

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 Evolution, 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: laura.jane@graham@gmail.com (Laura J. Graham)

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

25 biodiversity planning within UK local authorities (Lott et al., 2006). It allows
26 large quantities of occurrence data to be collected at larger spatial and temporal
27 extents than would be feasible through individual field studies. Species-level
28 data are available from local recording schemes, as well as from large reposi-
29 tories, examples of which are Global Biodiversity Information Facility globally
30 (Global Biodiversity Information Facility, 2014) and National Biodiversity Net-
31 work (NBN) Gateway in the UK (National Biodiversity Network, 2014). There
32 are, however, some problems with volunteer-collected data. There are con-
33 cerns about the quality of data collected by non-specialists (see Bird et al. 2014
34 and Isaac et al. 2014 for discussions of these issues and some of the potential
35 solutions). Specific to the IFM, there are two prevalent issues in data avail-
36 able from major citizen science schemes. First, the data are typically available
37 at grid-square level (for example the finest resolution of data available on the
38 NBN Gateway is 100 x 100m, but a greater coverage of data is available at the 2
39 km resolution), whereas the IFM requires information about patch-level occu-
40 pancies (Hanski, 1999). Although some patches may cover a 100 x 100m grid
41 cell, in a highly fragmented landscape such as an urban or heavily managed
42 landscape, the patches are likely to be smaller than this. Additionally, even
43 if the sizes match, the grid cell boundaries are unlikely to be coincident with
44 the patch boundaries. Secondly, the data tend to be presented as species lists,
45 which only give information about species' presences. In a study by Moilanen
46 (2002), it was found that false absences can bias parameter estimates in all com-
47 ponents of the model; therefore, the higher the confidence in true absence, the
48 better fitting the model will be (but see Kéry et al. 2010). If volunteer-collected
49 data are to be useful for the IFM, or stochastic patch occupancy models more
50 widely, methods are needed for downscaling these data to patch level, and
51 for determining confidence in species' absences. Here we present methods to
52 address both of these issues.

53 Current approaches to downscaling atlas data for species tend to fall into
54 three categories: expert opinion, empirical models and spatial processes (Araújo
55 et al., 2005; Keil et al., 2013). The expert opinion approach typically involves

56 matching species to suitable land-cover classes. For a wide range of species,
57 however, the species–habitat relationship is not well known, and so this method
58 can only be applied to well-studied species (Araújo et al., 2005). This approach
59 also operates on the assumption that any suitable habitat is occupied by the
60 species, which is ecologically unrealistic because species range filling is gen-
61 erally discontinuous (Rapoport, 1982). The empirical approach uses environ-
62 mental variables such as climate, land-cover classes and normalised vegetation
63 difference indices to predict species’ occurrences (see Araújo et al. 2005 for an
64 example using general additive modelling and Keil et al. 2013 for one using
65 hierarchical Bayesian modelling). These methods are particularly appropriate
66 for broad-scale species mapping, for example national and continental studies
67 (Stockwell and Peterson, 2002). The spatial-processes approach divides coarse
68 grid cells into finer grid cells and uses statistical point-and-cluster processes
69 to randomly select cells at a fine grain. The environmental attributes from
70 these finer grid cells are used as predictors for species’ presences and absences.
71 These methods assume that all fine-grain grid cells within a coarse-grain cell of
72 known occupancy contain suitable habitat. To overcome this problem, Niamir
73 et al. (2011) proposed a method which combines expert knowledge and point
74 sampling.

75 The empirical and spatial-processes approaches to downscaling species at-
76 las data use environmental variables as predictors, drawing from species’ dis-
77 tribution modelling. The extent and grain of interest for a city-level biodi-
78 versity plan tend to be much smaller than in studies taking a species’ dis-
79 tribution modelling approach to downscaling, and the environmental gradi-
80 ents sampled therefore much narrower but with greater habitat heterogene-
81 ity. With their very limited variation, environmental factors such as climate
82 are not useful for predicting species’ occurrence at smaller extents. Instead,
83 land cover tends to be the most reliable predictor, and thus the empirical and
84 spatial-processes approaches collapse to species–habitat associations at the city
85 level and individual patch characteristics are likely to become important. The
86 method we outline below applies a combination of expert knowledge (through

87 literature review) and spatial factors. The method involves attributing species'
88 presence to a suitable habitat patch based on its spatial characteristics and
89 known ecological patterns (species–area relationships and the distance decay
90 of community similarity).

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

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

113 2. Methods

114 2.1. Study area

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

125 2.2. Citizen science species data

126 Data for bird species were provided by Nottinghamshire Birdwatchers. These
127 data comprised 12,110 records of 24 species in 44 2 km grid cells for the years
128 1998–2011. Bat species data were provided by Nottinghamshire Bat Group
129 and further records were downloaded from NBN Gateway. The combined
130 bat datasets, once duplicates had been removed, contained 421 records for 10
131 species in 109 1 km grid cells from 1983–2013. Amphibian and reptile data
132 were downloaded from NBN gateway. There were a total of 1116 records for 11
133 species in 56 2 km grid cells for the period 1984–2012. All downloads from the
134 NBN Gateway were performed using the R package ‘rnbn’ (Ball and August,
135 2013). The full list of data providers is supplied in Supplementary Materials,
136 Appendix A (Table A3).

137 2.3. Species–habitat associations and dispersal

138 It is common practice to use indicator species in biodiversity assessments
139 (Caro and O’Doherty, 2013) because constraints on time, funding and taxo-
140 nomic knowledge make collection of data on all species unfeasible (Blair, 1999;

141 Margules et al., 2002). We selected indicator species for modelling with the IFM
142 where sufficient data and information about habitat requirements and disper-
143 sal were available. We ensured that species with a range of habitat specialisms
144 and dispersal abilities were chosen, to maximise the species' validity as indica-
145 tors.

