Graham, Laura J. and Haines-Young, Roy H. and Field, Richard (2015) Using citizen science data for conservation planning: methods for quality control and downscaling for use in stochastic patch occupancy modelling. Biological Conservation, 192 . pp. 65-73. ISSN 0006-3207

# Using citizen science data for conservation planning: methods for quality control and downscaling for use in stochastic patch occupancy modelling

Laura J. Graham[a,b,*], Roy H. Haines-Young[a], Richard Field[a]

[a] *School of Geography, University of Nottingham, University Park, Nottingham, UK*
[b] *Department of Ecology and Evoluation, Stony Brook University, Stony Brook, NY, USA*

---

## Abstract

The Incidence Function Model (IFM) has been put forward as a tool for assessing conservation plans. A key benefit of the IFM is low data requirements: widely available species occurrence data and information about land cover. Citizen science is a promising source of such data; however, to use these data in the IFM there are typically two problems. First, the spatial resolution is too coarse, but existing approaches to downscaling species data tend not to extend to patch level (as required by the IFM). Second, widely available citizen science data typically report species' presences only. We devise ten different downscaling methods based on theoretical ecological relationships (the species–area relationship and the distance decay of similarity), and test them against each other. The better performing downscaling methods were based on patch area, rather than distance from other occupied patches. These methods allow data at a coarse resolution to be used in the IFM for comparing conservation management and development plans. Further field testing is required to establish the degree to which results of these new methods can be treated as definitive spatially-explicit predictions. To address the issue of false absences, we present a method to estimate the probability that all species have been listed (and thus that a species' absence from the list represents a true absence), using

---

[*] Corresponding author
*Email address:* `laurajaneegraham@gmail.com` (Laura J. Graham)

*Preprint submitted to Biological Conservation* *August 7, 2015*

the species-accumulation curve. This measure of confidence in absence helps both to objectively identify a habitat network for fitting the IFM, and to target areas for further species recording.

*Keywords:* citizen science, data quality, downscaling, spatial bias, urban conservation

---

## 1. Introduction

Stochastic patch occupancy models can be useful tools for incorporating biodiversity conservation into city planning because they allow for spatially explicit analysis of species' persistence under habitat fragmentation (Hanski, 1994; Opdam et al., 2002, 2003; Van Teeffelen et al., 2012). Species occurrence data at large spatial and temporal extents are necessary for both biodiversity planning (Williams et al., 2002) and for fitting stochastic patch occupancy models (Hanski, 1999; Opdam et al., 2003; Etienne et al., 2004). The Incidence Function Model (IFM) has been identified as particularly suitable for practical biodiversity planning (Lindenmayer et al., 1999; Graham et al., in press), in part as a result of its low data requirements: widely available species occurrence data can be used (Hanski, 1999; Etienne et al., 2004). Most studies tend to employ the IFM in a single-species approach, where the patch occupancies have been specifically surveyed for the purpose (e.g. Bulman et al. 2007; MacPherson and Bright 2011; Heard et al. 2013; Dolrenry et al. 2014). For the IFM to be useful for biodiversity assessment within a conservation or planning framework, multiple indicator species need to be studied. However, to collect occupancy data for a suite of species is costly in terms of time and resources and so other strategies are needed. Our contribution is to provide new strategies to address this lack of occupancy data.

Volunteer biological recording, or more broadly citizen science, is a useful source of data for ecological and conservation research over a large spatial extent (Silvertown, 2009; Devictor et al., 2010; Dickinson et al., 2010, 2012; Tulloch et al., 2013; Graham et al., 2014). These kinds of data are also regularly used for biodiversity planning within UK local authorities (Lott et al., 2006). It allows large quantities of occurrence data to be collected at larger spatial and temporal extents than would be feasible through individual field studies. Species-level data are available from local recording schemes, as well as from large repositories, examples of which are Global Biodiversity Information Facility globally (Global Biodiversity Information Facility, 2014) and National Biodiversity Network (NBN) Gateway in the UK (National Biodiversity Network, 2014). There are, however, some problems with volunteer-collected data. There are concerns about the quality of data collected by non-specialists (see Bird et al. 2014 and Isaac et al. 2014 for discussions of these issues and some of the potential solutions). Specific to the IFM, there are two prevalent issues in data available from major citizen science schemes. First, the data are typically available at grid-square level (for example the finest resolution of data available on the NBN Gateway is 100 x 100m, but a greater coverage of data is available at the 2 km resolution), whereas the IFM requires information about patch-level occupancies (Hanski, 1999). Although some patches may cover a 100 x 100m grid cell, in a highly fragmented landscape such as an urban or heavily managed landscape, the patches are likely to be smaller than this. Additionally, even if the sizes match, the grid cell boundaries are unlikely to be coincident with the patch boundaries. Secondly, the data tend to be presented as species lists, which only give information about species' presences. In a study by Moilanen (2002), it was found that false absences can bias parameter estimates in all components of the model; therefore, the higher the confidence in true absence, the better fitting the model will be (but see Kéry et al. 2010). If volunteer-collected data are to be useful for the IFM, or stochastic patch occupancy models more widely, methods are needed for downscaling these data to patch level, and for determining confidence in species' absences. Here we present methods to address both of these issues.

Current approaches to downscaling atlas data for species tend to fall into three categories: expert opinion, empirical models and spatial processes (Araújo et al., 2005; Keil et al., 2013). The expert opinion approach typically involves matching species to suitable land-cover classes. For a wide range of species, however, the species–habitat relationship is not well known, and so this method can only be applied to well-studied species (Araújo et al., 2005). This approach also operates on the assumption that any suitable habitat is occupied by the species, which is ecologically unrealistic because species range filling is generally discontinuous (Rapoport, 1982). The empirical approach uses environmental variables such as climate, land-cover classes and normalised vegetation difference indices to predict species' occurrences (see Araújo et al. 2005 for an example using general additive modelling and Keil et al. 2013 for one using hierarchical Bayesian modelling). These methods are particularly appropriate for broad-scale species mapping, for example national and continental studies (Stockwell and Peterson, 2002). The spatial-processes approach divides coarse grid cells into finer grid cells and uses statistical point-and-cluster processes to randomly select cells at a fine grain. The environmental attributes from these finer grid cells are used as predictors for species' presences and absences. These methods assume that all fine-grain grid cells within a coarse-grain cell of known occupancy contain suitable habitat. To overcome this problem, Niamir et al. (2011) proposed a method which combines expert knowledge and point sampling.

The empirical and spatial-processes approaches to downscaling species atlas data use environmental variables as predictors, drawing from species' distribution modelling. The extent and grain of interest for a city-level biodiversity plan tend to be much smaller than in studies taking a species' distribution modelling approach to downscaling, and the environmental gradients sampled therefore much narrower but with greater habitat heterogeneity. With their very limited variation, environmental factors such as climate are not useful for predicting species' occurrence at smaller extents. Instead, land cover tends to be the most reliable predictor, and thus the empirical and spatial-processes approaches collapse to species–habitat associations at the city level and individual patch characteristics are likely to become important. The method we outline below applies a combination of expert knowledge (through literature review) and spatial factors. The method involves attributing species' presence to a suitable habitat patch based on its spatial characteristics and known ecological patterns (species–area relationships and the distance decay of community similarity).

To return to the second issue with citizen science data — that they tend to report presence only, but the IFM parameters are sensitive to false absences — we show how this can be circumvented. The IFM parameters estimated for a species can be applied to a different patch network (Hanski et al., 1996) or those estimated on a contiguous subset of patches can be applied to the wider landscape (Bulman et al., 2007). If a core area can be identified within the landscape, with a high confidence in the species' absences, parameters can be estimated using the data from this subset. Species-accumulation curves are widely used to estimate species richness in sampled areas (e.g. Soberón M. and Llorente B. 1993; Colwell and Coddington 1994). This method has also been adapted to give a measure of how well an area has been surveyed (Hortal et al., 2004). Here, we used species accumulation curves to estimate confidence in true absence, and therefore identify subsets of the landscape for use in parameter estimation.

We aim to investigate the extent to which citizen science data are useful as inputs to the IFM. Firstly, we identified well-sampled grid cells within the landscape which can be used to parameterise the IFM. Secondly, we tested several downscaling methods based on spatial characteristics of the landscape and known ecological patterns to convert the species data to an appropriate resolution for the IFM. To achieve our aim, we use the study area of the city of Nottingham, UK and apply the methods to indicator species from the bird, herptile and mammal groups.

## 2. Methods

### 2.1. Study area

The Nottingham City unitary authority was used as a case-study area, with a 2km buffer around its boundary to allow for some effect of dispersal from outside. Nottingham is located in the East Midlands, UK and represents a typical large-to-medium sized urban area in the UK. The unitary authority boundary was chosen as this is the level at which planning decisions are generally made. The location of the study site and a breakdown of the Land Cover Map 2007 classes (Morton et al., 2011) is given in Appendix A (Figure A1, Table A1) with details for Nottingham, four nearby cities and the aggregate of ten similar-sized UK cities for comparison. This indicates that Nottingham is broadly representative of similarly sized UK cities.

### 2.2. Citizen science species data

Data for bird species were provided by Nottinghamshire Birdwatchers. These data comprised 12,110 records of 24 species in 44 2 km grid cells for the years 1998–2011. Bat species data were provided by Nottinghamshire Bat Group and further records were downloaded from NBN Gateway. The combined bat datasets, once duplicates had been removed, contained 421 records for 10 species in 109 1 km grid cells from 1983–2013. Amphibian and reptile data were downloaded from NBN gateway. There were a total of 1116 records for 11 species in 56 2 km grid cells for the period 1984–2012. All downloads from the NBN Gateway were performed using the R package 'rnbn' (Ball and August, 2013). The full list of data providers is supplied in Supplementary Materials, Appendix A (Table A3).

### 2.3. Species–habitat associations and dispersal

It is common practice to use indicator species in biodiversity assessments (Caro and O'Doherty, 2013) because constraints on time, funding and taxonomic knowledge make collection of data on all species unfeasible (Blair, 1999;

Margules et al., 2002). We selected indicator species for modelling with the IFM where sufficient data and information about habitat requirements and dispersal were available. We ensured that species with a range of habitat specialisms and dispersal abilities were chosen, to maximise the species' validity as indicators.

The bird species chosen for modelling with the IFM included five generalists (*Turdus merula*, *Prunella modularis*, *Carduelis carduelis*, *Carduelis chloris* and *Muscicapa striata*), three farmland specialists (*Emberiza calandra*, *Passer montanus* and *Emberiza citrinella*) and four woodland specialists (*Sylvia atricapilla*, *Dendrocopos major*, *Garrulus glandarius* and *Poecile palustris*). *E. citrinella* also uses heathland. The amphibian species selected were *Rana temporaria* and *Bufo bufo*. Common names for all species are given in Table 1. The species chosen for modelling were those which were from well-sampled groups and which had a high enough prevelence in the landscape. If a species is too rare in the landscape, there is little information about the effects of area and isolation to be gained; as such the IFM should only be used if more than around 20% of patches are occupied (Hanski, 1994).

We created a lookup between species and LCM 2007 land-cover types for birds based on Wernham et al. (2002) and Holden and Cleeves (2006) and amphibians based on Beebee and Griffiths (2000). Minimum area requirements for most species were taken from Hinsley et al. (1995), a study based on 151 woods in eastern England, with minimum patch size 0.02 ha. Not all species were included in Hinsley et al. (1995), so minimum area requirements for *E. calandra* were taken from Meyer et al. (2007) and *P. palustris* from Broughton et al. (2006). Minimum area requirements for the amphibian species were not found from a review of the literature, so an arbitrary value of 0.02 ha was chosen.

Dispersal distances for birds mainly came from Paradis et al. (1998), who used British Trust for Ornithology (BTO) ringing data to determine mean breeding and natal dispersal distances; We used the natal distances. Dispersal for *E. calandra* came from Wernham et al. (2002) and for *P. palustris* from Broughton et al. (2010). Dispersal distances for amphibians were taken from Gilioli et al.

(2008), a study on amphibian metapopulations, where the figure given was based on expert opinion. The range of dispersal distances is 700 m (*B. bufo*) to 41.2 km (*S. atricapilla*)

Full details of habitat associations, minimum area and dispersal distances are given in Table 1.

### *2.4. Habitat data*

Land Cover Map 2007 data (LCM 2007, Morton et al. 2011) were used to create maps of suitable habitat for each species. We filtered LCM 2007 data by land-cover class for each species (classes shown in Table 1). Amphibian species have the additional requirement of freshwater for breeding. Although *R. temporaria* will spawn in sites as small as large puddles and ditches, *B. bufo* requires the presence of large, permanent water bodies (Beebee and Griffiths, 2000). To account for the presence of such water bodies in the *B. bufo* habitat model, only habitat that fell within a buffer of its dispersal distance (700 m) from freshwater was considered suitable. We dissolved the artificial boundaries created by land ownership, demarcations between habitat types, and paths and small roads ($\leq 3$ m in width). Finally, we removed all habitat patches smaller than the species' minimum area requirement. This process created a map of suitable habitat patches for each species based on its habitat requirements.

### *2.5. Quantifying uncertainty in species' absence*

Measures of uncertainty in raw species distribution data should be mapped and made explicit as part of good practice (Rocchini et al., 2011). Not only can this provide a spatial account of the potential biases in the data, but it can also aid predictions of total species richness (Soberón M. and Llorente B., 1993; Colwell and Coddington, 1994) or identify well sampled areas for use in species distribution modelling (Hortal et al., 2004; Hortal and Lobo, 2005). We applied the following method to attach a confidence of true absences to each grid cell, and used this information to select areas for fitting the IFM. We followed established methods which fit smoothed species-accumulation curves to a Clench function, and identified the slope of this curve at the position of the last record (Soberón M. and Llorente B., 1993; Colwell and Coddington, 1994; Hortal et al., 2004; Hortal and Lobo, 2005). This slope gives a measure of rate of species accumulation with additional sampling effort. We translated this value to a measure of confidence in species absence, as outlined below.

First, we removed any grid cell with either only one species, or one record per species. The number of database records (ignoring reported abundance, because of the heterogeneity of ways in which this is measured) has previously been used successfully as a surrogate for sampling effort (Lobo, 2008). Using this approach, the species-accumulation curve was created by plotting the cumulative number of species against the number of records. This curve was smoothed by randomising the order of record entry 100 times (*sensu* Hortal et al. 2004; Lobo 2008). Each curve was fitted to a Clench function (equation 1) using non-linear least squares. $S_r$ is the number of species added with each new record, $r$ is the number of records and $a$ and $b$ are the parameters of the function.

$$S_r = \frac{ar}{1 + br} \tag{1}$$

The slope at the point of the last record was calculated using the first-order derivative of the Clench function (equation 2).

$$\frac{dS_r}{dr} = \frac{a}{(1 + br)^2} \tag{2}$$

The confidence in true species' absences was defined as the proportion $1 - \frac{dS_r}{dr}$. We used a threshold level of 95% confidence as the criterion for selecting grid cells suitable for model fitting. We calculated slopes for each of the species groups for each grid cell at the appropriate resolution (2 km for birds, and amphibians and reptiles, 1 km for bats).

### *2.6. Downscaling species atlas data*

The data detailed in Section 2.2 are available at 2 km level for birds, and reptiles and amphibians and at 1 km level for bats. To fit the IFM using these data, it is necessary to downscale them to patch level. Here, we outline and demonstrate a method that first identifies a patch-occupancy level within each grid cell, and then uses a suite of methods based on ecological theory to populate that proportion of patches.

The two specific ecological relationships which informed the downscaling methods are the species–area relationship, and the distance decay of similarity. Species richness increases with increasing island or habitat fragment area (Gleason and Jan, 1921). Based on this, and the fact that real assemblages are typically nested (Wright and Reeves, 1992), larger habitat patches are more likely to be occupied than smaller patches. Distance decay of community similarity follows on from the first law of geography, that "near things are more related than distant things" (Tobler, 1970). Distance decay of similarity in ecology is the negative relationship between geographic distance and community similarity (Nekola and White, 1999). Thus, it should hold that patches that are closer to occupied patches are themselves likely to be occupied.

We calculated the required patch characteristics using the R packages ’rgeos’ (Bivand and Rundel, 2013) and ’rgdal’ (Bivand et al., 2015). Distance was calculated as the minimum edge-to-edge distance between each patch and the nearest patch within a different occupied grid cell (using function ’gDistance’ from ’rgeos’). The proportion of the patch falling in each grid cell was also calculated (area of patch within the grid cell divided by total area of the patch). The purpose of the proportion is to ensure that patches which fall in two or more grid cells are weighted accordingly. Patch area was calculated using the function ’gArea’ from ’rgeos’.

For each grid cell, we assigned species to patches by weighted sampling. The number of patches sampled from each grid cell was proportional to the landscape occupancy level. For example if a species is present in 50% of the well-surveyed (i.e. confidence of completeness above 95%; Figure 1) 2 km cells at the landscape level, species would be assigned to 50% of the patches inside the grid cell. The weighting was the proportion of the patch falling in that particular grid cell multiplied by a weighting based on either area, distance or both. We tested ten different methods of weighting for the sampling. These fall into three categories: distance weighting only, area weighting only, and combined distance and area weighting. Four distance measures were calculated, where $d$ is distance: based on inverse distance weighting, $\frac{1}{d}$ and $\log \frac{1}{d}$; for a linear relationship $\max(d) - d$; to account for dispersal $e^{-\alpha d}$. The last measure in the list is taken from the incidence function model (see section 2.7). Two area measures were tested ($A$ represents area): $A$ and $\log A$. $\log A$ was tested because species richness and area are usually linearly related when both are log-transformed (Gleason and Jan, 1921; Preston, 1962), and also this allows for a threshold at which the patch area becomes large enough that the size is no longer important. Given the likelihood that both the area and distance relationships are important, for three of the distance measures ($\log \frac{1}{d}$, $\max(d) - d$, $e^{-\alpha d}$) sampling was further weighted by $\log A$. So that the measures could be used as probabilities, we standardised these to the range $\{0, 1\}$. As a null model, we employed a downscaling method where the only weighting applied was the proportion of the patch falling in that particular grid cell. This means that the probability of a patch being selected depends only on how much of the patch lies in the grid cell and not on its size or isolation.

### *2.7. Incidence Function Model*

The incidence function model (IFM), a stochastic patch occupancy model developed by Hanski (1994), allows long-term predictions of metapopulation persistence in a network of habitat patches to be made through estimation of colonisation and extinction rates. We followed methods outlined by Oksanen (2004) which are based on Hanski (1994). A full description of the IFM is given in the Supplementary Material, Appendix B.

The IFM has parameters $u$, $x$ and $y$, which are estimated from the data. For each species, we fitted the presence–absence data resulting from each downscaling method for an eight-year period to a logistic regression model (see Supplementary Material, Appendix B). The following two years of data were combined and kept to evaluate the performance of each downscaling method.

We ran 200 replicates of the IFM simulations for 300 time-steps for each downscaling method. For each replicate, we downscaled the species data as described above and generated a new set of parameters for each method. The realistic current occupancy configuration was identified as that at the time step after a suitable burn-in period. This was chosen as a point where species that reach a stable value had done so for most downscaling methods.

### *2.8. Comparing downscaling methods*

Data for validation do not exist at patch scale, so we scaled the results from the IFM for this time step back up to 2 km grid cells; for each replicate, if a patch contained within a grid cell was occupied, the grid cell for that generation and replicate was considered occupied. The probability of occurrence was created by taking the mean of the replicates. Data were available for the two actual years following those used for fitting the model, so we used these records for model evaluation. The up-scaled results for each species were filtered to create a presence dataset containing only grid cells with occurrences recorded in one or both of the two validation years, and an absence dataset containing only grid cells where no occurrence was recorded in either year. We used a Kruskal-Wallis test to determine a significant difference between downscaling methods, and the post-hoc multiple pairwise comparison test outlined by Siegel and Castellan (1988) using the R package 'pgirmess' (Giraudoux, 2014) to determine where these differences lay.

Further validation of the model output using an independent source of data was done for bird species using the 2 km grid data from the BTO Atlas of Breeding Birds (Gibbons et al., 1993). It should be noted that not all 2 km grid cells were visited and that the data do not reflect a complete species list, but instead the species seen in a 1–2 hour visit. We created a subset of the upscaled model output which included only 2 km cells visited for the BTO data set and then performed the validation in the same way as described above.

As an additional reality check of the model, we calculated the median proportion of occupied habitat for each downscaling method at 175 time-steps (after the burn-in period) and compared against the national occupancy level (% of 10 x 10 km grid cells occupied). Specifically, we identified whether nationally abundant species appear locally abundant after downscaling and modelling, and similarly for nationally rare species. We quantified the relationship between the simulated occupancy proportions and national occupancy proportions by fitting a linear regression model with national proportions as the dependent variable and simulated proportions as the independent variable. Although local proportions of species' occupancy do not necessarily reflect national ones, this test is a useful secondary check used in conjunction with the above validation.

