silander, J.A. (2014) Estimating uncertainty in daily weather interpolations: a Bayesian framework for developing climate surfaces. *International Journal of Climatology*. doi: 10.1002/joc.3859. in press.

Winkler, R.L. (1967) The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, **62**, 776–800.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. & Veran, S. (2012) Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, **4**, 236–243.

Zarnetske, P.L. (2006) *Predicting habitat suitability with extant presence points and ecologically-based pseudo-absence points for a lesser-known species: the American three-toed woodpecker (*Picoides dorsalis*)*. Thesis, Utah State University, Logan.

Zipkin, E.F., Campbell Grant, E.H. & Fagan, W.F. (2012) Evaluating the predictive abilities of community occupancy models using AUC while accounting for imperfect detection. *Ecological Applications*, **22**, 1962–1972.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 BUGS and R codes and data that were used in the four models presented in this study.

Appendix S2 Simple model for spatial scaling of false absences and supplementary figures presenting details of predictions of the Fine-grain model, Full 2-scale model, Downscaling models 2 and 3, all presented in the same way as in Fig. 6.

Figure S1 Simulation experiment exploring the spatial scaling of false negatives.

Figure S2 Predictions (at the 20 km × 20 km grain) of the Full 2-scale model that incorporated distributional data from two spatial scales and also the spatial random effects (but no informative priors on habitat preferences).

Figure S3 Predictions (at the 20 km × 20 km grain) of the Fine-grain model that incorporated no informative priors on habitat preferences, no spatial random effects and the distributional data only at the fine scale.

Figure S4 Predictions (at the 20 km × 20 km grain) of the Downscaling model 2 that incorporated no informative priors on habitat preferences but used spatial random effects.

Figure S5 Predictions (at the 20 km × 20 km grain) of the Downscaling model 3 that incorporated no informative priors on habitat preferences and no spatial random effects.

BIOSKETCHES

Petr Keil is interested in laying squares of different sizes on maps; he likes to learn what is spatial scale and how various ecological phenomena appear to us when different scales are used to measure them.

Adam M. Wilson is interested in the ecological implications of global environmental change. He uses remotely sensed observations, field data and other existing datasets to investigate shifting species distributions, climatic controls on ecosystem resilience, and changes in the timing of ecological events.

Walter Jetz is interested in the broad-scale ecology, biogeography and conservation of terrestrial vertebrates and plants in a changing world.

Editor: Lluís Brotons