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

158 We created a lookup between species and LCM 2007 land-cover types for
159 birds based on Wernham et al. (2002) and Holden and Cleaves (2006) and am-
160 phibians based on Beebee and Griffiths (2000). Minimum area requirements
161 for most species were taken from Hinsley et al. (1995), a study based on 151
162 woods in eastern England, with minimum patch size 0.02 ha. Not all species
163 were included in Hinsley et al. (1995), so minimum area requirements for *E. ca-*
164 *landra* were taken from Meyer et al. (2007) and *P. palustris* from Broughton et al.
165 (2006). Minimum area requirements for the amphibian species were not found
166 from a review of the literature, so an arbitrary value of 0.02 ha was chosen.

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

172 (2008), a study on amphibian metapopulations, where the figure given was
173 based on expert opinion. The range of dispersal distances is 700 m (*B. bufo*) to
174 41.2 km (*S. atricapilla*)

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

177 2.4. *Habitat data*

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

191 2.5. *Quantifying uncertainty in species' absence*

192 Measures of uncertainty in raw species distribution data should be mapped
193 and made explicit as part of good practice (Rocchini et al., 2011). Not only
194 can this provide a spatial account of the potential biases in the data, but it
195 can also aid predictions of total species richness (Soberón M. and Llorente B.,
196 1993; Colwell and Coddington, 1994) or identify well sampled areas for use in
197 species distribution modelling (Hortal et al., 2004; Hortal and Lobo, 2005). We
198 applied the following method to attach a confidence of true absences to each
199 grid cell, and used this information to select areas for fitting the IFM. We fol-
200 lowed established methods which fit smoothed species-accumulation curves

201 to a Clench function, and identified the slope of this curve at the position of the
202 last record (Soberón M. and Llorente B., 1993; Colwell and Coddington, 1994;
203 Hortal et al., 2004; Hortal and Lobo, 2005). This slope gives a measure of rate of
204 species accumulation with additional sampling effort. We translated this value
205 to a measure of confidence in species absence, as outlined below.

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

$$S_r = \frac{ar}{1 + br} \quad (1)$$

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

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

219 The confidence in true species' absences was defined as the proportion $1 -$
220 $\frac{dS_r}{dr}$. We used a threshold level of 95% confidence as the criterion for selecting
221 grid cells suitable for model fitting. We calculated slopes for each of the species
222 groups for each grid cell at the appropriate resolution (2 km for birds, and
223 amphibians and reptiles, 1 km for bats).

224 2.6. Downscaling species atlas data

225 The data detailed in Section 2.2 are available at 2 km level for birds, and
226 reptiles and amphibians and at 1 km level for bats. To fit the IFM using these

227 data, it is necessary to downscale them to patch level. Here, we outline and
228 demonstrate a method that first identifies a patch-occupancy level within each
229 grid cell, and then uses a suite of methods based on ecological theory to popu-
230 late that proportion of patches.

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

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

251 For each grid cell, we assigned species to patches by weighted sampling.
252 The number of patches sampled from each grid cell was proportional to the
253 landscape occupancy level. For example if a species is present in 50% of the
254 well-surveyed (i.e. confidence of completeness above 95%; Figure 1) 2 km cells
255 at the landscape level, species would be assigned to 50% of the patches inside
256 the grid cell. The weighting was the proportion of the patch falling in that
257 particular grid cell multiplied by a weighting based on either area, distance or

258 both. We tested ten different methods of weighting for the sampling. These
259 fall into three categories: distance weighting only, area weighting only, and
260 combined distance and area weighting. Four distance measures were calcu-
261 lated, where d is distance: based on inverse distance weighting, $\frac{1}{d}$ and $\log \frac{1}{d}$;
262 for a linear relationship $\max(d) - d$; to account for dispersal $e^{-\alpha d}$. The last
263 measure in the list is taken from the incidence function model (see section 2.7).
264 Two area measures were tested (A represents area): A and $\log A$. $\log A$ was
265 tested because species richness and area are usually linearly related when both
266 are log-transformed (Gleason and Jan, 1921; Preston, 1962), and also this allows
267 for a threshold at which the patch area becomes large enough that the size is no
268 longer important. Given the likelihood that both the area and distance relation-
269 ships are important, for three of the distance measures ($\log \frac{1}{d}$, $\max(d) - d$, $e^{-\alpha d}$)
270 sampling was further weighted by $\log A$. So that the measures could be used
271 as probabilities, we standardised these to the range $\{0, 1\}$. As a null model, we
272 employed a downscaling method where the only weighting applied was the
273 proportion of the patch falling in that particular grid cell. This means that the
274 probability of a patch being selected depends only on how much of the patch
275 lies in the grid cell and not on its size or isolation.

276 2.7. Incidence Function Model

277 The incidence function model (IFM), a stochastic patch occupancy model
278 developed by Hanski (1994), allows long-term predictions of metapopulation
279 persistence in a network of habitat patches to be made through estimation of
280 colonisation and extinction rates. We followed methods outlined by Oksanen
281 (2004) which are based on Hanski (1994). A full description of the IFM is given
282 in the Supplementary Material, Appendix B.