### *2.9. Sensitivity analysis of the occupancy assumption*

Our downscaling methods assume that the proportion of patches occupied by a species within each grid cell is the same as the proportion of grid cells occupied by the species within the full landscape. This is reasonable to the extent that species often display self-similar (fractal) distributions across adjacent scales (Ritchie and Olff 1999; Kunin 1998; but see Halley et al. 2004). However, because the relationship between scale and occupancy is not always linear (eg. Barwell et al., 2014), we performed a sensitivity analysis to test how important this assumption was when fitting the IFM by varying the starting occupancy proportions. The purpose of the sensitivity analysis was to examine how much impact a change of 10 percentage points either side of the landscape % occupancy would have on the results. For the area-weighted downscaling method, we generated 3 sets of 200 starting occupancies. The first set, Landscape (LS) occupancy, kept to the assumption that species' distributions are fractal across adjacent scales; for the high occupancy set we increased the percentage of suitable habitat occupied by 10 percentage points (eg. if LS occupancy was 50%, the high occupancy would be 60%); similarly, for the low occupancy set we decreased the percentage of suitable habitat occupied by 10 percentage points. The IFM was parameterised using each of these sets of starting occupancies and the species' patch occupancies simulated over 500 generations.

One of the useful outputs of the IFM is that quantified measures of metapopulation viability can be calculated, such as survival probability, minimum occupancy and current occupancy (after the burn-in period). We analysed the sensitivity of the IFM to the occupancy proportion of the starting condition by testing the sensitivity in these resulting measures. We used post-hoc multiple pairwise comparison test to determine whether the differences in the measures were significant by comparing the measures obtained from the LS starting occupancy to those from both the high and low occupancy starting conditions, and the measures resulting from the high and low starting occupancies to each other.

## 3. Results

### *3.1. Mapping uncertainty in presence data*

After fitting these curves to a Clench function and calculating the slope at the point of the last record, we identified 36% of cells were well-surveyed birds, 2% for bats and 52% for amphibians and reptiles (Figure 1).

To fit the IFM to a subset, that subset must be contiguous (Moilanen 2002 found that any missing patches in the network affected parameter estimates), and also must contain enough patches to fit the logistic regression model. We therefore selected the largest contiguous block of well-surveyed grid cells for fitting the IFM for each species group. These are shown in Figure 1 c) and i) as a bold outline. From our findings, bats are not well surveyed enough to use this dataset in the IFM.

### *3.2. Comparison of downscaling methods*

We fitted the IFM to eight years of species data. For birds, these were the first eight years of the dataset: 1998–2005. Although amphibians and reptiles were well sampled based on records from all years in the dataset, there were few records for the time period 1998–2005. We therefore fit the IFM to amphibian data from 1988–1995. The bird species data for the combined years

2006–2007, and the amphibian species data for 1996–1997 were used for model evaluation.

Based on the 200 simulations of the IFM for each species and downscaling method, we set the burn-in period to be 175 time steps (t). Plots showing the model trajectories are given in Supplementary Materials, Appendix C. These plots give results for both proportion of habitat patches occupied and proportion of suitable habitat area occupied; because both were qualitatively similar all following results show the latter which we considered to be the measure which had most practical value. The model predicted full occupancy within the first few generations for the longer dispersing species (*Carduelis carduelis*, *Muscicapa striata*, *Sylvia atricapilla* and *Dendrocopos major*). These were removed from further analysis because it is likely these species do not display metapopulation dynamics at the scale of study due to a lack of dispersal limitation.

Figure 2 a) shows, for all remaining species, the predicted probability of occupancy for each of the downscaling methods for those grid cells in which the species have been recorded in either of the two years of evaluation data. A similar analysis broken down by species is shown in the Supplementary Materials, Appendix D (Figures D1 and D2). Validation using the independent (BTO) data gave similar results (see Appendix E in the supplementary materials).

An overall significant difference between model performance based on the downscaling methods was detected using a Kruskal-Wallis test (Chi-square = 143.52, df = 9, $P < 0.001$). These results show that the area and, to some extent, log(area) weighted downscaling methods are most reliable in predicting species' presence overall.

Figure 2 b) shows the predicted probabilities of occurrence for grid cells where the species has not been recorded as present. It should be noted that these should not be considered as confirmed absences due to the ad hoc nature of citizen science biological recording (see Discussion). The analysis is shown by species in Supplementary Materials, Appendix D (Figures D3 and D4). A significant difference between model predictions based on starting conditions from each of the downscaling methods was detected here also (Chi-square = 53.4, df = 9, $p < 0.001$). Here, however, it is less clear which downscaling method is best at predicting the presumed absences, but the better methods for predicting presences are the worst for predicting absences. This suggests that when parameterising the model based on data downscaled by the area or log(area) weighted method, the IFM over predicts species' occupancy.

No downscaling method matched the national proportions closely (Figure 3), though both $A$ and $\log A$ are plausible at the finer scale in the study area. The area-weighted downscaling method best matched national occupancies, although the $R^2$ value was still low ($R^2 = 0.47$, $F(1, 8) = 6.99$, $P = 0.03$, residuals normally distributed). The regression model suggested a close to 1:1 relationship with national proportions (national occupancy proportion = 0.13 + 0.97 x predicted occupancy proportion, intercept not significantly different from 0 and slope not significantly different from 1). *P. palustris* was often below national levels for each method, whereas *G. glandarius* and *E. calandra* were frequently above (although not for the area-weighted downscaling). *P. palustris's* habitat also includes parks and farmland with woods and coppices (Holden and Cleeves, 2006), but it is not possible to identify these from the LCM 2007 data and so only broadleaved woodland was included.

### *3.3. Sensitivity analysis of the occupancy starting condition*

We performed a sensitivity analysis to evaluate whether changes in the initial starting occupancy had an impact on the measures of landscape persistence derived from the IFM results. A post-hoc multiple comparison test was used to determine whether the differences in the measure between occupancy start conditions were significant. Many were significant, but overall a change in 10% of occupancy starting condition resulted in a change of less than 10% in the resulting metapopulation measure (see Table 2). Minimum occupancy % seemed to be the measure most sensitive to starting condition, with the largest differences between the value based on high vs low occupancy being 18.97% for *T. merula* and 22.27% for *R. temporaria*.

## 4. Discussion

For the IFM to have wide use in conservation management and planning, it is important that accessible sources of data are available. Our results show that citizen science data offer the potential to meet the data requirements of the IFM, but with some important caveats. Citizen science and collections data can be beneficial for studies at large spatial and temporal extents but are often sparse and geographically biased (Funk and Richardson, 2002; Boakes et al., 2010) or at a coarser spatial resolution than required for planning purposes (Araújo et al., 2005; Keil et al., 2013). The data available for the study area may not be a complete reflection of species' occupancies (Figure 1). Despite this, there are sufficient records for both bird and amphibian species to fit the IFM.

We used a method to quantify uncertainty in species distribution data, which gives a measure of confidence in true absence, a further issue with presence-only species lists. This method is useful here to select a core subset of the landscape for fitting the IFM with minimal subjectivity. This method also shows spatially explicit biases in the data, which can be incorporated into any statistical modelling of the data. These kinds of biases are prevalent in citizen science data, and methods to identify them are necessary (Bird et al., 2014). The maps showing levels of uncertainty in grid cells can be useful for volunteer recorders to help identify where further recording effort is necessary and also for conservation planners to evaluate areas to target future research effort.

The downscaling method we present herein deals with the issue of the data being available at a coarser resolution than required. Of the downscaling methods tested, we found that the methods which involved weighting by $A$ and $\log A$ were both the closest match to the known presences in the evaluation data set and the national occupancy proportions. Based on the above, weighting by either $A$ or $\log A$ would be the most appropriate method for downscaling species' distribution data for use in the IFM. These results suggest that the species–area relationship has more influence than the distance decay of similarity at this spatial scale. The shape of the distance decay relationship depends on the sampling extent and grain: when sampled at a large extent and grain, the relationship is usually exponential; whereas when sampled at a small extent and grain with limited environmental gradient, the distance decay generally forms a power-law relationship (Nekola and McGill, 2014). Of the remaining methods, the closest to national occupancies was the method weighted by $\log \frac{1}{d} \log A$, which fits with an exponential relationship for distance decay. It is possible that for our study area, the slope of the power-law relationship is very shallow due to the limited environmental gradient, hence why those methods which weight by patch area alone fit national occupancy levels more closely.

It should be noted that when comparing against the grid cells where the species had not been recorded as present, the downscaling methods which weighted by $A$ and $\log A$ performed relatively poorly, if we presume these were actually absences. Well-designed recording would be necessary to understand whether downscaling using these weightings leads to over prediction of species occupancy when simulating using the IFM, or whether these were in fact unrecorded presences. It is possible that there are many false absences in the dataset. For example *T. merula* was recorded as present in 16 of the 61 2 km grid cells, however in the 2 km grid data from the fieldwork for the BTO Atlas of Breeding Birds (Gibbons et al., 1993), *T. merula* is recorded in 98% of the grids visited. Similarly, *C. chloris* was recorded in 18 of the 61 2 km grid cells, but was present in 89% of the 2 km cells visited for the BTO Atlas. It is possible that this is a result of bias in recorder motivations away from recording very common species (Isaac and Pocock, 2015). Validation using the relatively few BTO cells in our study area gave very similar results (see above and Appendix E).

Refining species–habitat associations may improve the performance of the downscaling methods. It is possible that datasets whose land-cover classes are broad and non-specific (e.g. LCM 2007, as used here) are not entirely sufficient to identify suitable habitat patches and need to be combined with other data that offer further description of habitat types (e.g. Ordnance Survey Master Map for this study area). Species also depend on structure as well as type of habitat, for example Broughton et al. (2006) found an effect of canopy height on marsh tit occupancy. Incorporating remotely sensed data will allow for habitat structure, and perhaps quality, to be considered. We have used LCM 2007 because of its availability (both to us, and to planners who may wish to use this method), however remote sensing data are increasing in quantity and quality, and pushes to make these data open access will mean habitat data are more accessible and accurate (Sutherland et al., 2014; Turner et al., 2015). For the marsh tit, which appears to be consistently under-predicted by all methods, it is possible that improving the habitat association will improve prediction. It is possible that there are no issues with the habitat association for the corn bunting and they are more prevalent in Nottingham than nationally. Nottingham has quite a high proportion of arable land compared to some other cities (see Table A1).

Long-dispersing species were consistently over-predicted and removed from the final analysis. It may be that these do not exhibit metapopulation dynamics at the scale we are studying. The metapopulation approach is only applicable if a species is sufficiently dispersal-limited (Hanski, 1994). Equally, we have not tested our method for very rare species because this again goes against the assumptions of the IFM.

We assumed a fractal distribution of the species when setting the initial proportion of occupied patches to which the downscaling methods allocated presences. The results of the sensitivity analysis were mixed, suggesting that the impact of this assumption could be greater for some species — particularly when using the minimum occupancy measure of metapopulation persistence. Our results suggest that in most cases, assuming a fractal distribution of species occupancy is reasonable. However, refining this part of the procedure could potentially improve the model predictions. Occupancy–area curves are linear when the species distribution is fractal (Kunin, 1998). The relationship between sampling scale and occupancy has been found to be non-linear in some plant (Kunin, 1998) and dragonfly species (Barwell et al., 2014). Incorporating tested estimations of the occupancy–area curve from these methods may improve the predictions from the downscaling methods outlined above.

The predicted patch occupancies from the methods presented above would require rigorous testing with field data before they can be taken as definitive spatially explicit predictions. For the purposes of fitting the IFM for use in planning and conservation management scenario comparison, however, appropriately chosen downscaling methods should suffice. This is because the inherent uncertainty in metapopulation models means estimates derived from models should be viewed as relative comparisons rather than absolute predictions (Grimm et al., 2004).

Downscaling species atlas data using a combination of habitat associations and patch characteristics has the potential to fill a gap in existing downscaling methods. The method proposed and tested herein is useful for study areas that are too large for full surveys, but small enough that the environmental gradients are limited or non-existent, making traditional species distribution modelling approaches (e.g. Araújo et al. 2005; Keil et al. 2013) inappropriate.

## Acknowledgements

We thank Joaquín Hortal for very helpful discussion of the methods for maps of uncertainty. Thanks also to Rob Johnson and Peter Acton at Nottinghamshire Biological and Geological Records Centre, Michael Walker at Nottinghamshire Wildlife Trust and Andy Hall from Not- tinghamshire Birdwatchers for supplying species data, and giving advice on the use of these data. Finally, we thank Adam Algar and the BEER (Biodiversity, Ecology and Evolution Research) Group at the University of Nottingham for helpful comments on an earlier version of this manuscript. This research was funded by the ESRC.

## References

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. doi:`10.1111/j.1466-822X.2004.00128.x`.

Ball, S., August, T., 2013. rnbn: Access NBN data. URL: `https://github.com/JNCC-UK/rnbn`.

Barwell, L.J., Azaele, S., Kunin, W.E., Isaac, N.J.B., 2014. Can coarse-grain patterns in insect atlas data predict local occupancy? Diversity and Distributions 20, 895–907. doi:`10.1111/ddi.12203`.

Beebee, T.J.C., Griffiths, R.a., 2000. Amphibians and reptiles: a natural history of the British herpetofauna. Harper Collins, London.

Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., Pecl, G.T., Barrett, N., Frusher, S., 2014. Statistical solutions for error and bias in global citizen science datasets. Biological Conservation 173, 144–154. doi:`10.1016/j.biocon.2013.07.037`.

Bivand, R., Keitt, T., Rowlingson, B., 2015. The rgdal Package: Bindings for the geospatial data abstraction library. URL: `http://cran.r-project.org/package=rgdal`.

Bivand, R., Rundel, C., 2013. rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-2. URL: `http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Interface+to+Geometry+Engine+-+Open+Source+(GEOS)#0`.

Blair, R.B., 1999. Birds and butterflies along an urban gradient: Surrogate taxa for assessing biodiversity? Ecological Applications 9, 164–170.

Boakes, E.H., McGowan, P.J.K., Fuller, R.a., Chang-Qing, D., Clark, N.E., O'Connor, K., Mace, G.M., 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. PLoS Biology 8, e1000385. doi:`10.1371/journal.pbio.1000385`.

Broughton, R.K., Hill, R.A., Bellamy, P.E., Hinsley, S.A., 2010. Dispersal, ranging and settling behaviour of marsh tits *Poecile palustris* in a fragmented landscape in lowland England. Bird Study 57, 458–472. doi:`10.1080/00063657.2010.489316`.

Broughton, R.K., Hinsley, S.A., Bellamy, P.E., Hill, R.A., Rothery, P., 2006. Marsh tit *Poecile palustris* territories in a British broad-leaved wood. Ibis 148, 744–752. doi:`10.1111/j.1474-919X.2006.00583.x`.

Bulman, C.R., Wilson, R.J., Holt, A.R., Bravo, L.G., Early, R.I., Warren, M.S., Thomas, C.D., 2007. Minimum viable metapopulation size, extinction debt, and the conservation of a declining species. Ecological Applications 17, 1460–1473. doi:`10.1890/06-1032.1`.

Caro, T.M., O'Doherty, G., 2013. On the use of surrogate species in conservation biology. Conservation Biology 13, 805–814.

Colwell, R.K., Coddington, J.a., 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 345, 101–118.

Devictor, V., Whittaker, R.J., Beltrame, C., 2010. Beyond scarcity: Citizen science programmes as useful tools for conservation biogeography. Diversity and Distributions 16, 354–362. doi:`10.1111/j.1472-4642.2009.00615.x`.

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T., Purcell, K., 2012. The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment 10, 291–297. doi:`10.1890/110236`.

Dickinson, J.L., Zuckerberg, B., Bonter, D.N., 2010. Citizen science as an ecological research tool: Challenges and benefits. Annual Review of Ecology, Evolution, and Systematics 41, 149–172. doi:`10.1146/annurev-ecolsys-102209-144636`.

Dolrenry, S., Stenglein, J., Hazzah, L., Lutz, R.S., Frank, L., 2014. A metapopulation approach to African lion (*Panthera leo*) conservation. PLoS ONE 9, e88081. doi:`10.1371/journal.pone.0088081`.

Etienne, R.S., ter Braak, C.J.F., Vos, C.C., 2004. Application of Stochastic Patch Occupancy Models to Real Metapopulations, in: Hanski, I., Gaggiotti, O.E. (Eds.), Ecology, Genetics and Evolution of Metapopulations. Academic Press, Burlington, MA, USA, pp. 105–132. doi:`10.1016/B978-012323448-3/50007-6`.

Funk, V.A., Richardson, K.S., 2002. Systematic data in biodiversity studies: use it or lose it. Systematic biology 51, 303–316. doi:`10.1080/10635150252899789`.

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

Gilioli, G., Bodini, A., Baumgärtner, J., Weidmann, P., Hartmann, J., 2008. A novel approach based on information theory to rank conservation strategies: An application to amphibian metapopulations. Animal Conservation 11, 453–462. doi:`10.1111/j.1469-1795.2008.00190.x`.

Giraudoux, P., 2014. pgirmess: Data analysis in ecology. URL: `http://cran.r-project.org/package=pgirmess`.

Gleason, H.a., Jan, N., 1921. Species and Area. Journal of Ecology 9, 95–99.

Global Biodiversity Information Facility, 2014. GBIF.org. URL: `http://www.gbif.org`.

Graham, L.J., Bailey, J.J., Algar, A.C., Field, R., 2014. Where next for macroecology: citizen macroecology? Frontiers of Biogeography 6, 16–19.

Graham, L.J., Field, R., Haines-Young, R., in press. Strengthening Urban Landscape Planning: a metapopulation modelling framework, in: Francis, R. (Ed.), Urban Landscape Ecology: Science, Policy and Practice. Routledge.

Grimm, V., Lorek, H., Finke, J., Koester, F., Malachinski, M., Sonnenschein, M., Moilanen, A., Storch, I., Singer, A., Wissel, C., Frank, K., 2004. META-X: Generic software for metapopulation viability analysis. Biodiversity and Conservation 13, 165–188. doi:`10.1023/B:BIOC.0000004317.42949.f7`.

Halley, J.M., Hartley, S., Kallimanis, A.S., Kunin, W.E., Lennon, J.J., Sgardelis, S.P., 2004. Uses and abuses of fractal method in ecology. Ecology Letters 7, 254–271. doi:`10.1111/j.1461-0248.2004.00568.x`.

Hanski, I., 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology 63, 151–162. doi:`10.2307/5591`.

Hanski, I., 1999. Metapopulation ecology. Oxford University Press, Oxford. doi:`10.1002/9780470015902.a0021905`.

Hanski, I., Moilanen, a., Pakkala, T., Kuussaari, M., 1996. The quantitative incidence function model and persistence of an endangered butterfly metapopulation. Conservation Biology 10, 578–590. doi:`10.1046/j.1523-1739.1996.10020578.x`.

Heard, G.W., McCarthy, M.A., Scroggie, M.P., Baumgartner, J.B., Parris, K.M., 2013. A Bayesian model of metapopulation viability, with application to an endangered amphibian. Diversity and Distributions 19, 555–566. doi:`10.1111/ddi.12052`.

Hinsley, S.A., Bellamy, P.E., Newton, I., Sparks, T.H., 1995. Habitat and landscape factors influencing the presence of individual breeding bird species in woodland fragments. Journal of Avian Biology 26, 94–104. doi:`10.2307/3677057`.

Holden, P., Cleeves, T., 2006. RSPB Handbook of British Birds. 2nd ed., Christopher Helm, London.

Hortal, J., Garcia-Pereira, P., García-Barros, E., 2004. Butterfly species richness in mainland Portugal: Predictive models of geographic distribution patterns. Ecography 27, 68–82. doi:`10.1111/j.0906-7590.2004.03635.x`.

Hortal, J., Lobo, J.M., 2005. An ED-based protocol for optimal sampling of biodiversity. Biodiversity and Conservation 14, 2913–2947. doi:`10.1007/s10531-004-0224-z`.

Isaac, N.J.B., Pocock, M.J.O., 2015. Bias and information in biological records. Biological Journal of the Linnean Society 115, 522–531. doi:`10.1111/bij.12517`.

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P., Roy, D.B., 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. Methods in Ecology and Evolution , 1052–1060URL: `http://doi.wiley.com/10.1111/2041-210X.12254`, doi:`10.1111/2041-210X.12254`.

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. doi:`10.1111/j.2041-210x.2012.00264.x`.

Kéry, M., Royle, J.A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., Zbinden, N., 2010. Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. Conservation Biology 24, 1388–1397. doi:`10.1111/j.1523-1739.2010.01479.x`.

Kunin, W.E., 1998. Extrapolating species abundance across spatial scales. Science 281, 1513–1515. doi:`10.1126/science.281.5382.1513`.

Lindenmayer, D.B., McCarthy, M.A., Pope, M.L., 1999. Arboreal marsupial incidence in eucalypt patches in south- eastern Australia: a test of Hanski's incidence function metapopulation model for patch occupancy. Oikos 84, 99–109. doi:`10.2307/3546870`.

Lobo, J.M., 2008. Database records as a surrogate for sampling effort provide higher species richness estimations. Biodiversity and Conservation 17, 873–881. doi:`10.1007/s10531-008-9333-4`.

Lott, D., Moon, S., Cooper, A., 2006. Biodiversity needs for Local Authorities and National Park Authorities. Technical Report. Association of Local Government Ecologists. URL: `http://www.alge.org.uk/publications/`.

MacPherson, J.L., Bright, P.W., 2011. Metapopulation dynamics and a landscape approach to conservation of lowland water voles (*Arvicola amphibius*). Landscape Ecology 26, 1395–1404. doi:`10.1007/s10980-011-9669-0`.

Margules, C.R., Pressey, R.L., Williams, P.H., 2002. Representing biodiversity: data and procedures for identifying priority areas for conservation. Journal of Biosciences 27, 309–326. doi:`10.1007/BF02704962`.

Meyer, B.C., Mammen, K., Grabaum, R., 2007. A spatially explicit model for integrating species assessments into landscape planning as exemplified by the Corn Bunting (*Emberiza calandra*). Journal for Nature Conservation 15, 94–108. doi:`10.1016/j.jnc.2007.01.001`.

Moilanen, A., 2002. Implications of empirical data quality to metapopulation model parameter estimation and application. Oikos 96, 516–530. doi:`10.1034/j.1600-0706.2002.960313.x`.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R., Simpson, I., 2011. Final Report for LCM2007 - the new UK land cover map. Countryside Survey Technical Report No 11/07. Technical Report. doi:`doi:10.5285/1d78e01a-a9c1-4371-8482-1c1b57d9661f`.

National Biodiversity Network, 2014. NBN Gateway. URL: `http://data.nbn.org.uk`.

Nekola, J.C., McGill, B.J., 2014. Scale dependency in the functional form of the distance decay relationship. Ecography 37, 309–320. doi:`10.1111/j.1600-0587.2013.00407.x`.

Nekola, J.C., White, P.S., 1999. The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26, 867–878. doi:`10.1046/j.1365-2699.1999.00305.x`.

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. doi:`10.1111/j.1472-4642.2011.00793.x`.

Oksanen, J., 2004. Incidence Function Model in R. URL: `http://cc.oulu.fi/$\sim$jarioksa/opetus/openmeta/metafit.pdf`.

Opdam, P., Foppen, R.P.B., Vos, C.C., 2002. Bridging the gap between ecology and spatial planning in landscape ecology. Landscape Ecology 16, 767–779.

Opdam, P., Verboom, J., Pouwels, R., 2003. Landscape cohesion: An index for the conservation potential of landscapes for biodiversity. Landscape Ecology 18, 113–126. doi:`10.1023/A:1024429715253`.

Paradis, E., Baillie, S.R., Sutherland, W.J., Gregory, R.D., 1998. Patterns of natal and breeding dispersal in birds. Journal of Animal Ecology 67, 518–536. doi:`10.1046/j.1365-2656.1998.00215.x`.

Preston, F.W., 1962. The Canonical Distribution of Commonness and Rarity: Part I. Ecology 43, 185–215. URL: `http://www.esajournals.org/doi/abs/10.2307/1931976`, doi:`10.2307/1931976`.

Rapoport, E., 1982. Areogeography: Geographical Strategies of Species. Pergamon Press, Oxford.

Ritchie, M.E., Olff, H., 1999. Spatial scaling laws yield a synthetic theory of biodiversity. Nature 400, 557–560. doi:`10.1038/23010`.

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. doi:`10.1177/0309133311399491`.

Siegel, S., Castellan, N.J., 1988. Nonparametric Statistics for the Behavioral Sciences. 2nd ed., McGraw-Hill, New York.

Silvertown, J., 2009. A new dawn for citizen science. Trends in ecology & evolution 24, 467–471. doi:`10.1016/j.tree.2009.03.017`.

Soberón M., J., Llorente B., J., 1993. The use of species accumulation functions for the prediction of species richness. Conservation Biology 7, 480–488. doi:`10.1046/j.1523-1739.1993.07030480.x`.

Stockwell, D.R.B., Peterson, a.T., 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148, 1–13. doi:`10.1016/S0304-3800(01)00388-X`.

Sutherland, W.J., Aveling, R., Brooks, T.M., Clout, M., Dicks, L.V., Fellman, L., Fleishman, E., Gibbons, D.W., Keim, B., Lickorish, F., Monk, K.A., Mortimer, D., Peck, L.S., Pretty, J., Rockström, J., Rodríguez, J.P., Smith, R.K., Spalding, M.D., Tonneijck, F.H., Watkinson, A.R., 2014. A horizon scan of global conservation issues for 2014. Trends in Ecology and Evolution 29, 15–22. doi:`10.1016/j.tree.2013.11.004`.

Tobler, W.R., 1970. A computer movie simulation urban growth in the Detroit region. Economic Geography 46, 234–240. doi:`10.2307/143141`.

Tulloch, A.I., Possingham, H.P., Joseph, L.N., Szabo, J., Martin, T.G., 2013. Realising the full potential of citizen science monitoring programs. Biological Conservation 165, 128–138. doi:`10.1016/j.biocon.2013.05.025`.

Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A.K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koh, L.P., Leimgruber, P., Taubenboeck, H., Wegmann, M., Wikelski, M., Woodcock, C.E., 2015. Free and open-access satellite data are key to biodiversity conservation. Biological Conservation 182, 173–176. doi:`10.1016/j.biocon.2014.11.048`.

Van Teeffelen, A.J.A., Vos, C.C., Opdam, P., 2012. Species in a dynamic world: Consequences of habitat network dynamics on conservation planning. Biological Conservation 153, 239–253. doi:`10.1016/j.biocon.2012.05.001`.

Wernham, C., Siriwardena, G.M., Toms, M., Marchant, J., Clark, J.A., Baillie, S., 2002. The Migration Atlas: Movements of the Birds of Britian and Ireland. T. & A.D. Poyser, London.

Williams, P.H., Margules, C.R., Hilbert, D.W., 2002. Data requirements and data sources for biodiversity priority area selection. Journal of Biosciences 27, 327–338. doi:`10.1007/BF02704963`.

Wright, D.H., Reeves, J.H., 1992. On the meaning and measurement of nestedness of species assemblages. Oecologia 92, 416–428. doi:`10.1007/BF00317469`.

# Tables

Table 1: Mean dispersal distance, broad habitat type (based on LCM 2007), and minimum habitat requirement for each modelled species. LCM classes: 1. Broadleaved Woodland, 2. Coniferous Woodland, 3. Arable and Horticulture, 4. Improved Grassland, 5. Rough Grassland, 6. Neutral Grassland, 8. Acid Grassland, 9. Fen, Marsh and Swamp, 10. Heather, 11. Heather Grassland, 16. Freshwater

| Species | Common name | Dispersal (km) | LCM class | Min. area (ha) |
|---|---|---:|---|---:|
| *Turdus merula* | Blackbird | 3.3 | 1, 2, 3, 4, 5, 6, 7, 8 | 0.02 |
| *Prunella modularis* | Dunnock | 2.1 | 1, 2, 3, 4, 5, 6, 7, 8 | 0.02 |
| *Carduelis carduelis* | Goldfinch | 11.1 | 3, 5 | 0.12 |
| *Carduelis chloris* | Greenfinch | 4.2 | 1, 2, 3 | 0.25 |
| *Muscicapa striata* | Spotted flycatcher | 12.8 | 1, 2 | 0.1 |
| *Emberiza calandra* | Corn bunting | 4 | 3, 4, 5, 6, 8 | 2.5 |
| *Passer montanus* | Tree sparrow | 8 | 1, 2, 3 | 0.12 |
| *Emberiza citrinella* | Yellowhammer | 8.4 | 3, 5, 10, 11 | 0.03 |
| *Sylvia atricapilla* | Blackcap | 41.2 | 1 | 0.2 |
| *Dendrocopos major* | Great spotted woodpecker | 16.5 | 1, 2 | 0.26 |
| *Garrulus glandarius* | Jay | 3.5 | 1, 2 | 0.32 |
| *Poecile palustris* | Marsh tit | 0.885 | 1 | 2.1 |
| *Rana temporaria* | Common frog | 1 | 1, 2, 3, 4, 5, 6, 8, 9, 16 | 0.02 |
| *Bufo bufo* | Common toad | 0.7 | 1, 2, 5, 6, 8, 9, 16 | 0.02 |

Table 2: Sensitivity analysis of the occupancy starting condition. The landscape occupancy column shows the results from a starting occupancy % equal to that of the % of 2km x 2km grid cells in which the species is present in the landscape. The high occupancy column shows the result where an additional 10% of the suitable habitat is assumed to be occupied for the starting condition; low occupancy is the results from starting with 10% less suitable patches occupied than landscape occupancy. For the three end columns, an x represents a significant difference in the measure between starting occupancies based on a post-hoc multiple pairwise comparison test.

| Species | Metapopulation measure | Landscape occupancy | Low occupancy | High occupancy | Low vs LS | High vs LS | High vs Low |
|---|---|---:|---:|---:|---:|---:|---:|
| *Turdus merula* | Minimum occupancy % | 88.27 | 85.22 | 66.25 | x | x | x |
| *Turdus merula* | Surviving replicates % | 100.00 | 100.00 | 100.00 | | | |
| *Turdus merula* | Occupancy % after burn-in | 96.21 | 94.74 | 96.37 | x | x | x |
| *Prunella modularis* | Minimum occupancy % | 64.85 | 63.17 | 65.60 | x | x | x |
| *Prunella modularis* | Surviving replicates % | 100.00 | 100.00 | 100.00 | | | |
| *Prunella modularis* | Occupancy % after burn-in | 81.61 | 76.21 | 85.47 | x | x | x |
| *Carduelis chloris* | Minimum occupancy % | 70.86 | 71.11 | 70.00 | | | |
| *Carduelis chloris* | Surviving replicates % | 100.00 | 100.00 | 100.00 | | | |
| *Carduelis chloris* | Occupancy % after burn-in | 95.36 | 93.25 | 94.82 | x | x | x |
| *Emberiza calandra* | Minimum occupancy % | 40.06 | 41.00 | 31.91 | | x | x |
| *Emberiza calandra* | Surviving replicates % | 100.00 | 100.00 | 100.00 | | | |
| *Emberiza calandra* | Occupancy % after burn-in | 59.24 | 53.44 | 61.99 | x | | x |
| *Passer montanus* | Minimum occupancy % | 50.38 | 49.72 | 34.17 | x | x | x |
| *Passer montanus* | Surviving replicates % | 100.00 | 100.00 | 98.00 | | | |

Continued on next page

Table 2: Continued from previous page

| Species | Metapopulation measure | Landscape occupancy | Low occupancy | High occupancy | Low vs LS | High vs LS | High vs Low |
|---|---|---|---|---|---|---|---|
| *Passer montanus* | Occupancy % after burn-in | 91.26 | 89.77 | 89.39 | x | x | x |
| | | | | | | | |
| *Emberiza citrinella* | Minimum occupancy % | 58.13 | 56.83 | 54.90 | | x | x |
| *Emberiza citrinella* | Surviving replicates % | 100.00 | 99.50 | 99.00 | | | |
| *Emberiza citrinella* | Occupancy % after burn-in | 93.25 | 91.38 | 91.49 | x | | x |
| | | | | | | | |
| *Garrulus glandarius* | Minimum occupancy % | 42.79 | 40.44 | 44.77 | x | x | x |
| *Garrulus glandarius* | Surviving replicates % | 100.00 | 100.00 | 100.00 | | | |
| *Garrulus glandarius* | Occupancy % after burn-in | 89.89 | 89.07 | 91.15 | x | x | x |
| | | | | | | | |
| *Poecile palustris* | Minimum occupancy % | 0.84 | 0.44 | 0.41 | | | |
| *Poecile palustris* | Surviving replicates % | 22.50 | 12.50 | 12.00 | | | |
| *Poecile palustris* | Occupancy % after burn-in | 5.81 | 4.58 | 4.56 | | | |
| | | | | | | | |
| *Rana temporaria* | Minimum occupancy % | 56.00 | 71.53 | 49.26 | x | x | x |
| *Rana temporaria* | Surviving replicates % | 100.00 | 100.00 | 100.00 | | | |
| *Rana temporaria* | Occupancy % after burn-in | 80.16 | 85.09 | 76.02 | x | | x |
| | | | | | | | |
| *Bufo bufo* | Minimum occupancy % | 6.00 | 4.69 | 4.71 | x | x | |
| *Bufo bufo* | Surviving replicates % | 70.00 | 61.50 | 59.00 | | | |
| *Bufo bufo* | Occupancy % after burn-in | 18.29 | 13.45 | 17.96 | x | | x |

# Figures

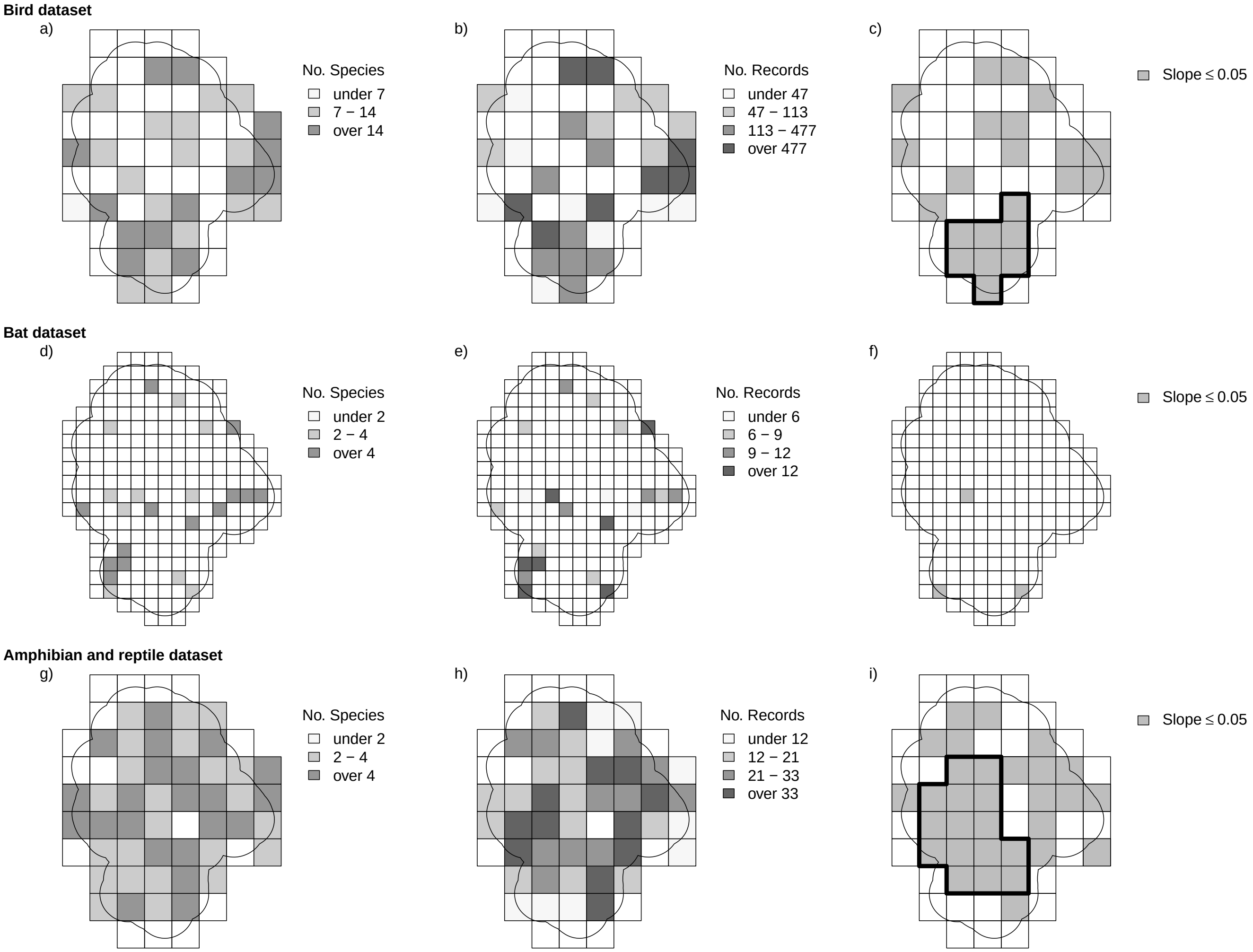

Figure 1: Species data quality maps. Species richness (a, d, g), number of records (b, e, h) and well sampled grids (c, f, i) shown for the three species datasets. The slope in c, f and i is calculated by creating a species-accumulation curve by plotting the record number against the cummulative number ofspecies; smoothing this curve by randomising the order of record entry 100 times; fitting to a clench function; and calculating the slope of the curve between the last two records. A slope of 0.05 represents a 95% confidence that all species have been recorded. Core area for use in model fitting outlined in bold.

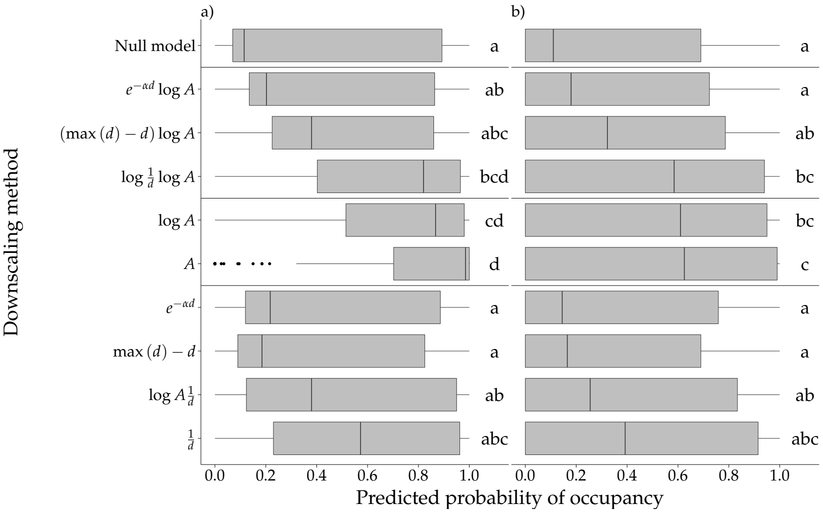

Figure 2: Boxplots showing the probability of occupancy for all species after the burn-in period (t = 175 time steps) for all grid cells recorded as a) present and b) absent in the evaluation data. For each species, there are n = 200 replicates of IFM simulations. Horizontal lines separate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species' dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

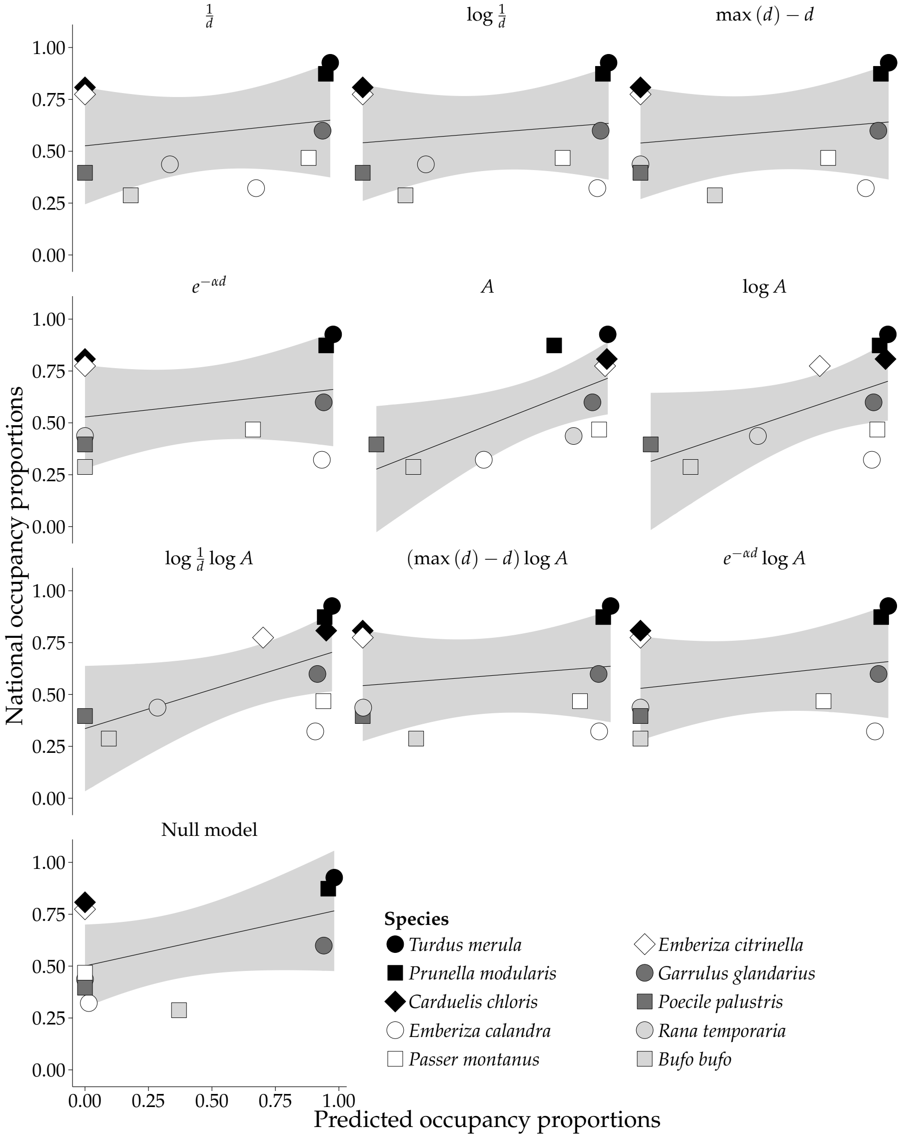

Figure 3: Comparison of mean (of n = 200 replicates of IFM simulations) predicted proportion of occupied habitat after the burn-in period (t = 175 time steps) against the national proportion of 10 km x 10 km grid cells occupied. National data are from the 2nd BTO Atlas (Gibbons et al., 1993) and Biological Records Centre Reptiles and Amphibians Dataset. Black lines are fitted linear regression lines, shaded area is ±1 SE of the regression line.