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

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

294 2.8. Comparing downscaling methods

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

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

316 As an additional reality check of the model, we calculated the median pro-
317 portion of occupied habitat for each downscaling method at 175 time-steps (af-

318 ter the burn-in period) and compared against the national occupancy level (%
319 of 10 x 10 km grid cells occupied). Specifically, we identified whether nation-
320 ally abundant species appear locally abundant after downscaling and mod-
321 elling, and similarly for nationally rare species. We quantified the relationship
322 between the simulated occupancy proportions and national occupancy pro-
323 portions by fitting a linear regression model with national proportions as the
324 dependent variable and simulated proportions as the independent variable.
325 Although local proportions of species' occupancy do not necessarily reflect na-
326 tional ones, this test is a useful secondary check used in conjunction with the
327 above validation.

328 2.9. *Sensitivity analysis of the occupancy assumption*

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

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

358 **3. Results**

359 *3.1. Mapping uncertainty in presence data*

360 After fitting these curves to a Clench function and calculating the slope at
361 the point of the last record, we identified 36% of cells were well-surveyed birds,
362 2% for bats and 52% for amphibians and reptiles (Figure 1).

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

370 *3.2. Comparison of downscaling methods*

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

376 2006–2007, and the amphibian species data for 1996–1997 were used for model
377 evaluation.

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

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

396 An overall significant difference between model performance based on the
397 downscaling methods was detected using a Kruskal-Wallis test (Chi-square =
398 143.52, df = 9, $P < 0.001$). These results show that the area and, to some ex-
399 tent, log(area) weighted downscaling methods are most reliable in predicting
400 species' presence overall.

401 Figure 2 b) shows the predicted probabilities of occurrence for grid cells
402 where the species has not been recorded as present. It should be noted that
403 these should not be considered as confirmed absences due to the ad hoc nature
404 of citizen science biological recording (see Discussion). The analysis is shown
405 by species in Supplementary Materials, Appendix D (Figures D3 and D4). A
406 significant difference between model predictions based on starting conditions

407 from each of the downscaling methods was detected here also (Chi-square =
408 53.4, $df = 9$, $p < 0.001$). Here, however, it is less clear which downscaling
409 method is best at predicting the presumed absences, but the better methods
410 for predicting presences are the worst for predicting absences. This suggests
411 that when parameterising the model based on data downscaled by the area or
412 log(area) weighted method, the IFM over predicts species' occupancy.

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

426 3.3. Sensitivity analysis of the occupancy starting condition

427 We performed a sensitivity analysis to evaluate whether changes in the ini-
428 tial starting occupancy had an impact on the measures of landscape persistence
429 derived from the IFM results. A post-hoc multiple comparison test was used
430 to determine whether the differences in the measure between occupancy start
431 conditions were significant. Many were significant, but overall a change in
432 10% of occupancy starting condition resulted in a change of less than 10% in
433 the resulting metapopulation measure (see Table 2). Minimum occupancy %
434 seemed to be the measure most sensitive to starting condition, with the largest
435 differences between the value based on high vs low occupancy being 18.97%
436 for *T. merula* and 22.27% for *R. temporaria*.

437 4. Discussion

438 For the IFM to have wide use in conservation management and planning,
439 it is important that accessible sources of data are available. Our results show
440 that citizen science data offer the potential to meet the data requirements of
441 the IFM, but with some important caveats. Citizen science and collections data
442 can be beneficial for studies at large spatial and temporal extents but are often
443 sparse and geographically biased (Funk and Richardson, 2002; Boakes et al.,
444 2010) or at a coarser spatial resolution than required for planning purposes
445 (Araújo et al., 2005; Keil et al., 2013). The data available for the study area may
446 not be a complete reflection of species' occupancies (Figure 1). Despite this,
447 there are sufficient records for both bird and amphibian species to fit the IFM.

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

458 The downscaling method we present herein deals with the issue of the data
459 being available at a coarser resolution than required. Of the downscaling meth-
460 ods tested, we found that the methods which involved weighting by A and
461 $\log A$ were both the closest match to the known presences in the evaluation
462 data set and the national occupancy proportions. Based on the above, weight-
463 ing by either A or $\log A$ would be the most appropriate method for downscal-
464 ing species' distribution data for use in the IFM. These results suggest that the
465 species–area relationship has more influence than the distance decay of simi-
466 larity at this spatial scale. The shape of the distance decay relationship depends

467 on the sampling extent and grain: when sampled at a large extent and grain,
468 the relationship is usually exponential; whereas when sampled at a small ex-
469 tent and grain with limited environmental gradient, the distance decay gener-
470 ally forms a power-law relationship (Nekola and McGill, 2014). Of the remain-
471 ing methods, the closest to national occupancies was the method weighted by
472 $\log \frac{1}{A} \log A$, which fits with an exponential relationship for distance decay. It is
473 possible that for our study area, the slope of the power-law relationship is very
474 shallow due to the limited environmental gradient, hence why those methods
475 which weight by patch area alone fit national occupancy levels more closely.

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

492 Refining species–habitat associations may improve the performance of the
493 downscaling methods. It is possible that datasets whose land-cover classes are
494 broad and non-specific (e.g. LCM 2007, as used here) are not entirely sufficient
495 to identify suitable habitat patches and need to be combined with other data
496 that offer further description of habitat types (e.g. Ordnance Survey Master
497 Map for this study area). Species also depend on structure as well as type of

498 habitat, for example Broughton et al. (2006) found an effect of canopy height on
499 marsh tit occupancy. Incorporating remotely sensed data will allow for habitat
500 structure, and perhaps quality, to be considered. We have used LCM 2007 be-
501 cause of its availability (both to us, and to planners who may wish to use this
502 method), however remote sensing data are increasing in quantity and quality,
503 and pushes to make these data open access will mean habitat data are more
504 accessible and accurate (Sutherland et al., 2014; Turner et al., 2015). For the
505 marsh tit, which appears to be consistently under-predicted by all methods,
506 it is possible that improving the habitat association will improve prediction.
507 It is possible that there are no issues with the habitat association for the corn
508 bunting and they are more prevalent in Nottingham than nationally. Notting-
509 ham has quite a high proportion of arable land compared to some other cities
510 (see Table A1).