# Appendix A - Additional study site and data information

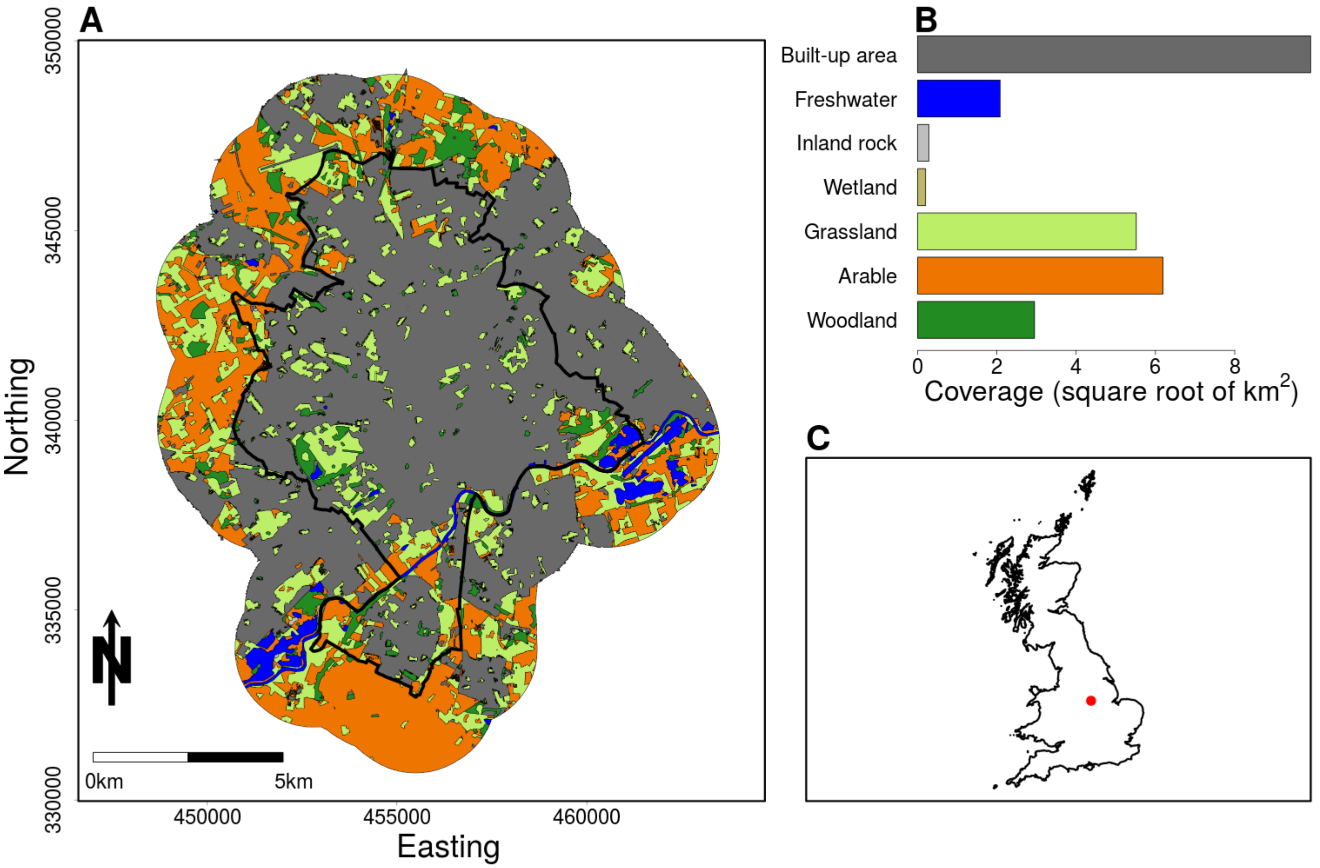

Figure A1: Study site of Nottingham, UK. Nottingham City Administrative Boundary (bold lines) with 2km buffer showing coarse land-cover classes. Inset map shows the location of Nottingham within Great Britain. Land-cover classes based on LCM 2007 (Morton et al., 2011)

Table A1: Proportions of the LCM 2007 broad habitat types in Nottingham, three nearby cities (Birmingham, Leicester and Sheffield) and for similar sized UK cities. *UK Cities is the total proportion for the 5 cities smaller in area, and the 5 cities larger in area than Nottingham (Derby, Hull, Leicester, Southampton, Dundee, Poole, Southend-on-Sea, Brighton and Hove, Portsmouth and Plymouth). Boundaries were defined using the Ordnance Survey Boundary Line shapefile 'District Borough Unitary' and adding a 2km buffer. N.B. Unitary authority boundaries are not strictly related to the size of the urban area.

| Code | Habitat Type | Nottingham | Birmingham | Leicester | Sheffield | UK Cities |
|---|---|---|---|---|---|---|
| 1 | Broadleaved Woodland | 4.77 | 4.66 | 2.50 | 9.67 | 5.03 |
| 2 | Coniferous Woodland | 0.28 | 0.59 | 0.04 | 3.02 | 0.74 |
| 3 | Arable and Horticulture | 21.66 | 15.32 | 23.13 | 10.02 | 20.92 |
| 4 | Improved Grassland | 14.63 | 16.12 | 18.71 | 20.66 | 18.85 |
| 5 | Rough Grassland | 1.86 | 1.27 | 0.74 | 4.41 | 2.02 |
| 6 | Neutral Grassland | 1.09 | 0.35 | 1.38 | 0.13 | 0.74 |
| 7 | Calcareous Grassland | - | - | - | - | 0.03 |
| 8 | Acid Grassland | 0.03 | 0.00 | - | 4.15 | 0.03 |
| 9 | Fen, Marsh and Swamp | 0.02 | - | - | 0.00 | 0.13 |
| 10 | Heather | 0.16 | 0.30 | - | 5.87 | 0.35 |
| 11 | Heather Grassland | 0.09 | 0.17 | - | 6.91 | 0.21 |
| 12 | Bog | - | 0.02 | - | 12.07 | - |
| 14 | Inland Rock | 0.05 | 0.28 | 0.39 | 0.17 | 0.25 |
| 15 | Saltwater | - | - | - | - | 6.94 |
| 16 | Freshwater | 2.41 | 0.76 | 0.87 | 1.18 | 0.51 |
| 17 | Supra-littoral Rock | - | - | - | - | 0.01 |
| 18 | Supra-littoral Sediment | - | - | - | - | 0.39 |
| 19 | Littoral Rock | - | - | - | - | 0.26 |
| 20 | Littoral Sediment | - | - | - | - | 4.36 |
| 21 | Saltmarsh | - | - | - | - | 0.5 |
| 22 | Urban | 8.21 | 14.15 | 12.31 | 6.05 | 9.11 |
| 23 | Suburban | 44.77 | 46.01 | 39.92 | 15.69 | 28.62 |

Table A2: Distributions of patch sizes for the LCM classes which make up the species' habitat.

| LCM Class | Mean area (ha) | Min. area (ha) | Max. area (ha) | # patches |
|---|---|---|---|---|
| Broadleaved Woodland | 3.23 | 0.00087 | 87.63 | 254 |
| Coniferous Woodland | 1.68 | 0.00004 | 8.08 | 29 |
| Arable and Horticulture | 19.08 | 0.00175 | 1015.37 | 201 |
| Improved Grassland | 6.35 | 0.00047 | 118.68 | 392 |
| Rough Grassland | 2.52 | 0.05196 | 11.00 | 128 |
| Neutral Grassland | 2.52 | 0.00178 | 15.97 | 76 |
| Acid Grassland | 1.56 | 0.73430 | 2.63 | 3 |
| Fen, Marsh and Swamp | 1.03 | 0.61178 | 1.68 | 4 |
| Heather | 0.79 | 0.19012 | 1.56 | 28 |
| Heather Grassland | 3.15 | 0.44680 | 10.10 | 5 |
| Freshwater | 12.07 | 0.06755 | 131.07 | 36 |

Table A3: Full list of organisations whose data were used. All data accessed from NBN Gateway except data from Nottinghamshire Birdwatchers.

| Data provider | Contact name |
|---|---|
| Amphibian and Reptile Conservation | John Wilkinson |
| Biological Records Centre | Dr David Roy |
| British Trust for Ornithology | Peter Lack |
| Derbyshire Biological Records Centre | Kieron Huston |
| Nottinghamshire Biological and Geological Records Centre | Rob Johnson |
| The Bat Conservation Trust | Philip Briggs |
| Nottinghamshire Birdwatchers | Andy Hall |

## Appendix B - Additional methods information

The incidence function model (IFM), a stochastic patch occupancy model developed by Hanski (1994), allows long-term predictions of metapopulation persistence in a network of habitat patches to be made through estimation of colonisation and extinction rates. The occupancy of a patch $i$ is given by $J_i$ where $J_i$ is a balance of colonisations ($C_i$) and extinctions ($E_i$).

$$J_i = \frac{C_i}{C_i + E_i - C_i E_i} \tag{B1}$$

The extinction probability (Equation B2) is calculated following the assumption that the species richness is directly proportional to the area of the patch they occupy. The species–area relationship is a well established concept, and as such the assumption can be justified on both empirical and theoretical grounds (Hanski, 1999). Extinction is affected by population size, so $E_i$ can therefore be expressed as a function of $A_i$:

$$E_i = min\left[\frac{u}{A_i^x}, 1\right] \tag{B2}$$

Here, $u$ and $x$ are species specific parameters, and patch $i$ is currently occupied. The critical patch size, below which a species cannot survive in the patch, is given by $u^{\frac{1}{x}}$ (all patches $\leq u^{\frac{1}{x}}$ have extinction probability 1). Parameter $x$ represents the extent to which a species' survival is dependent on patch size (larger $x$ represents weaker dependence).

The colonisation probability (Equation B3) is a function of patch connectivity $S_i$ (Equation B4). Species-specific parameter $y$ represents the level of connectivity required by a species to achieve colonisation.

$$C_i = \frac{S_i^2}{S_i^2 + y^2} \tag{B3}$$

Finally, connectivity (Equation B4) is a function of the distance from patch $i$ to patch $j$ ($d_{ij}$), the occupancy and area of patch $j$ ($p_j$ and $A_j$ respectively) and the species' mean natal dispersal ability, $\frac{1}{\alpha}$. Mean natal dispersal distance is derived from the literature.

$$S_i = \sum_{j=1}^{n} e^{-\alpha d_{ij}} p_j A_j \tag{B4}$$

Parameters $u$, $x$ and $y$ are estimated from the data. For each species, we fitted the presence–absence data resulting from each downscaling method for an eight-year period to a logistic regression model derived from the above equations (Equation B5). The following two years of data were combined and kept to evaluate the performance of each downscaling method.

$$\operatorname{logit}(J_i) = \beta_0 + 2 \log S + \beta_1 \log A \tag{B5}$$

Here $\beta_0 = -\log(uy)$ and $\beta_1 = x$. Parameter $u$ is estimated by assuming that for the smallest of all occupied patches, $E_i = 1$, then solving equation B2 for $u$. This value is then substituted into $\beta_0$ to solve for $y$. This follows the method outlined by Oksanen (2004).

## Appendix C - IFM outputs

The IFM was used to simulate the patch occupancies for 14 species for each of the 10 downscaling methods. 100 iterations of 500 timesteps were simulated. This Appendix gives plots of the model output calculated using the proportion of the number of suitable patches occupied, and the proportion of the total suitable area occupied. The 175th timestep was chosen as the burn-in period; this timestep is shown on each of the plots.

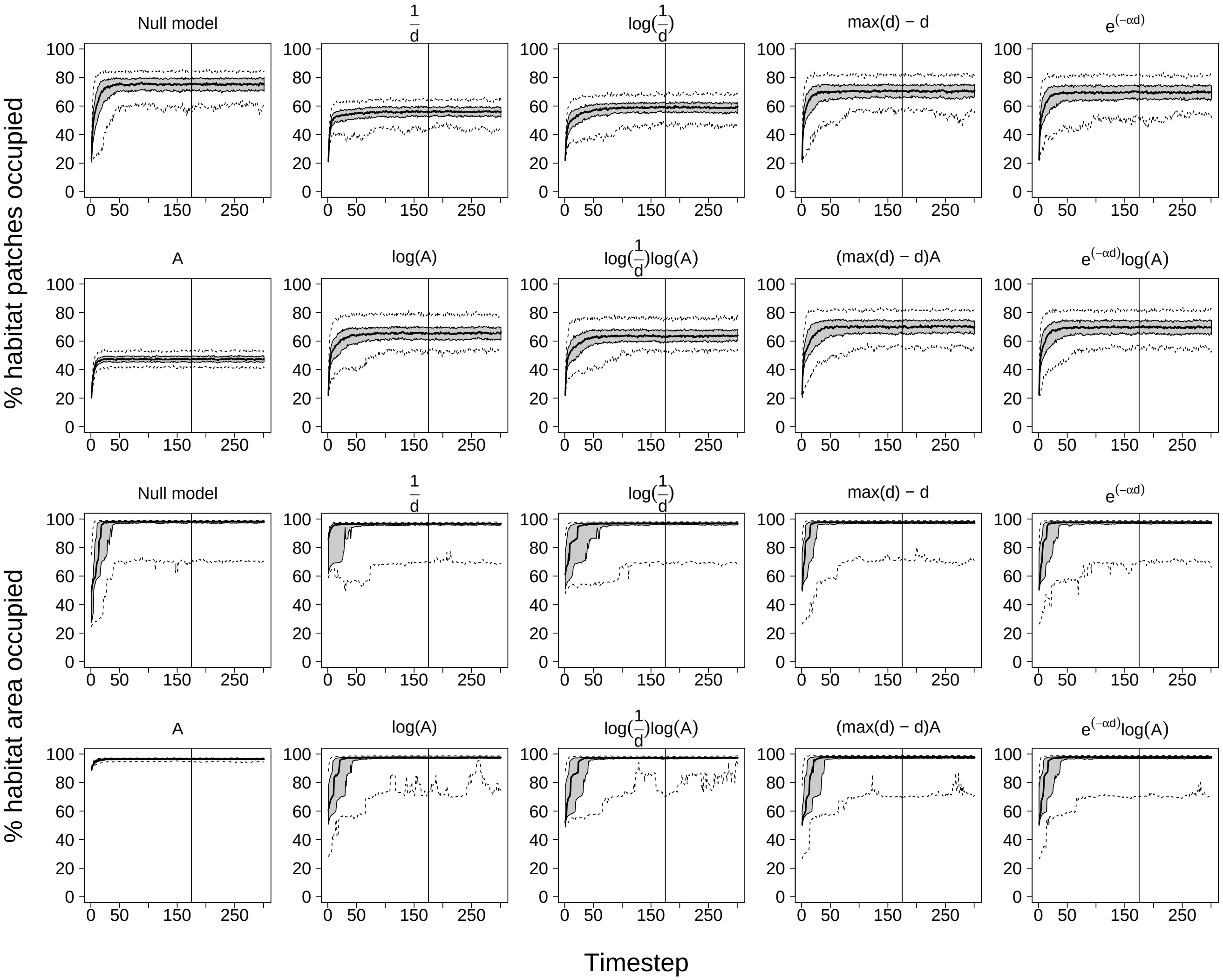

Figure C1: IFM output trajectories for *Turdus merula* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

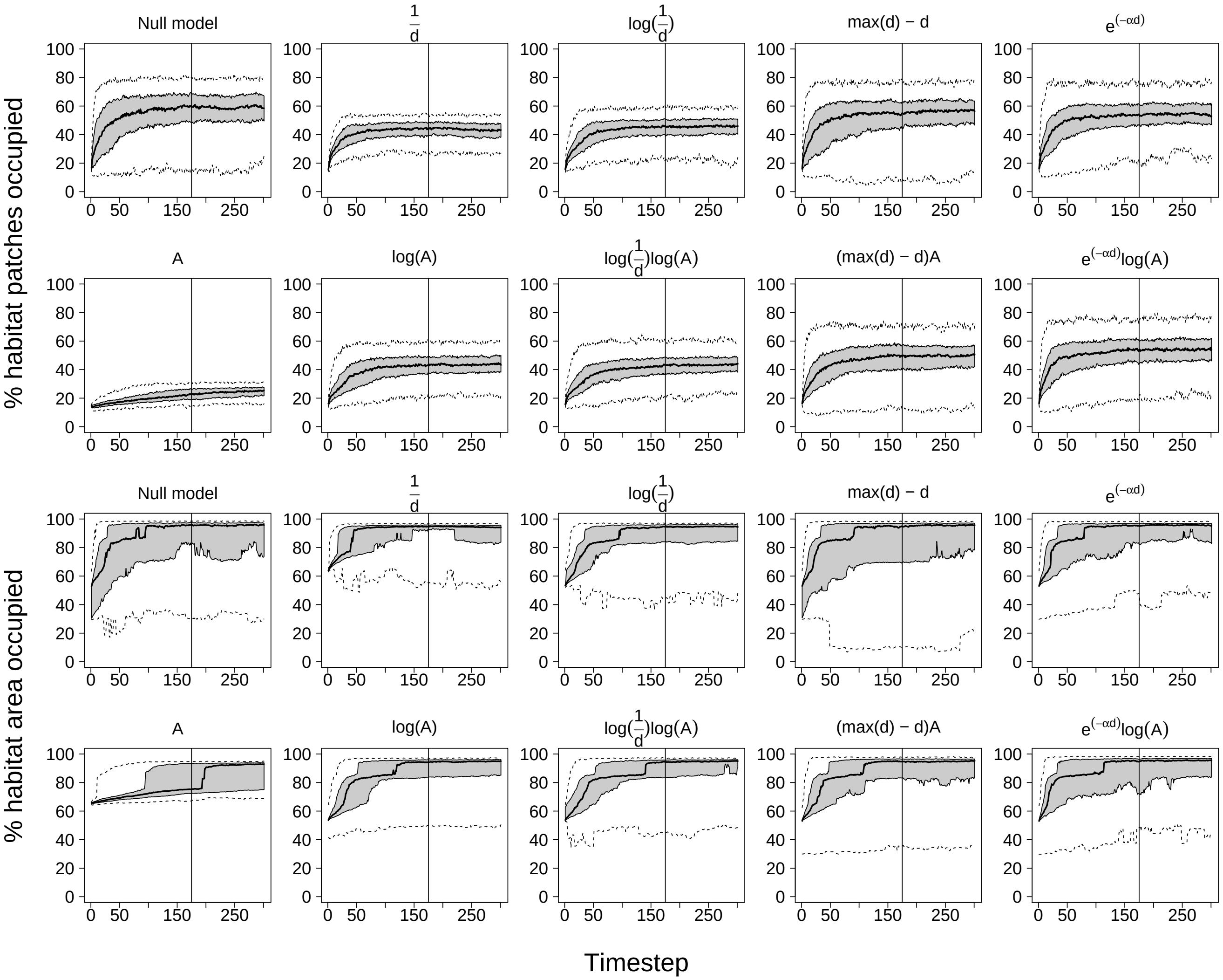

Figure C2: IFM output trajectories for *Prunella modularis* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

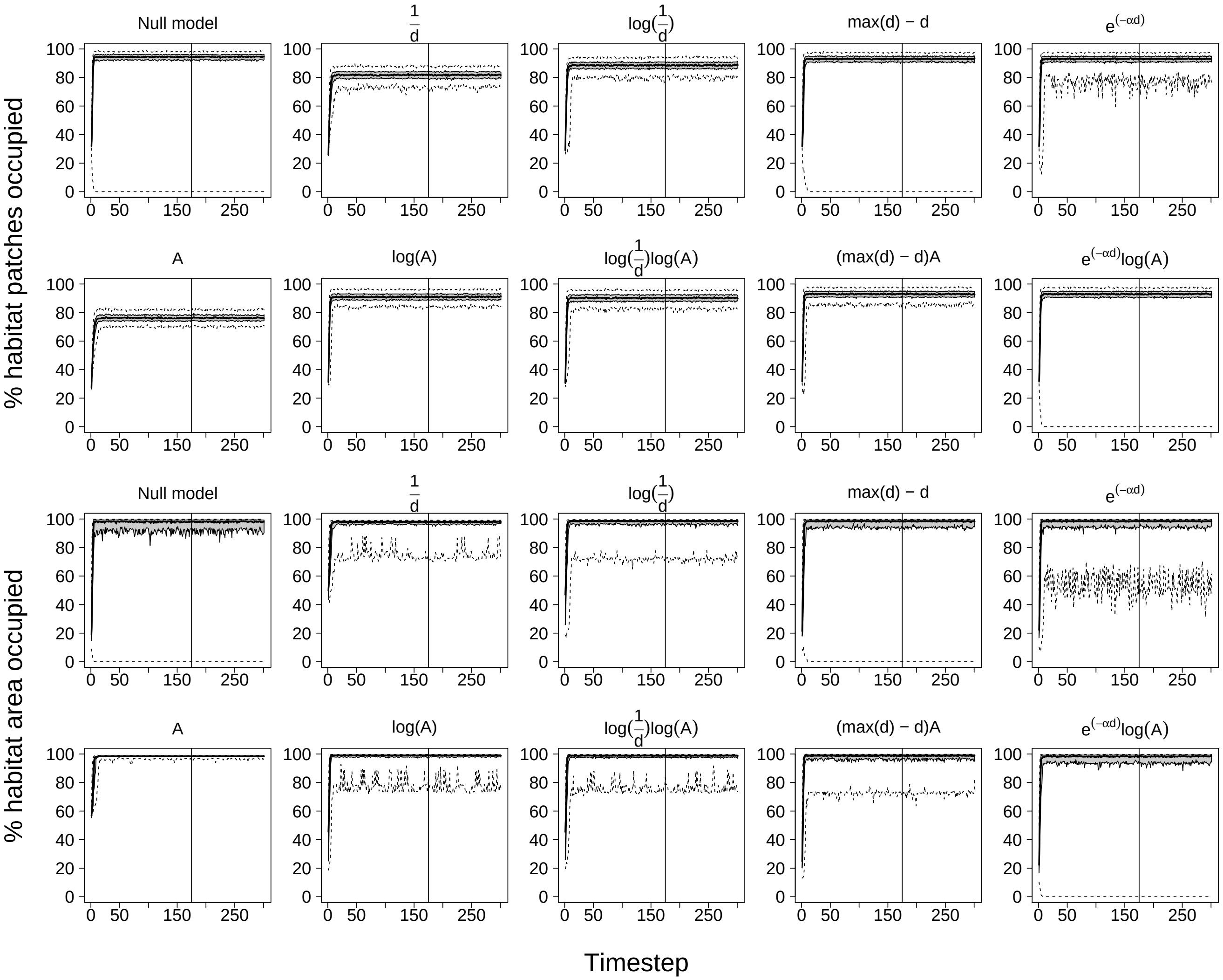

Figure C3: IFM output trajectories for *Carduelis carduelis* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

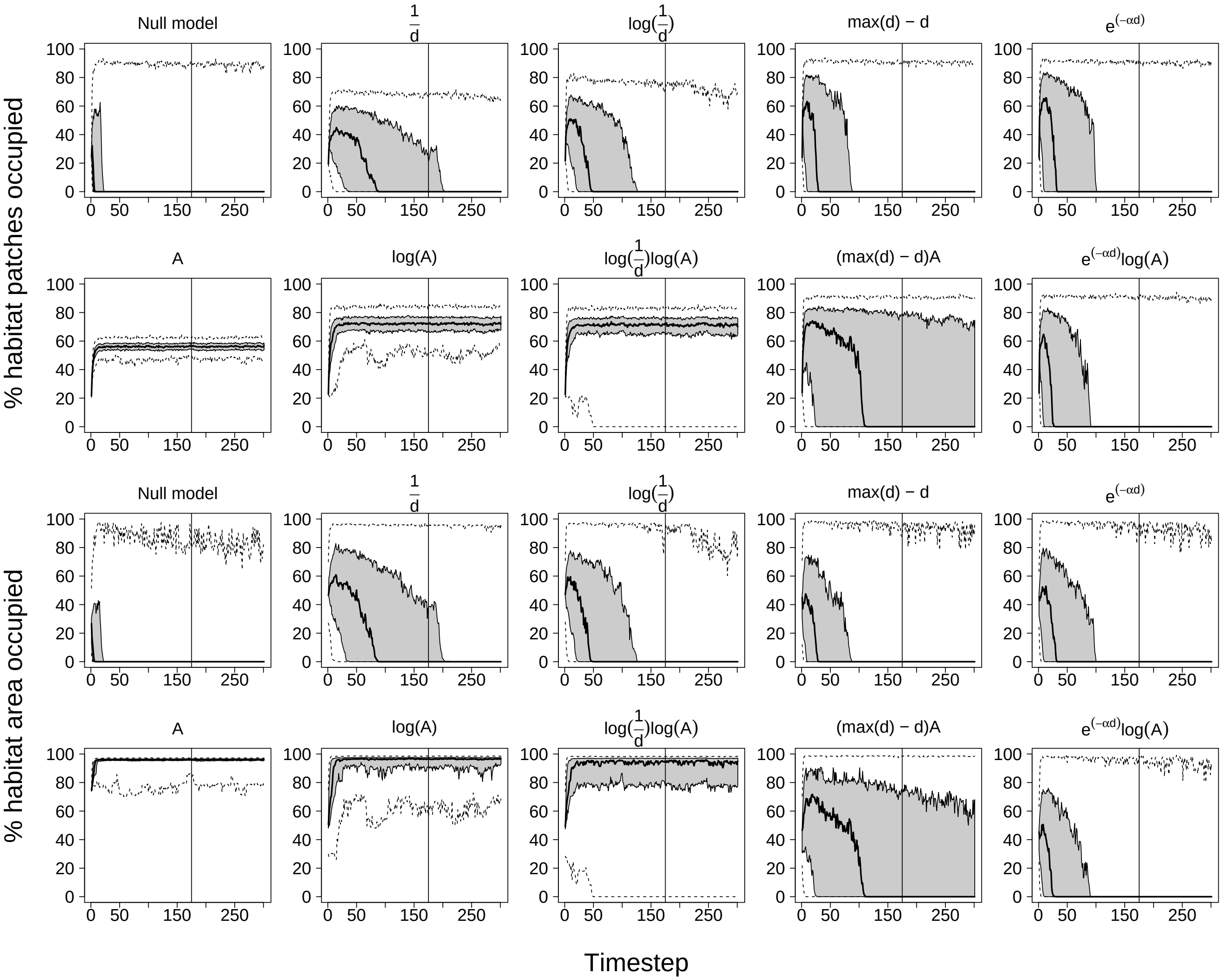

Figure C4: IFM output trajectories for *Carduelis chloris* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

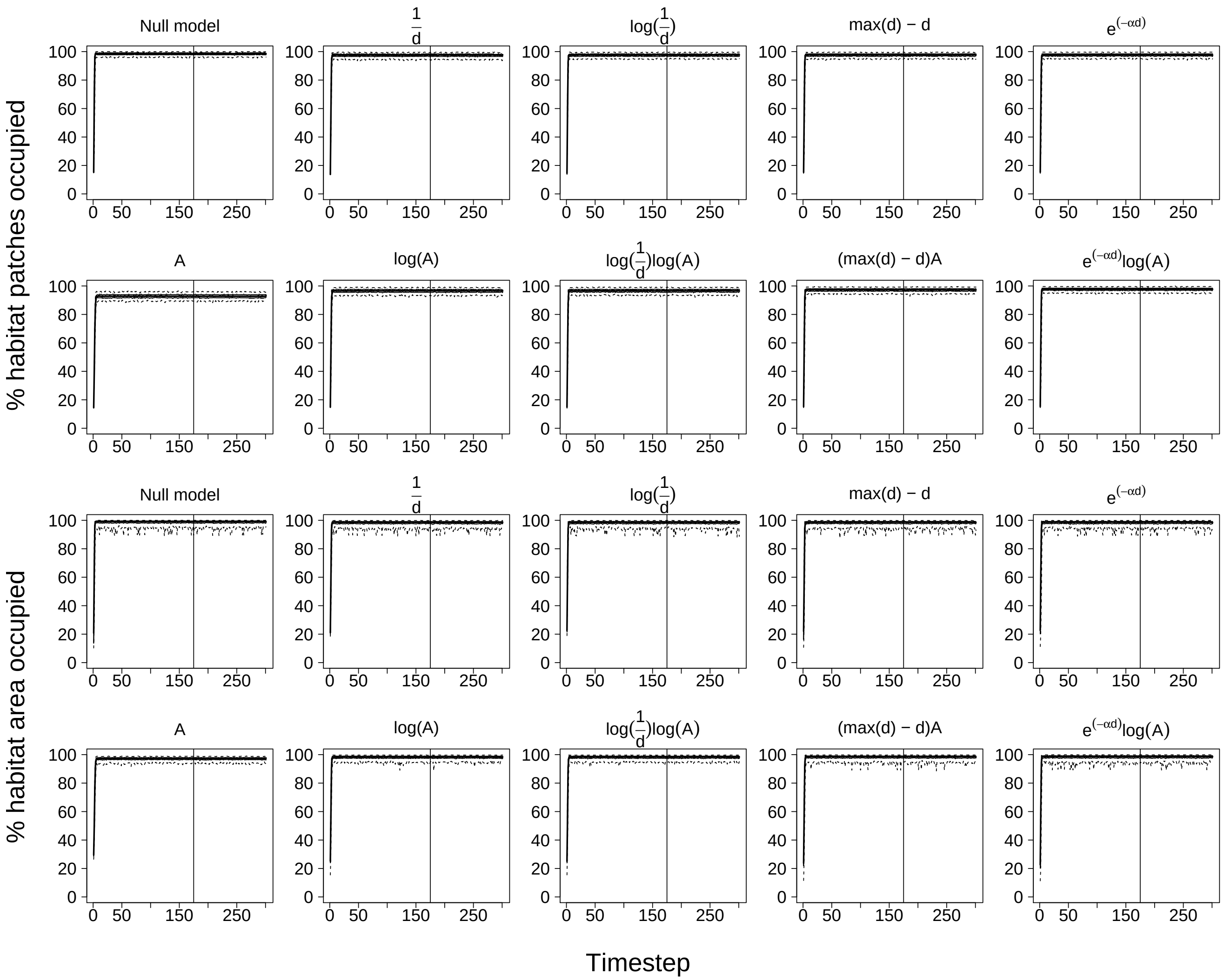

Figure C5: IFM output trajectories for *Muscicapa striata* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

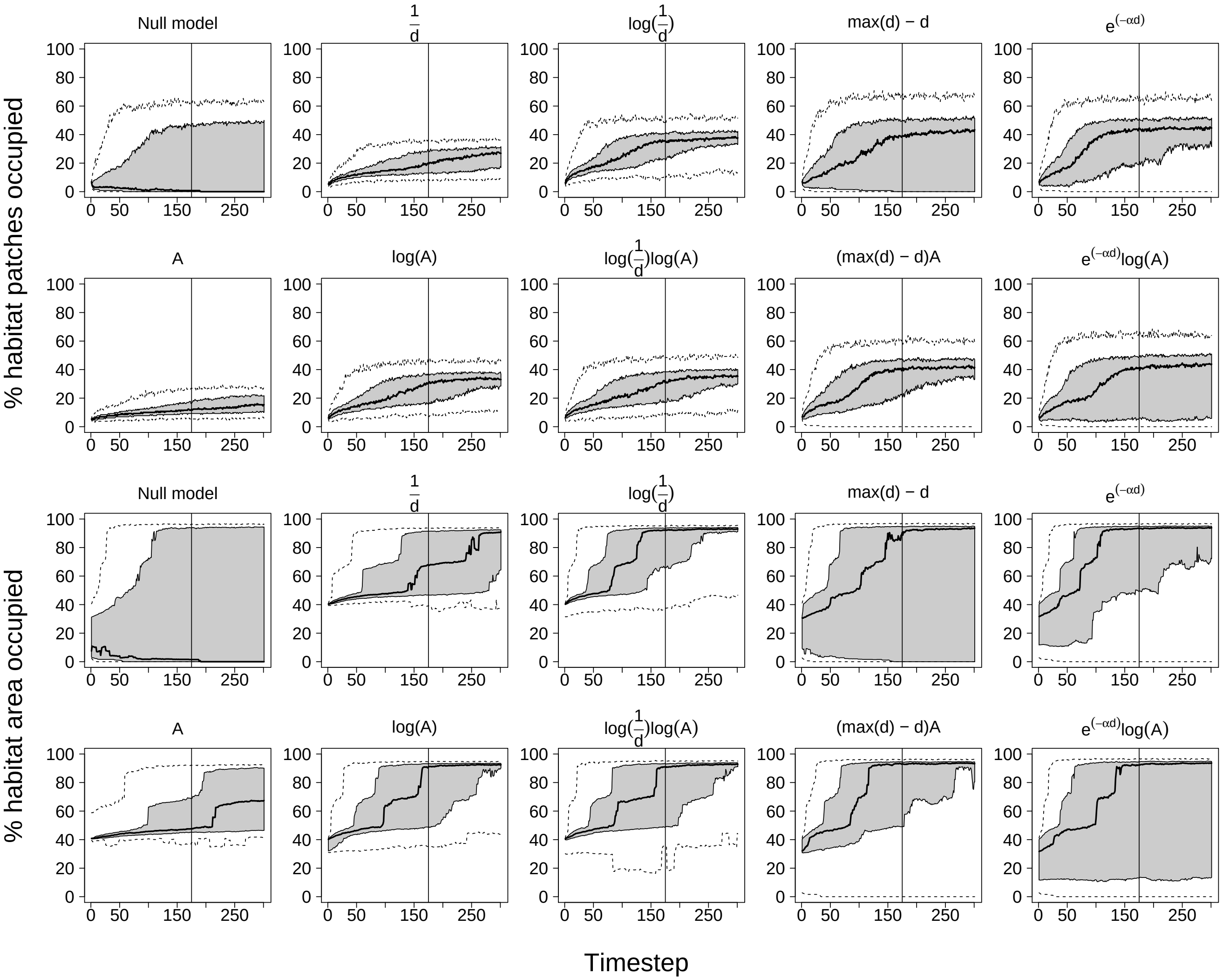

Figure C6: IFM output trajectories for *Emberiza calandra* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

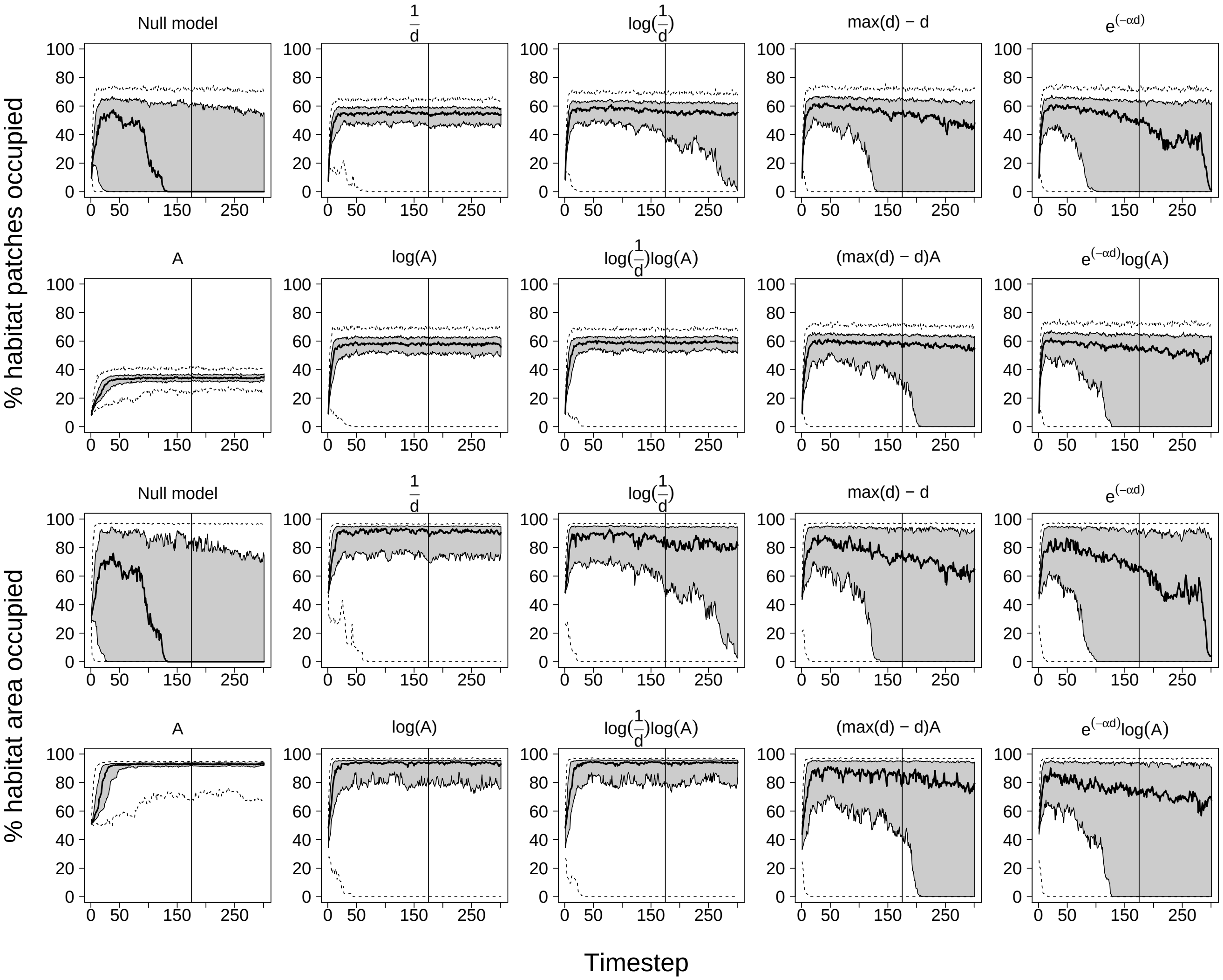

Figure C7: IFM output trajectories for *Passer montanus* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