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

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

529 predictions from the downscaling methods outlined above.

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

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

544 **Acknowledgements**

545 We thank Joaquín Hortal for very helpful discussion of the methods for
546 maps of uncertainty. Thanks also to Rob Johnson and Peter Acton at Notting-
547 hamshire Biological and Geological Records Centre, Michael Walker at Not-
548 tinghamshire Wildlife Trust and Andy Hall from Not-tinghamshire Birdwatch-
549 ers for supplying species data, and giving advice on the use of these data. Fi-
550 nally, we thank Adam Algar and the BEER (Biodiversity, Ecology and Evolu-
551 tion Research) Group at the University of Nottingham for helpful comments on
552 an earlier version of this manuscript. This research was funded by the ESRC.

553 **References**

- 554 Araújo, M.B., Thuiller, W., Williams, P.H., Reginster, I., 2005. Downscaling
555 European species atlas distributions to a finer resolution: Implications for
556 conservation planning. *Global Ecology and Biogeography* 14, 17–30. doi:10.
557 1111/j.1466-822X.2004.00128.x.
- 558 Ball, S., August, T., 2013. rbn: Access NBN data. URL: [https://github.com/
559 JNCC-UK/rbn](https://github.com/JNCC-UK/rbn).
- 560 Barwell, L.J., Azaele, S., Kunin, W.E., Isaac, N.J.B., 2014. Can coarse-grain pat-
561 terns in insect atlas data predict local occupancy? *Diversity and Distribu-
562 tions* 20, 895–907. doi:10.1111/ddi.12203.
- 563 Beebee, T.J.C., Griffiths, R.a., 2000. *Amphibians and reptiles: a natural history*
564 *of the British herpetofauna*. Harper Collins, London.
- 565 Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J.,
566 Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., Pecl,
567 G.T., Barrett, N., Frusher, S., 2014. Statistical solutions for error and bias
568 in global citizen science datasets. *Biological Conservation* 173, 144–154.
569 doi:10.1016/j.biocon.2013.07.037.
- 570 Bivand, R., Keitt, T., Rowlingson, B., 2015. The rgdal Package: Bindings for
571 the geospatial data abstraction library. URL: [http://cran.r-project.org/
572 package=rgdal](http://cran.r-project.org/package=rgdal).
- 573 Bivand, R., Rundel, C., 2013. rgeos: Interface to Geometry En-
574 gine - Open Source (GEOS). R package version 0.3-2. URL:
575 [http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:
576 Interface+to+Geometry+Engine+-+Open+Source+\(GEOS\)#0](http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Interface+to+Geometry+Engine+-+Open+Source+(GEOS)#0).
- 577 Blair, R.B., 1999. Birds and butterflies along an urban gradient: Surrogate taxa
578 for assessing biodiversity? *Ecological Applications* 9, 164–170.

- 579 Boakes, E.H., McGowan, P.J.K., Fuller, R.a., Chang-Qing, D., Clark, N.E.,
580 O'Connor, K., Mace, G.M., 2010. Distorted views of biodiversity: Spatial
581 and temporal bias in species occurrence data. *PLoS Biology* 8, e1000385.
582 doi:10.1371/journal.pbio.1000385.
- 583 Broughton, R.K., Hill, R.A., Bellamy, P.E., Hinsley, S.A., 2010. Dispersal, rang-
584 ing and settling behaviour of marsh tits *Poecile palustris* in a fragmented land-
585 scape in lowland England. *Bird Study* 57, 458–472. doi:10.1080/00063657.
586 2010.489316.
- 587 Broughton, R.K., Hinsley, S.A., Bellamy, P.E., Hill, R.A., Rothery, P., 2006.
588 Marsh tit *Poecile palustris* territories in a British broad-leaved wood. *Ibis* 148,
589 744–752. doi:10.1111/j.1474-919X.2006.00583.x.
- 590 Bulman, C.R., Wilson, R.J., Holt, A.R., Bravo, L.G., Early, R.I., Warren, M.S.,
591 Thomas, C.D., 2007. Minimum viable metapopulation size, extinction debt,
592 and the conservation of a declining species. *Ecological Applications* 17,
593 1460–1473. doi:10.1890/06-1032.1.
- 594 Caro, T.M., O'Doherty, G., 2013. On the use of surrogate species in conservation
595 biology. *Conservation Biology* 13, 805–814.
- 596 Colwell, R.K., Coddington, J.a., 1994. Estimating terrestrial biodiversity
597 through extrapolation. *Philosophical Transactions of the Royal Society of*
598 *London Series B-Biological Sciences* 345, 101–118.
- 599 Devictor, V., Whittaker, R.J., Beltrame, C., 2010. Beyond scarcity: Citizen sci-
600 ence programmes as useful tools for conservation biogeography. *Diversity*
601 *and Distributions* 16, 354–362. doi:10.1111/j.1472-4642.2009.00615.x.
- 602 Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips,
603 T., Purcell, K., 2012. The current state of citizen science as a tool for ecological
604 research and public engagement. *Frontiers in Ecology and the Environment*
605 10, 291–297. doi:10.1890/110236.