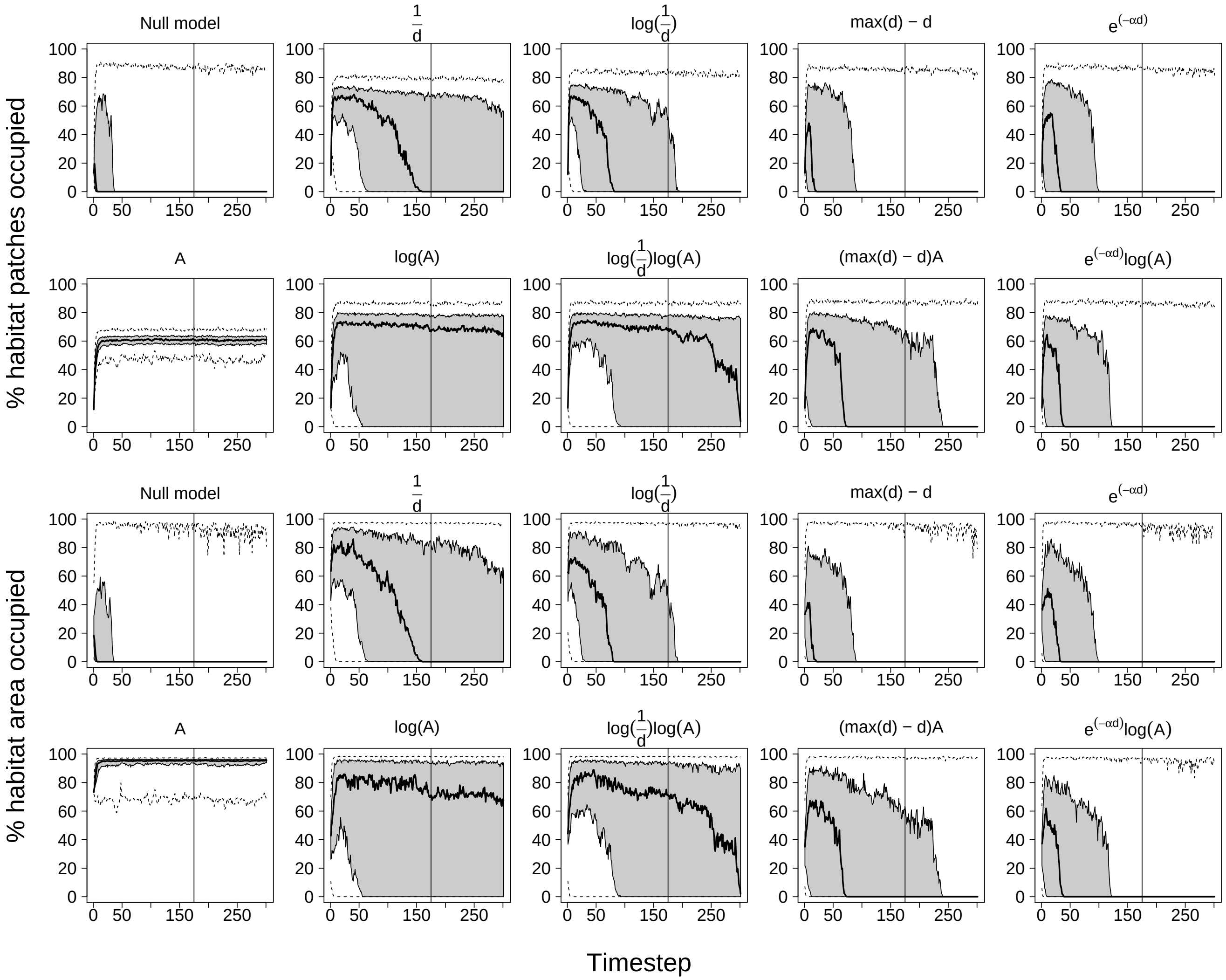

Figure C8: IFM output trajectories for *Emberiza citrinella* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

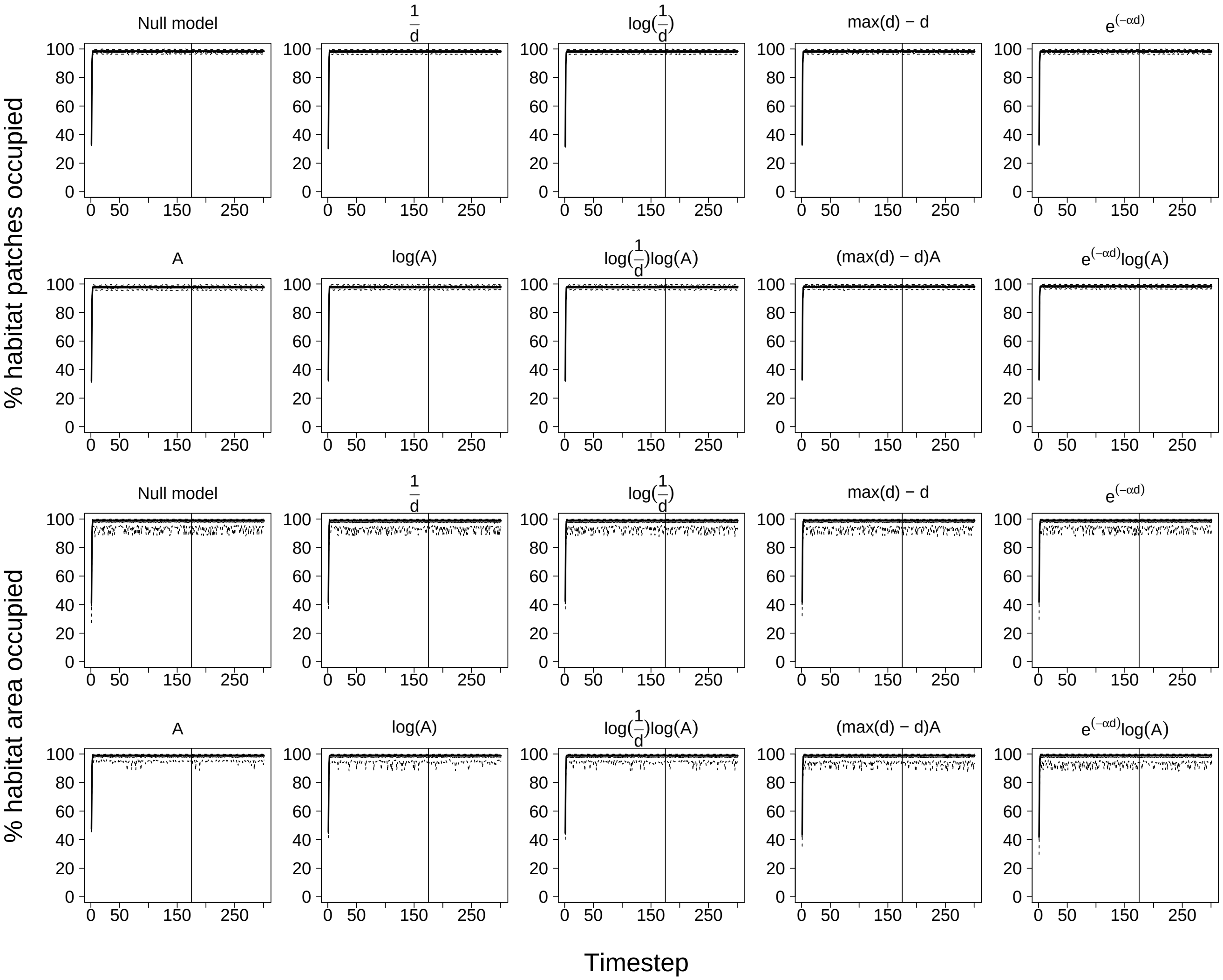

Figure C9: IFM output trajectories for *Sylvia atricapilla* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

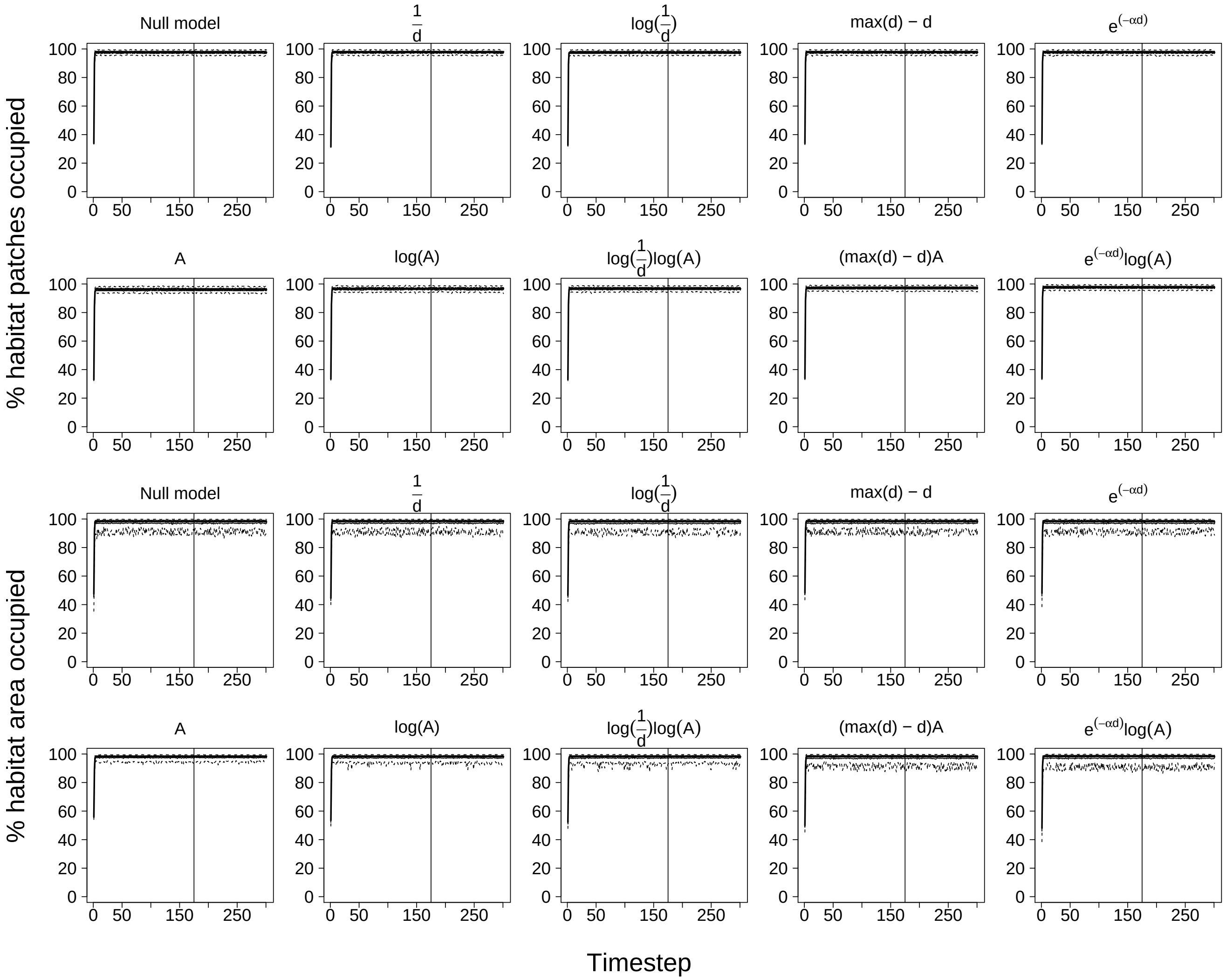

Figure C10: IFM output trajectories for *Dendrocopos major* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

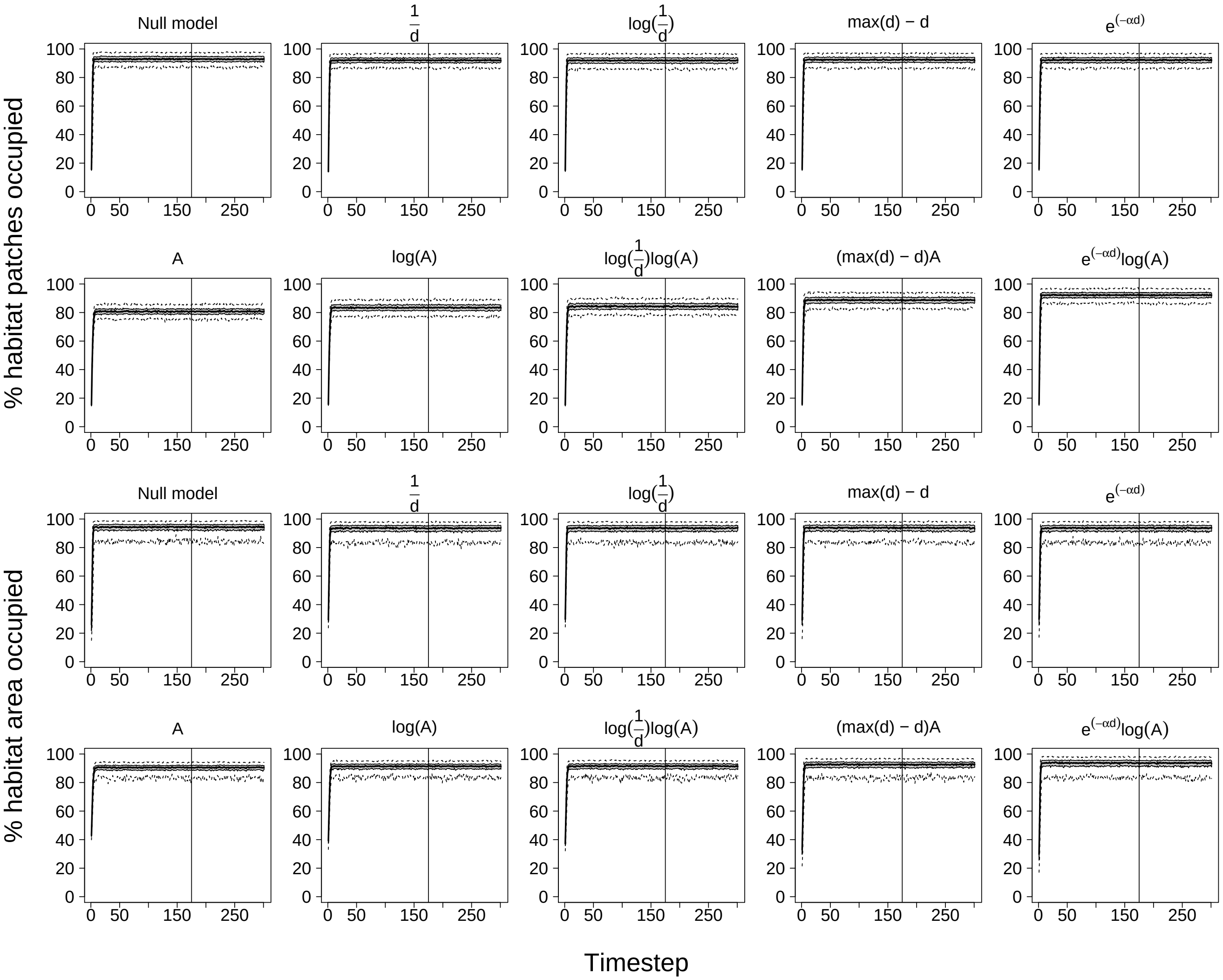

Figure C11: IFM output trajectories for *Garrulus glandarius* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

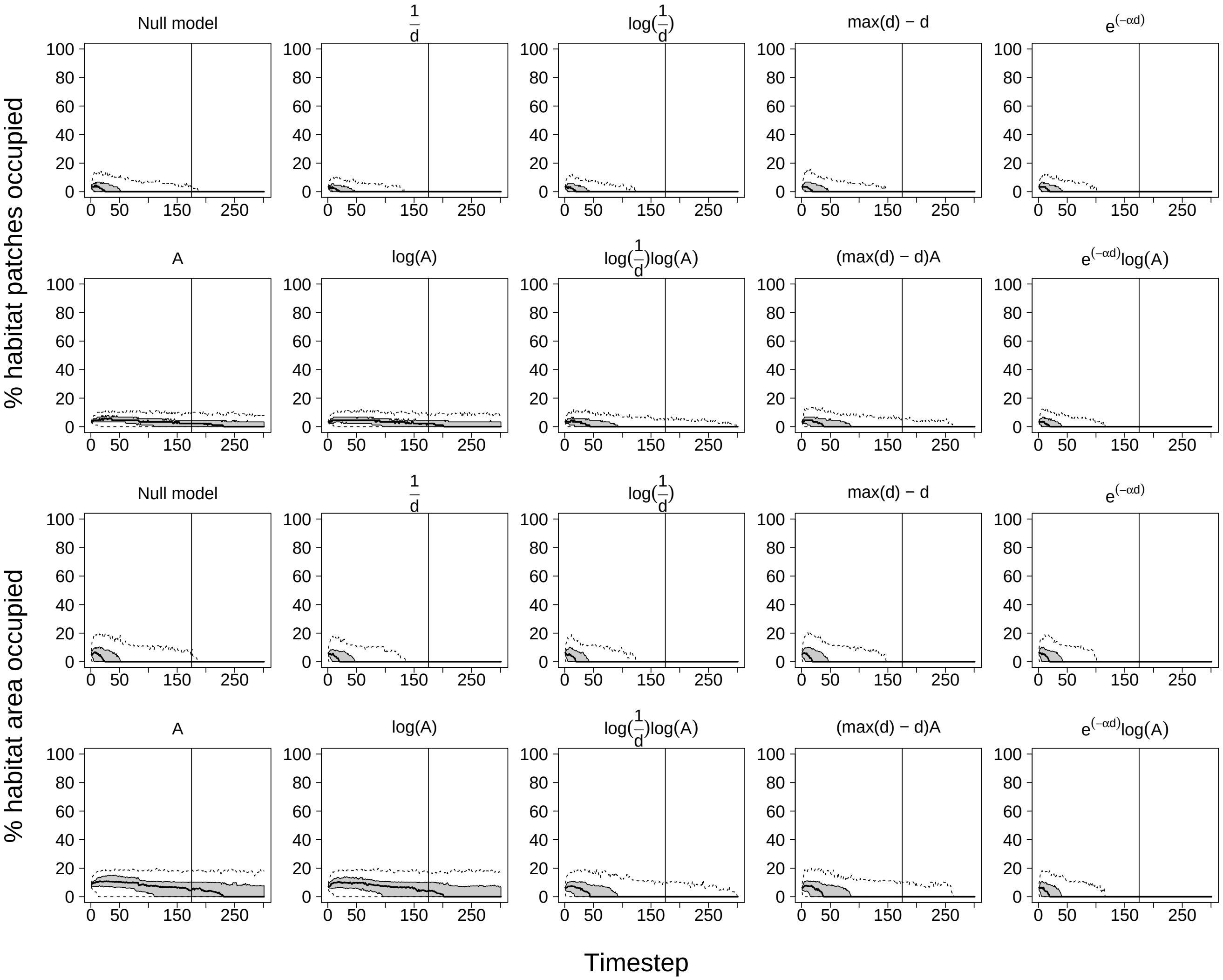

Figure C12: IFM output trajectories for *Poecile palustris* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

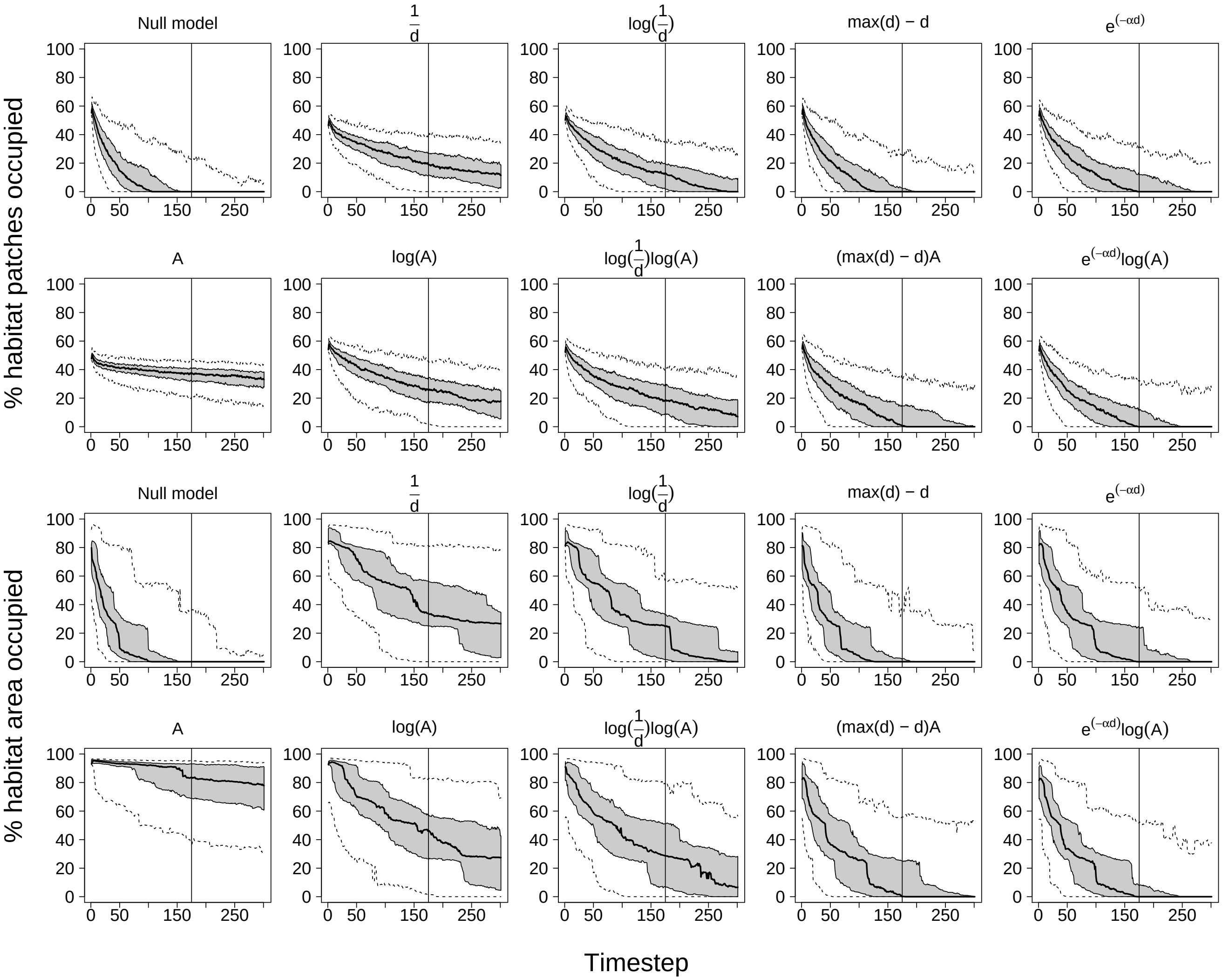

Figure C13: IFM output trajectories for *Rana temporaria* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

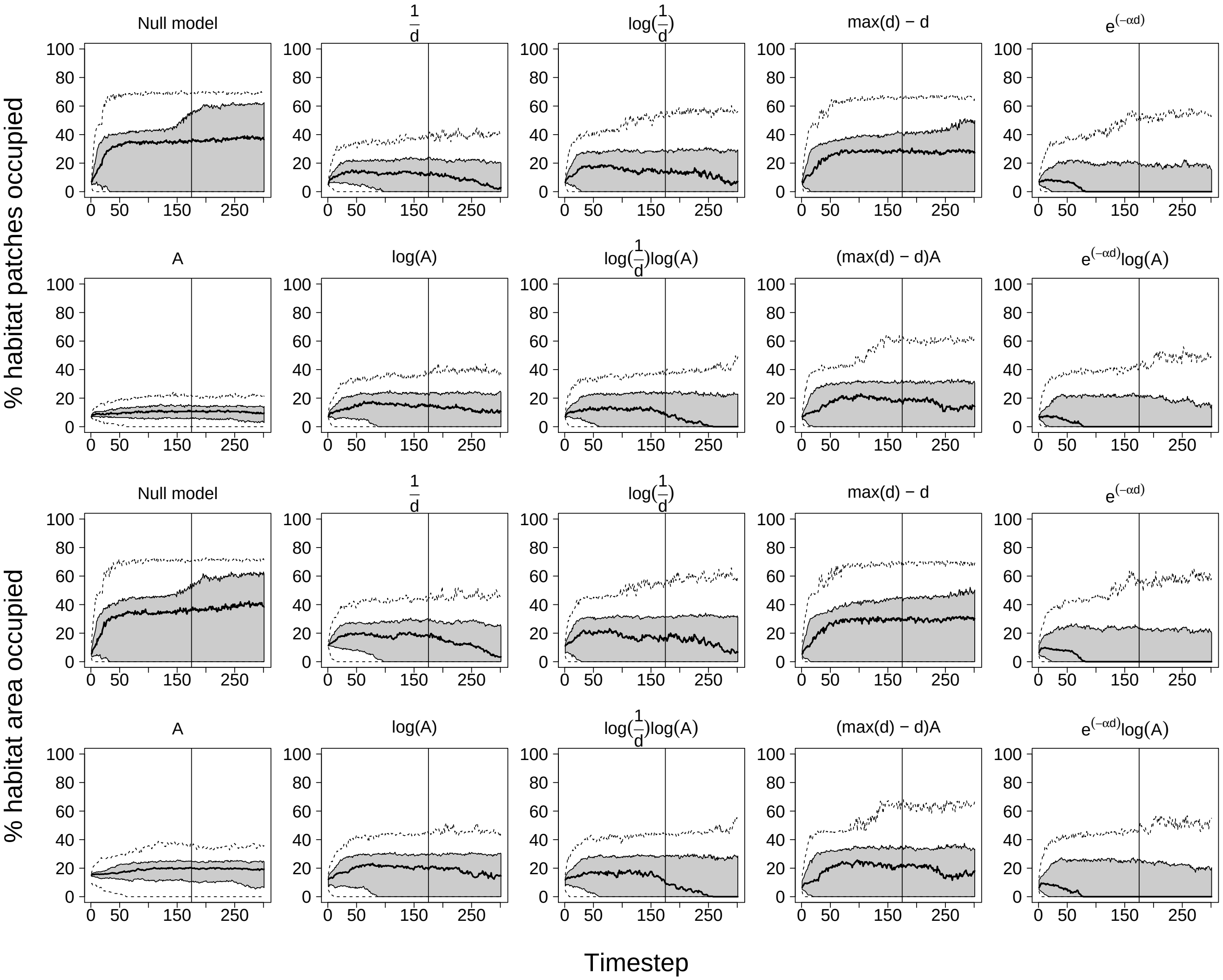

Figure C14: IFM output trajectories for *Bufo bufo* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, n = 200 iterations of the IFM. Red solid line represents the end of the burn-in period.

## Appendix D - Summary of model and downscaling performance by species

The below boxplots show the predicted probability of occupancy at t=175 for known presences (Figures D1 and D2) and presumed absences (Figures D3 and D4) obtained by running the IFM on starting occupancies based on each of the ten downscaling methods. Results are separated out by species. Although the comparison against known presences when broken down by species are not as clear cut as the combined results shown in Figure 2 (for example no significant difference is detected between downscaling methods for *T. merula*, *P. modularis*, *G. glandarius* or *B. bufo*), running the IFM based on an occupancy pattern created from the area or log(area) weighted downscaling method still yields the most reliable predictions of presences. It is possible, however, that parameterising the IFM with data from the area or log(area) weighted downscaling method will cause overpredictions.

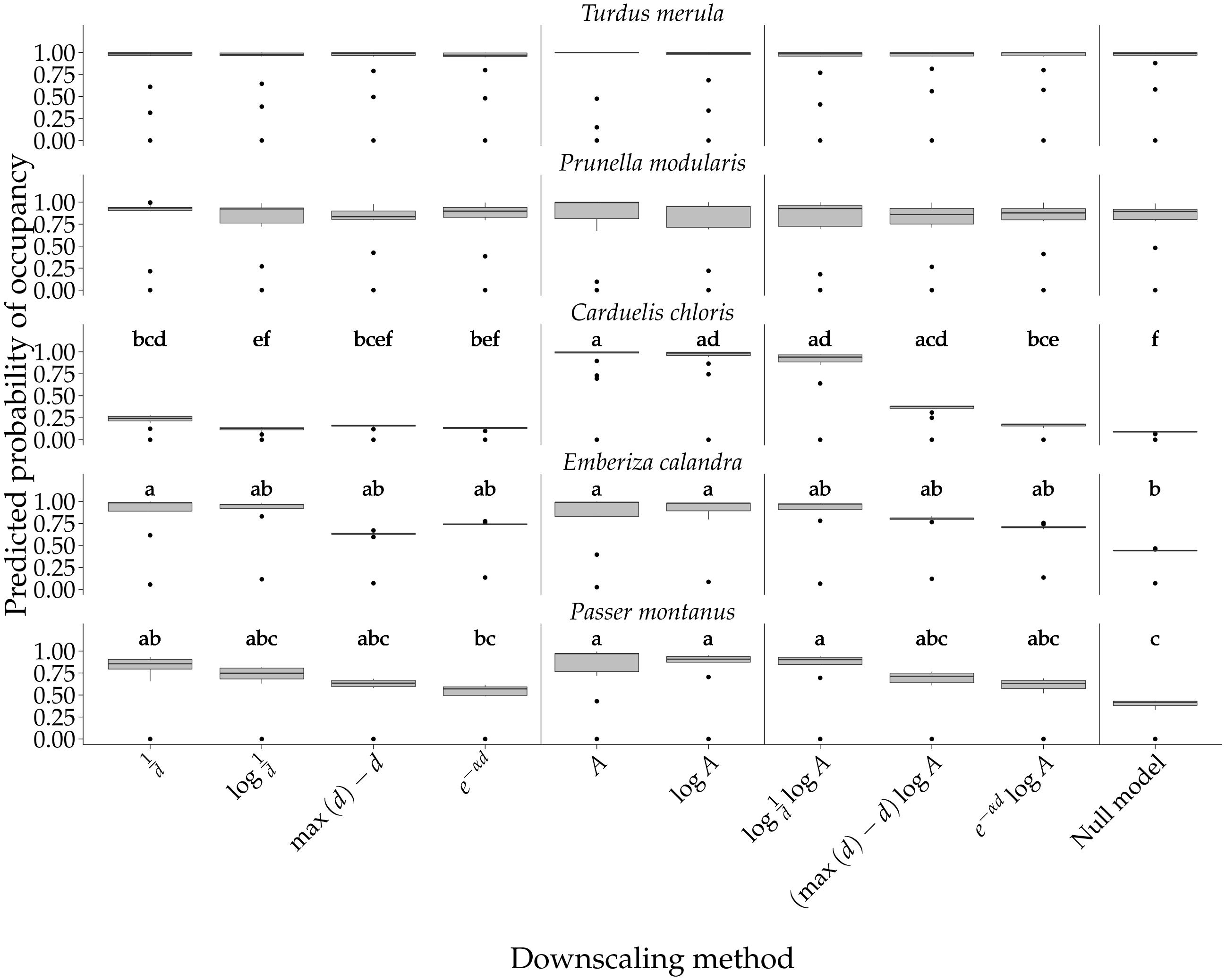

Figure D1: Boxplots showing the probability of occupancy for individual species after the burn-in period (t = 175) for all grid cells recorded as present in the evaluation data. Vertical lines seperate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

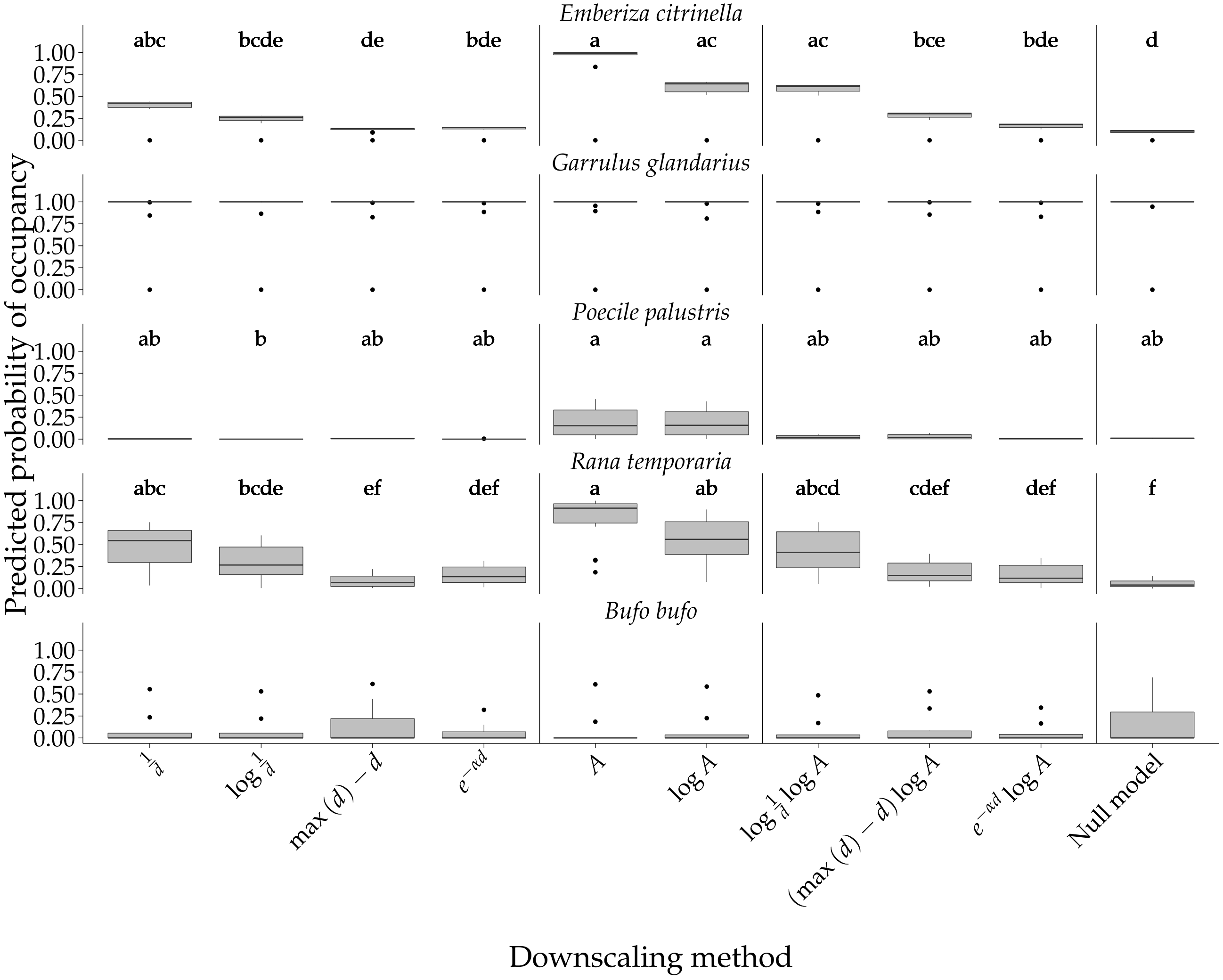

Figure D2: Boxplots showing the probability of occupancy for individual species after the burn-in period (t = 175) for all grid cells recorded as present in the evaluation data. Vertical lines seperate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

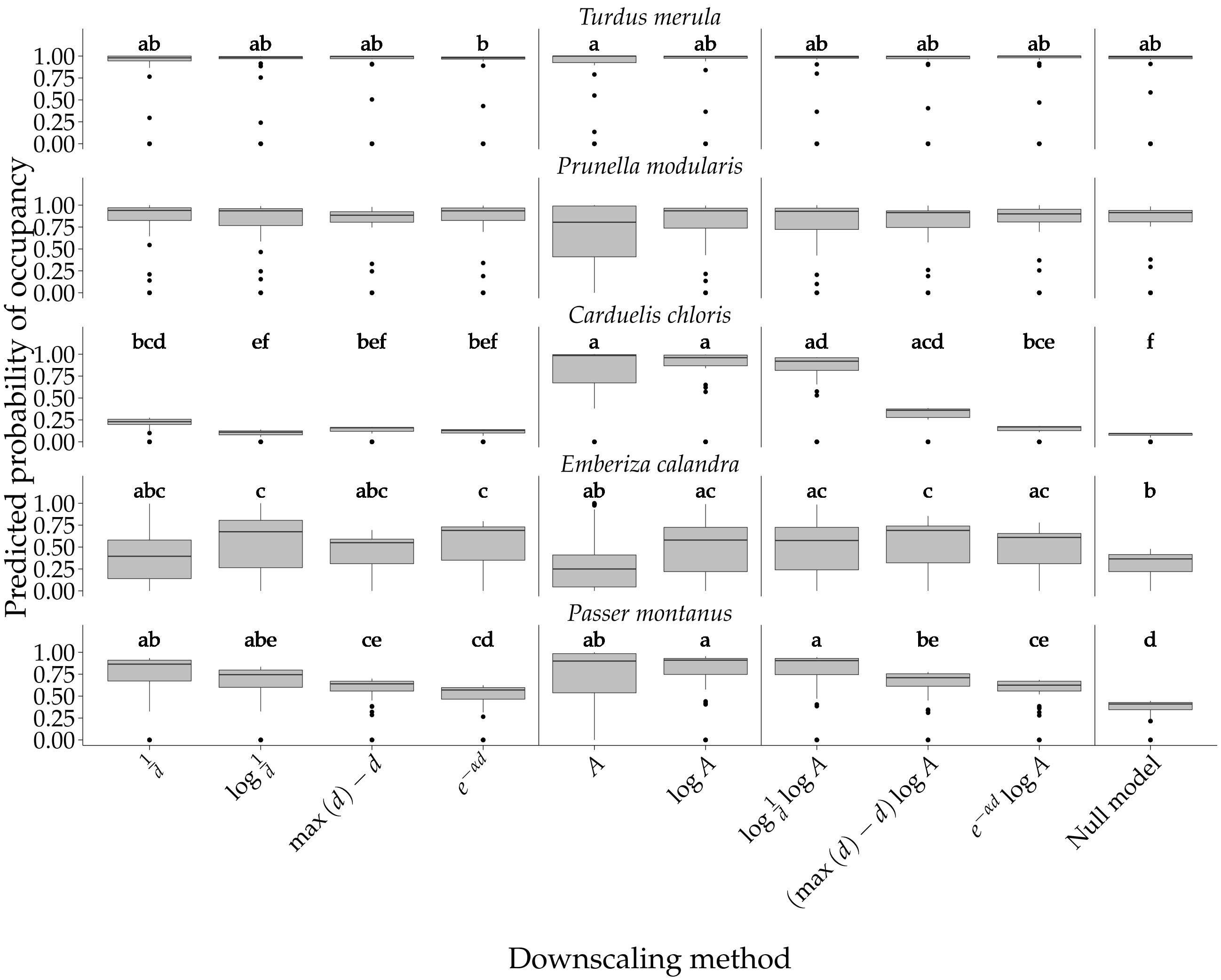

Figure D3: Boxplots showing the probability of occupancy for individual species after the burn-in period (t = 175) for all grid cells recorded as present in the evaluation data. Vertical lines seperate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

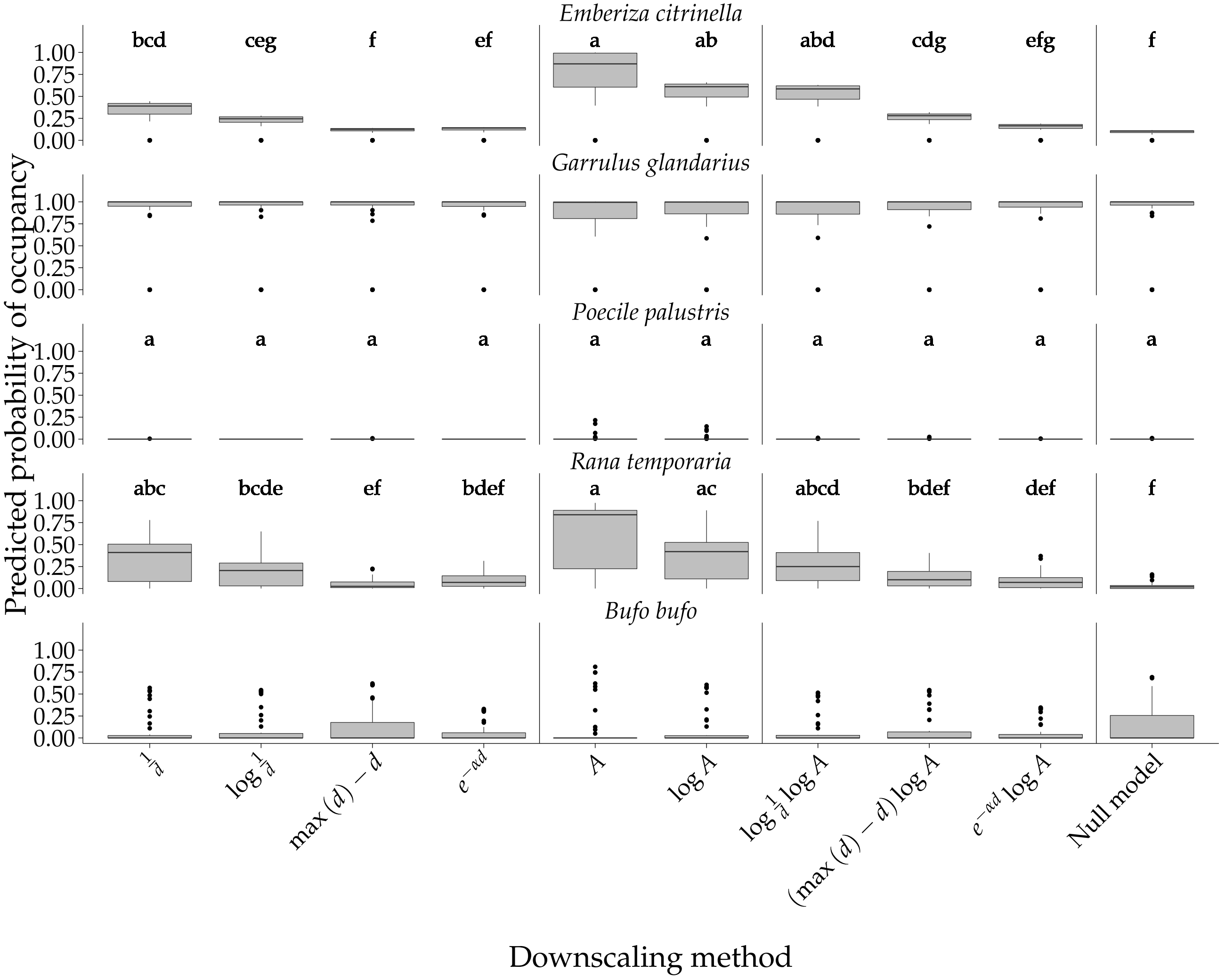

Figure D4: Boxplots showing the probability of occupancy for individual species after the burn-in period (t = 175) for all grid cells recorded as present in the evaluation data. Vertical lines seperate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

## Appendix E - Validation with independent data from the British Trust for Ornithology

Validation against an independent data set — the 2 km data from the fieldwork for the 2nd British Trust for Ornithology Atlas (1988–1991, Gibbons et al. 1993) — was performed in the same way as the validation against the Nottinghamshire Birdwatchers dataset, but limited to only grid cells visited in the BTO data.