- 606 Dickinson, J.L., Zuckerberg, B., Bonter, D.N., 2010. Citizen science as
607 an ecological research tool: Challenges and benefits. *Annual Re-*
608 *view of Ecology, Evolution, and Systematics* 41, 149–172. doi:10.1146/
609 *annurev-ecolsys-102209-144636*.
- 610 Dolrenry, S., Stenglein, J., Hazzah, L., Lutz, R.S., Frank, L., 2014. A metapop-
611 ulation approach to African lion (*Panthera leo*) conservation. *PLoS ONE* 9,
612 e88081. doi:10.1371/journal.pone.0088081.
- 613 Etienne, R.S., ter Braak, C.J.F., Vos, C.C., 2004. Application of Stochastic
614 Patch Occupancy Models to Real Metapopulations, in: Hanski, I., Gaggiotti,
615 O.E. (Eds.), *Ecology, Genetics and Evolution of Metapopulations*. Academic
616 Press, Burlington, MA, USA, pp. 105–132. doi:10.1016/B978-012323448-3/
617 50007-6.
- 618 Funk, V.A., Richardson, K.S., 2002. Systematic data in biodiversity stud-
619 ies: use it or lose it. *Systematic biology* 51, 303–316. doi:10.1080/
620 10635150252899789.
- 621 Gibbons, D.W., Reid, J.B., Chapman, R.a., 1993. *The new atlas of breeding birds*
622 *in Britain and Ireland:1988-1991*. T. & A.D. Poyser, London.
- 623 Gilioli, G., Bodini, A., Baumgärtner, J., Weidmann, P., Hartmann, J., 2008. A
624 novel approach based on information theory to rank conservation strategies:
625 An application to amphibian metapopulations. *Animal Conservation* 11,
626 453–462. doi:10.1111/j.1469-1795.2008.00190.x.
- 627 Giraudoux, P., 2014. *pgirmess: Data analysis in ecology*. URL: <http://cran.r-project.org/package=pgirmess>.
- 629 Gleason, H.a., Jan, N., 1921. Species and Area. *Journal of Ecology* 9, 95–99.
- 630 Global Biodiversity Information Facility, 2014. GBIF.org. URL: <http://www.gbif.org>.
- 631

- 632 Graham, L.J., Bailey, J.J., Algar, A.C., Field, R., 2014. Where next for macroecol-
633 ogy: citizen macroecology? *Frontiers of Biogeography* 6, 16–19.
- 634 Graham, L.J., Field, R., Haines-Young, R., in press. Strengthening Urban Land-
635 scape Planning: a metapopulation modelling framework, in: Francis, R.
636 (Ed.), *Urban Landscape Ecology: Science, Policy and Practice*. Routledge.
- 637 Grimm, V., Lorek, H., Finke, J., Koester, F., Malachinski, M., Sonnenschein,
638 M., Moilanen, A., Storch, I., Singer, A., Wissel, C., Frank, K., 2004. META-
639 X: Generic software for metapopulation viability analysis. *Biodiversity and*
640 *Conservation* 13, 165–188. doi:10.1023/B:BI0C.0000004317.42949.f7.
- 641 Halley, J.M., Hartley, S., Kallimanis, A.S., Kunin, W.E., Lennon, J.J., Sgardelis,
642 S.P., 2004. Uses and abuses of fractal method in ecology. *Ecology Letters* 7,
643 254–271. doi:10.1111/j.1461-0248.2004.00568.x.
- 644 Hanski, I., 1994. A practical model of metapopulation dynamics. *Journal of*
645 *Animal Ecology* 63, 151–162. doi:10.2307/5591.
- 646 Hanski, I., 1999. *Metapopulation ecology*. Oxford University Press, Oxford.
647 doi:10.1002/9780470015902.a0021905.
- 648 Hanski, I., Moilanen, a., Pakkala, T., Kuussaari, M., 1996. The quantitative inci-
649 dence function model and persistence of an endangered butterfly metapopu-
650 lation. *Conservation Biology* 10, 578–590. doi:10.1046/j.1523-1739.1996.
651 10020578.x.
- 652 Heard, G.W., McCarthy, M.A., Scroggie, M.P., Baumgartner, J.B., Parris, K.M.,
653 2013. A Bayesian model of metapopulation viability, with application to an
654 endangered amphibian. *Diversity and Distributions* 19, 555–566. doi:10.
655 1111/ddi.12052.
- 656 Hinsley, S.A., Bellamy, P.E., Newton, I., Sparks, T.H., 1995. Habitat and land-
657 scape factors influencing the presence of individual breeding bird species in
658 woodland fragments. *Journal of Avian Biology* 26, 94–104. doi:10.2307/
659 3677057.