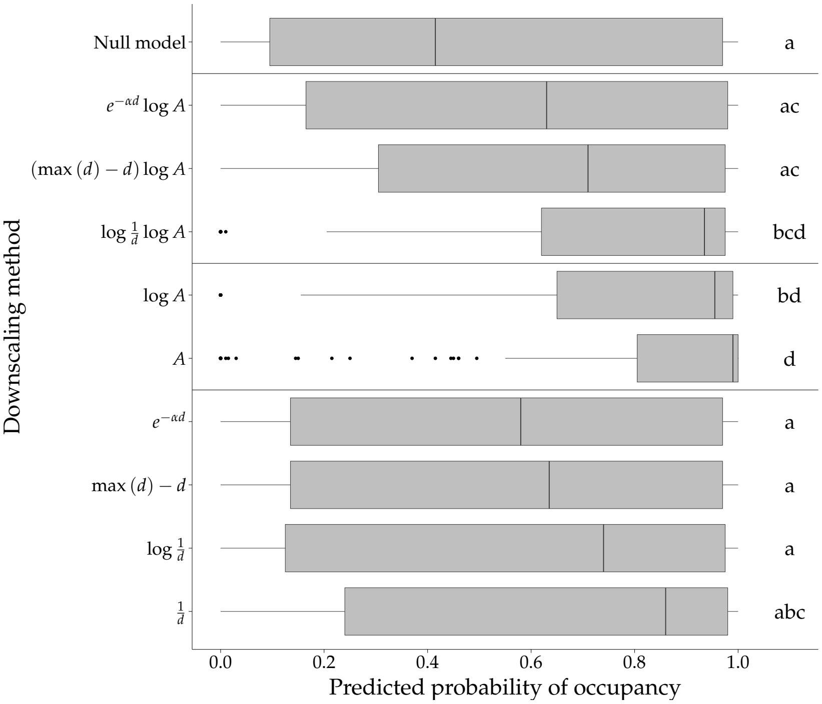

Figure E1: Boxplots showing the probability of occupancy for all species after the burn-in period (t = 175 time steps) for all grid cells recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). For each species, there are n = 200 replicates of IFM simulations. Horizontal lines separate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species' dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$). Kruskal-Wallis Chi-square = 119.9, df = 9, $P < 0.001$.

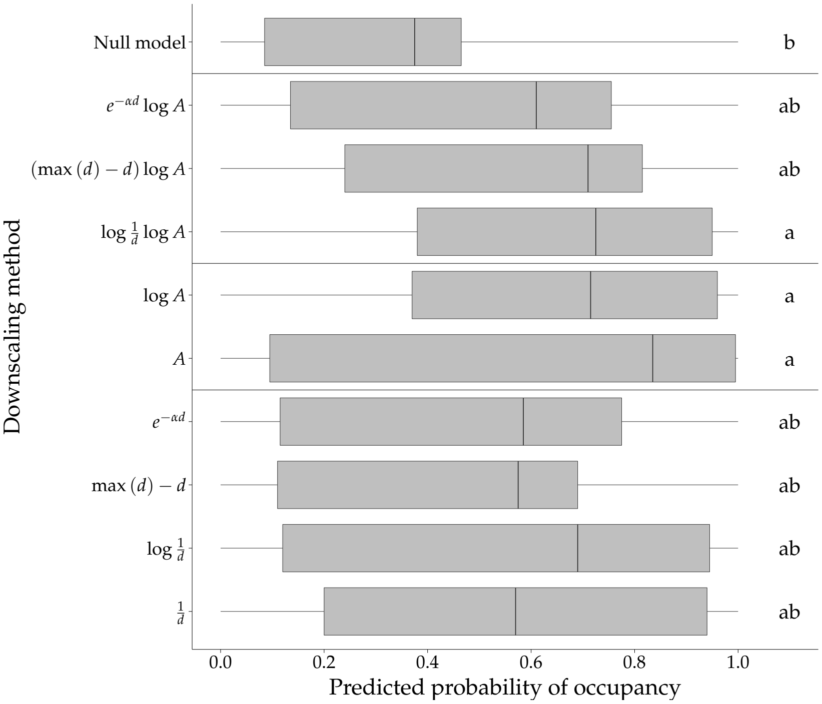

Figure E2: Boxplots showing the probability of occupancy for all species after the burn-in period (t = 175 time steps) for all grid cells which were visited but the species not recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). For each species, there are n = 200 replicates of IFM simulations. Horizontal lines separate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species' dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$). Kruskal-Wallis Chi-square = 40.7, df = 9, $P < 0.001$.

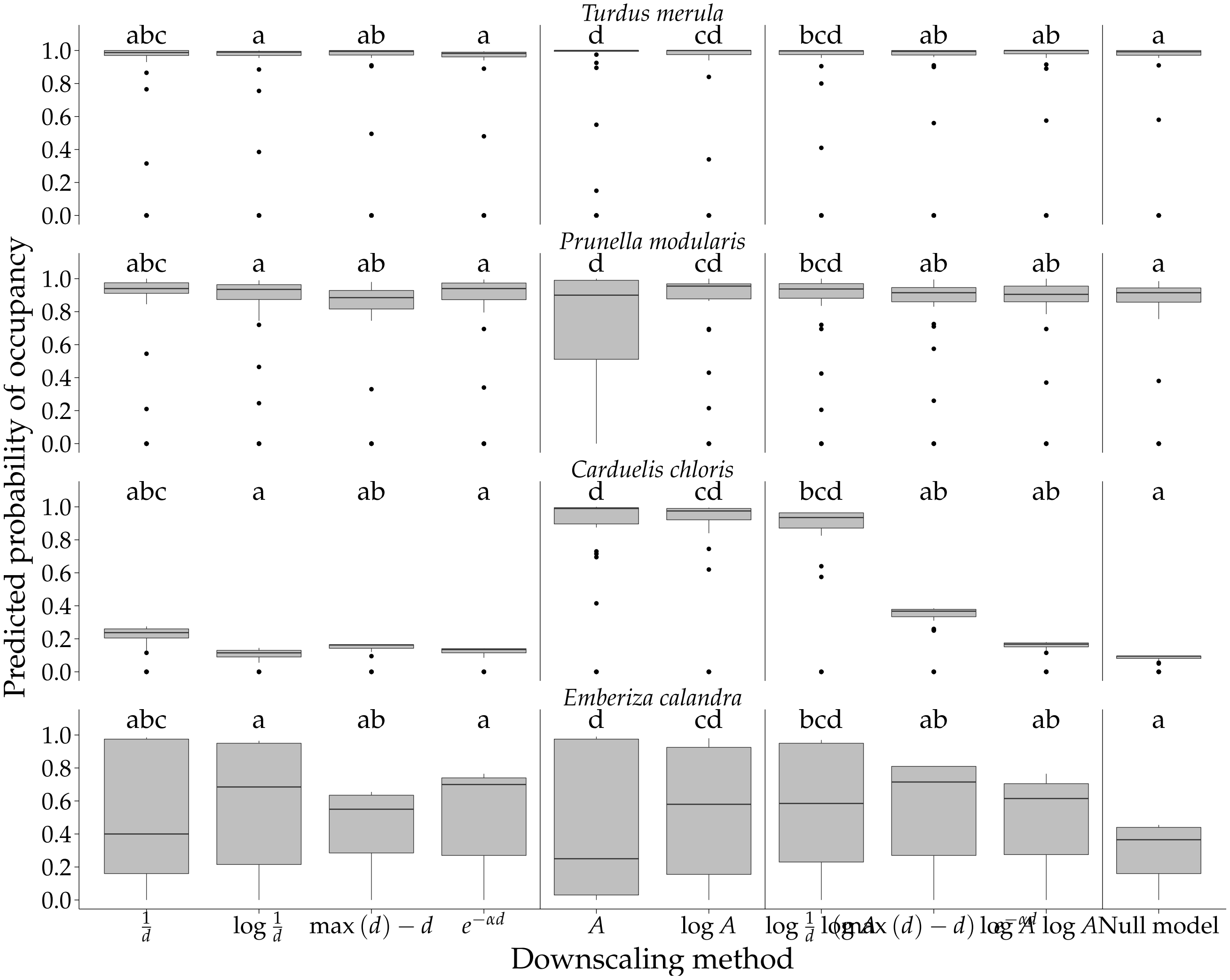

Figure E3: Boxplots showing the probability of occupancy for individual species after the burn-in period (t = 175) for all grid cells recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). Vertical lines seperate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

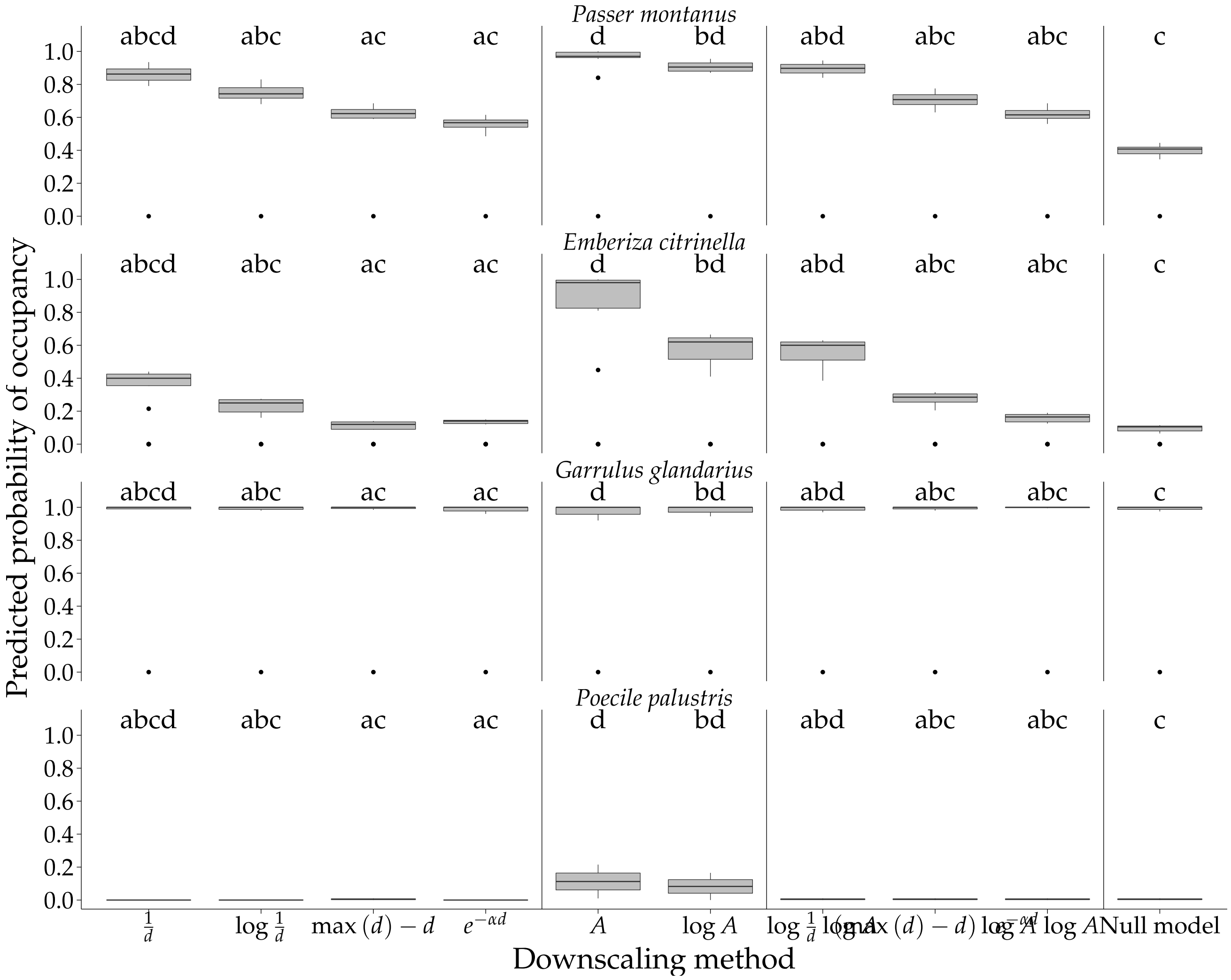

Figure E4: Boxplots showing the probability of occupancy for individual species after the burn-in period (t = 175) for all grid cells recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). Vertical lines seperate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

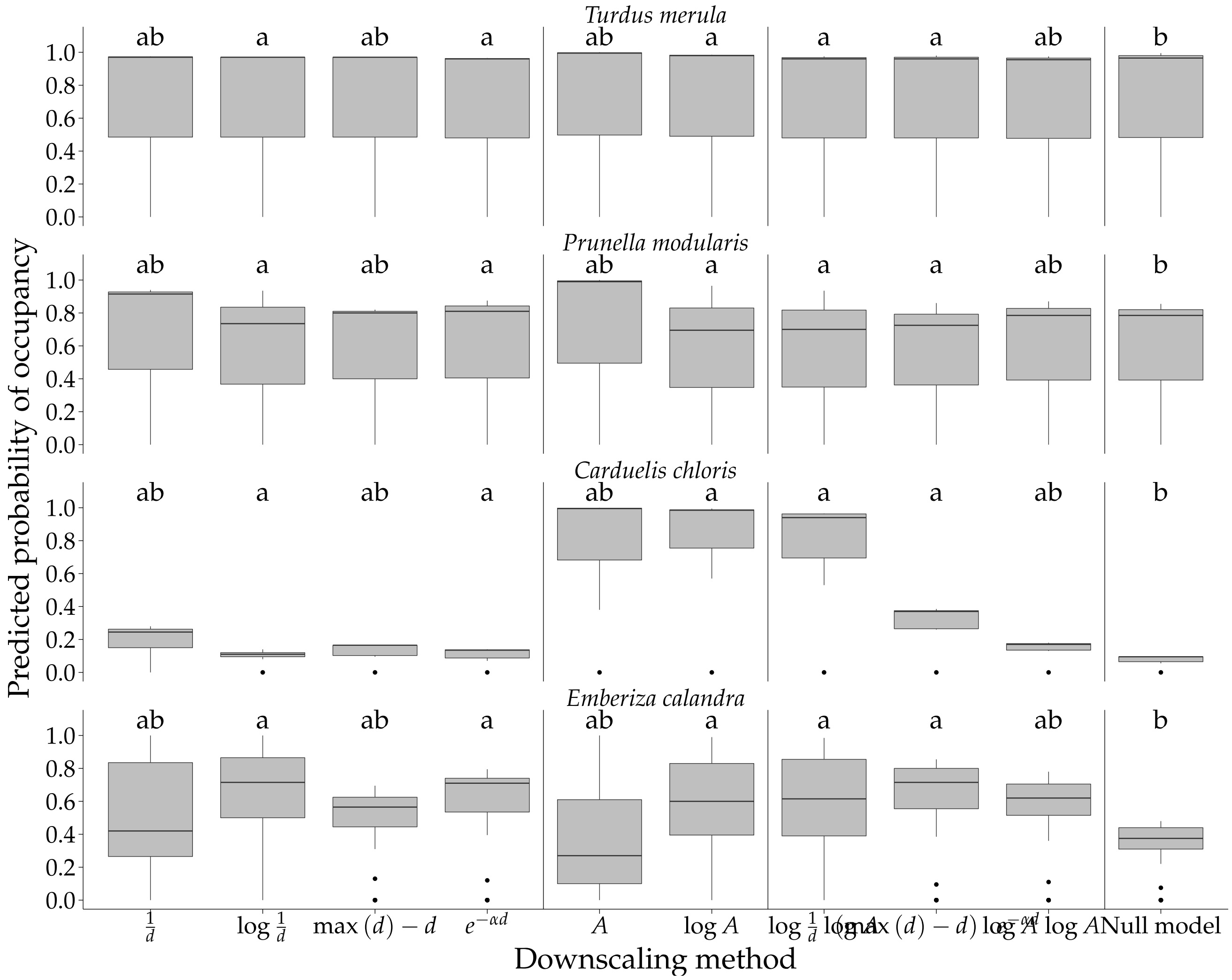

Figure E5: Boxplots showing the probability of occupancy for individual species after the burn-in period (t = 175) for all grid cells which were visited but the species not recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). Vertical lines seperate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

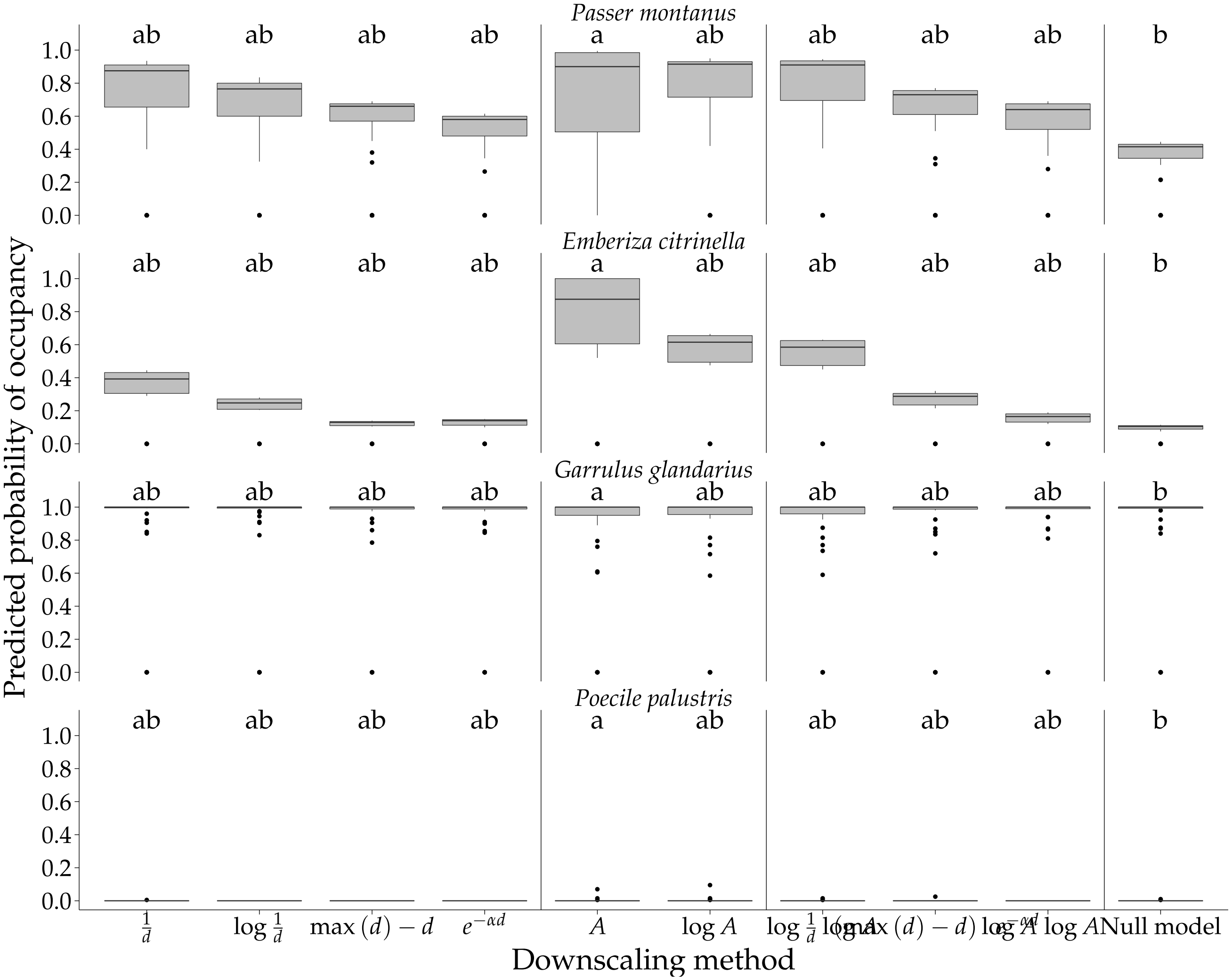

Figure E6: Boxplots showing the probability of occupancy for individual species after the burn-in period (t = 175) for all grid cells which were visited but the species not recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). Vertical lines seperate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling methods, $d$ represents the distance between patches, $A$ patch area, and $\alpha$ the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).