- 660 Holden, P., Cleeves, T., 2006. RSPB Handbook of British Birds. 2nd ed., Christo-
661 pher Helm, London.
- 662 Hortal, J., Garcia-Pereira, P., García-Barros, E., 2004. Butterfly species richness
663 in mainland Portugal: Predictive models of geographic distribution patterns.
664 *Ecography* 27, 68–82. doi:10.1111/j.0906-7590.2004.03635.x.
- 665 Hortal, J., Lobo, J.M., 2005. An ED-based protocol for optimal sampling of
666 biodiversity. *Biodiversity and Conservation* 14, 2913–2947. doi:10.1007/
667 s10531-004-0224-z.
- 668 Isaac, N.J.B., Pocock, M.J.O., 2015. Bias and information in biological records.
669 *Biological Journal of the Linnean Society* 115, 522–531. doi:10.1111/bij.
670 12517.
- 671 Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P., Roy, D.B., 2014.
672 Statistics for citizen science: extracting signals of change from noisy ecolog-
673 ical data. *Methods in Ecology and Evolution*, 1052–1060 URL: <http://doi.wiley.com/10.1111/2041-210X.12254>, doi:10.1111/2041-210X.12254.
- 674
- 675 Keil, P., Belmaker, J., Wilson, A.M., Unitt, P., Jetz, W., 2013. Downscaling of
676 species distribution models: a hierarchical approach. *Methods in Ecology
677 and Evolution* 4, 82–94. doi:10.1111/j.2041-210x.2012.00264.x.
- 678 Kéry, M., Royle, J.A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., Zbinden,
679 N., 2010. Site-occupancy distribution modeling to correct population-trend
680 estimates derived from opportunistic observations. *Conservation Biology*
681 24, 1388–1397. doi:10.1111/j.1523-1739.2010.01479.x.
- 682 Kunin, W.E., 1998. Extrapolating species abundance across spatial scales. *Sci-
683 ence* 281, 1513–1515. doi:10.1126/science.281.5382.1513.
- 684 Lindenmayer, D.B., McCarthy, M.A., Pope, M.L., 1999. Arboreal marsupial
685 incidence in eucalypt patches in south-eastern Australia: a test of Hanski's
686 incidence function metapopulation model for patch occupancy. *Oikos* 84,
687 99–109. doi:10.2307/3546870.

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

691 Lott, D., Moon, S., Cooper, A., 2006. Biodiversity needs for Local Authorities
692 and National Park Authorities. Technical Report. Association of Local Gov-
693 ernment Ecologists. URL: <http://www.alge.org.uk/publications/>.

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

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

700 Meyer, B.C., Mammen, K., Grabaum, R., 2007. A spatially explicit model for
701 integrating species assessments into landscape planning as exemplified by
702 the Corn Bunting (*Emberiza calandra*). *Journal for Nature Conservation* 15,
703 94–108. doi:10.1016/j.jnc.2007.01.001.

704 Moilanen, A., 2002. Implications of empirical data quality to metapopulation
705 model parameter estimation and application. *Oikos* 96, 516–530. doi:10.
706 1034/j.1600-0706.2002.960313.x.

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

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

712

713 Nekola, J.C., McGill, B.J., 2014. Scale dependency in the functional form of
714 the distance decay relationship. *Ecography* 37, 309–320. doi:10.1111/j.
715 1600-0587.2013.00407.x.

- 716 Nekola, J.C., White, P.S., 1999. The distance decay of similarity in biogeog-
717 raphy and ecology. *Journal of Biogeography* 26, 867–878. doi:10.1046/j.
718 1365-2699.1999.00305.x.
- 719 Niamir, A., Skidmore, A.K., Toxopeus, A.G., Muñoz, A.R., Real, R., 2011. Fi-
720 neness atlas data for species distribution models. *Diversity and Distribu-*
721 *tions* 17, 1173–1185. doi:10.1111/j.1472-4642.2011.00793.x.
- 722 Oksanen, J., 2004. Incidence Function Model in R. URL: [http://cc.oulu.fi/
723 \\$\sim\\$jarioksa/opetus/openmeta/metafit.pdf](http://cc.oulu.fi/\simjarioksa/opetus/openmeta/metafit.pdf).
- 724 Opdam, P., Foppen, R.P.B., Vos, C.C., 2002. Bridging the gap between ecology
725 and spatial planning in landscape ecology. *Landscape Ecology* 16, 767–779.
- 726 Opdam, P., Verboom, J., Pouwels, R., 2003. Landscape cohesion: An index for
727 the conservation potential of landscapes for biodiversity. *Landscape Ecology*
728 18, 113–126. doi:10.1023/A:1024429715253.
- 729 Paradis, E., Baillie, S.R., Sutherland, W.J., Gregory, R.D., 1998. Patterns of natal
730 and breeding dispersal in birds. *Journal of Animal Ecology* 67, 518–536.
731 doi:10.1046/j.1365-2656.1998.00215.x.
- 732 Preston, F.W., 1962. The Canonical Distribution of Commonness and Rarity:
733 Part I. *Ecology* 43, 185–215. URL: [http://www.esajournals.org/doi/abs/
734 10.2307/1931976](http://www.esajournals.org/doi/abs/10.2307/1931976), doi:10.2307/1931976.
- 735 Rapoport, E., 1982. *Areogeography: Geographical Strategies of Species*. Perg-
736 amon Press, Oxford.
- 737 Ritchie, M.E., Olff, H., 1999. Spatial scaling laws yield a synthetic theory of
738 biodiversity. *Nature* 400, 557–560. doi:10.1038/23010.
- 739 Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta,
740 C., Bacaro, G., Chiarucci, A., 2011. Accounting for uncertainty when map-
741 ping species distributions: The need for maps of ignorance. *Progress in Phys-*
742 *ical Geography* 35, 211–226. doi:10.1177/0309133311399491.

- 743 Siegel, S., Castellan, N.J., 1988. *Nonparametric Statistics for the Behavioral*
744 *Sciences*. 2nd ed., McGraw-Hill, New York.
- 745 Silvertown, J., 2009. A new dawn for citizen science. *Trends in ecology &*
746 *evolution* 24, 467–471. doi:10.1016/j.tree.2009.03.017.
- 747 Soberón M., J., Llorente B., J., 1993. The use of species accumulation func-
748 tions for the prediction of species richness. *Conservation Biology* 7, 480–488.
749 doi:10.1046/j.1523-1739.1993.07030480.x.
- 750 Stockwell, D.R.B., Peterson, a.T., 2002. Effects of sample size on accuracy of
751 species distribution models. *Ecological Modelling* 148, 1–13. doi:10.1016/
752 S0304-3800(01)00388-X.
- 753 Sutherland, W.J., Aveling, R., Brooks, T.M., Clout, M., Dicks, L.V., Fellman, L.,
754 Fleishman, E., Gibbons, D.W., Keim, B., Lickorish, F., Monk, K.A., Mortimer,
755 D., Peck, L.S., Pretty, J., Rockström, J., Rodríguez, J.P., Smith, R.K., Spald-
756 ing, M.D., Tonneijck, F.H., Watkinson, A.R., 2014. A horizon scan of global
757 conservation issues for 2014. *Trends in Ecology and Evolution* 29, 15–22.
758 doi:10.1016/j.tree.2013.11.004.
- 759 Tobler, W.R., 1970. A computer movie simulation urban growth in the Detroit
760 region. *Economic Geography* 46, 234–240. doi:10.2307/143141.
- 761 Tulloch, A.I., Possingham, H.P., Joseph, L.N., Szabo, J., Martin, T.G., 2013. Re-
762 alising the full potential of citizen science monitoring programs. *Biological*
763 *Conservation* 165, 128–138. doi:10.1016/j.biocon.2013.05.025.
- 764 Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A.K., Szantoi, Z.,
765 Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koh, L.P., Leimgruber, P.,
766 Taubenboeck, H., Wegmann, M., Wikelski, M., Woodcock, C.E., 2015. Free
767 and open-access satellite data are key to biodiversity conservation. *Biologi-
768 cal Conservation* 182, 173–176. doi:10.1016/j.biocon.2014.11.048.

- 769 Van Teeffelen, A.J.A., Vos, C.C., Opdam, P., 2012. Species in a dynamic world:
770 Consequences of habitat network dynamics on conservation planning. *Bio-*
771 *logical Conservation* 153, 239–253. doi:10.1016/j.biocon.2012.05.001.
- 772 Wernham, C., Siriwardena, G.M., Toms, M., Marchant, J., Clark, J.A., Baillie, S.,
773 2002. *The Migration Atlas: Movements of the Birds of Britian and Ireland*. T.
774 & A.D. Poyser, London.
- 775 Williams, P.H., Margules, C.R., Hilbert, D.W., 2002. Data requirements and
776 data sources for biodiversity priority area selection. *Journal of Biosciences*
777 27, 327–338. doi:10.1007/BF02704963.
- 778 Wright, D.H., Reeves, J.H., 1992. On the meaning and measurement of
779 nestedness of species assemblages. *Oecologia* 92, 416–428. doi:10.1007/
780 BF00317469.

781 **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)
<i>Turdus merula</i>	Blackbird	3.3	1, 2, 3, 4, 5, 6, 7, 8	0.02
<i>Prunella modularis</i>	Dunnock	2.1	1, 2, 3, 4, 5, 6, 7, 8	0.02
<i>Carduelis carduelis</i>	Goldfinch	11.1	3, 5	0.12
<i>Carduelis chloris</i>	Greenfinch	4.2	1, 2, 3	0.25
<i>Muscicapa striata</i>	Spotted flycatcher	12.8	1, 2	0.1
<i>Emberiza calandra</i>	Corn bunting	4	3, 4, 5, 6, 8	2.5
<i>Passer montanus</i>	Tree sparrow	8	1, 2, 3	0.12
<i>Emberiza citrinella</i>	Yellowhammer	8.4	3, 5, 10, 11	0.03
<i>Sylvia atricapilla</i>	Blackcap	41.2	1	0.2
<i>Dendrocopos major</i>	Great spotted woodpecker	16.5	1, 2	0.26
<i>Garrulus glandarius</i>	Jay	3.5	1, 2	0.32
<i>Poecile palustris</i>	Marsh tit	0.885	1	2.1
<i>Rana temporaria</i>	Common frog	1	1, 2, 3, 4, 5, 6, 8, 9, 16	0.02
<i>Bufo bufo</i>	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
<i>Turdus merula</i>	Minimum occupancy %	88.27	85.22	66.25	x	x	x
<i>Turdus merula</i>	Surviving replicates %	100.00	100.00	100.00			
<i>Turdus merula</i>	Occupancy % after burn-in	96.21	94.74	96.37	x	x	x
<i>Prunella modularis</i>	Minimum occupancy %	64.85	63.17	65.60	x	x	x
<i>Prunella modularis</i>	Surviving replicates %	100.00	100.00	100.00			
<i>Prunella modularis</i>	Occupancy % after burn-in	81.61	76.21	85.47	x	x	x
<i>Carduelis chloris</i>	Minimum occupancy %	70.86	71.11	70.00			
<i>Carduelis chloris</i>	Surviving replicates %	100.00	100.00	100.00			
<i>Carduelis chloris</i>	Occupancy % after burn-in	95.36	93.25	94.82	x	x	x
<i>Emberiza calandra</i>	Minimum occupancy %	40.06	41.00	31.91		x	x
<i>Emberiza calandra</i>	Surviving replicates %	100.00	100.00	100.00			
<i>Emberiza calandra</i>	Occupancy % after burn-in	59.24	53.44	61.99	x		x
<i>Passer montanus</i>	Minimum occupancy %	50.38	49.72	34.17	x	x	x
<i>Passer montanus</i>	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
<i>Passer montanus</i>	Occupancy % after burn-in	91.26	89.77	89.39	x	x	x
<i>Emberiza citrinella</i>	Minimum occupancy %	58.13	56.83	54.90		x	x
<i>Emberiza citrinella</i>	Surviving replicates %	100.00	99.50	99.00			
<i>Emberiza citrinella</i>	Occupancy % after burn-in	93.25	91.38	91.49	x		x
<i>Garrulus glandarius</i>	Minimum occupancy %	42.79	40.44	44.77	x	x	x
<i>Garrulus glandarius</i>	Surviving replicates %	100.00	100.00	100.00			
<i>Garrulus glandarius</i>	Occupancy % after burn-in	89.89	89.07	91.15	x	x	x
<i>Poecile palustris</i>	Minimum occupancy %	0.84	0.44	0.41			
<i>Poecile palustris</i>	Surviving replicates %	22.50	12.50	12.00			
<i>Poecile palustris</i>	Occupancy % after burn-in	5.81	4.58	4.56			
<i>Rana temporaria</i>	Minimum occupancy %	56.00	71.53	49.26	x	x	x
<i>Rana temporaria</i>	Surviving replicates %	100.00	100.00	100.00			
<i>Rana temporaria</i>	Occupancy % after burn-in	80.16	85.09	76.02	x		x
<i>Bufo bufo</i>	Minimum occupancy %	6.00	4.69	4.71	x	x	
<i>Bufo bufo</i>	Surviving replicates %	70.00	61.50	59.00			
<i>Bufo bufo</i>	Occupancy % after burn-in	18.29	13.45	17.96	x		x

782 **Figures**

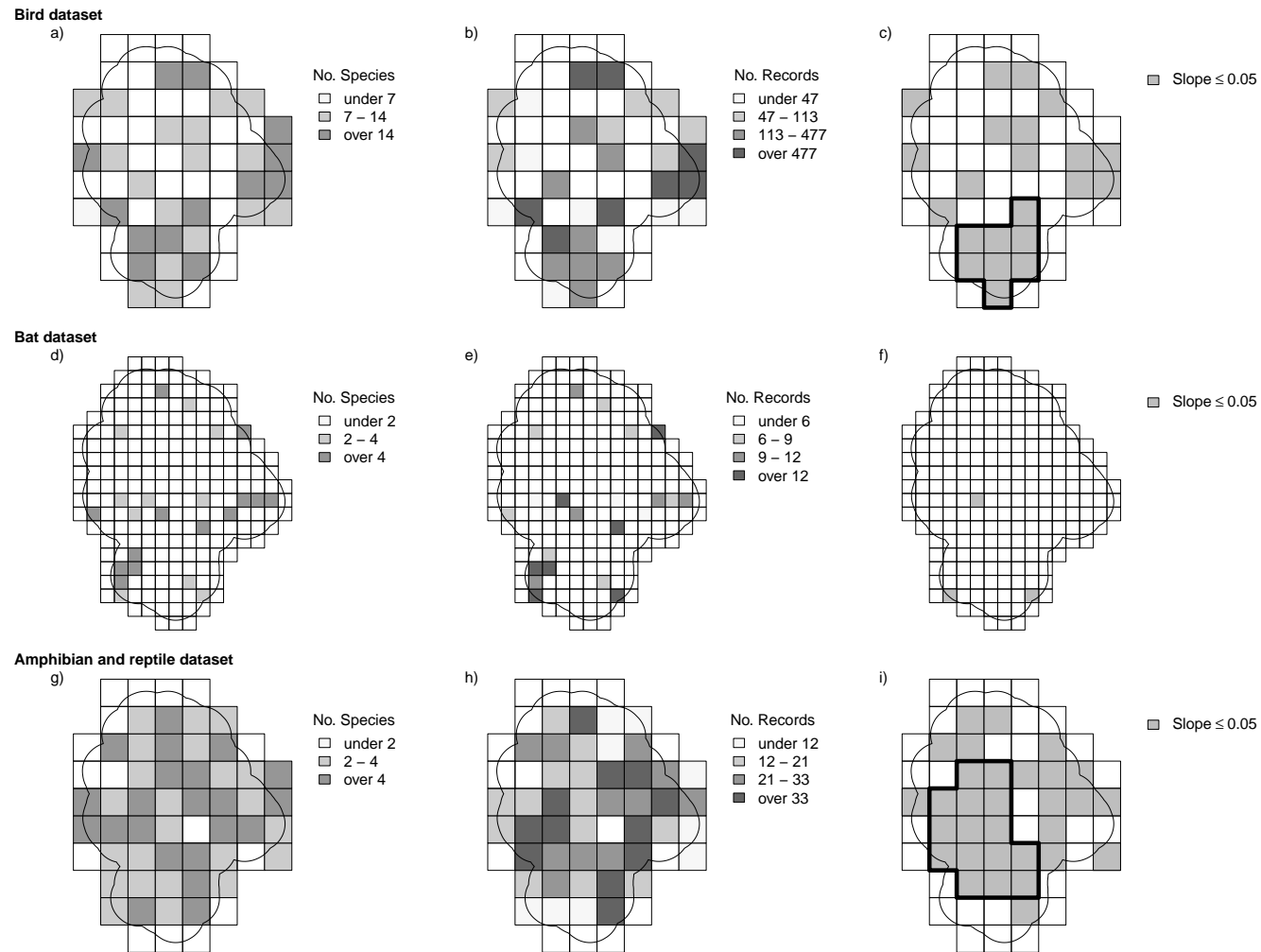


Figure 1: Species data quality maps. Species richness (a, d, g), number of records (b, e, h) and well sampled grids (c, f, i) shown for the three species datasets. The slope in c, f and i is calculated by creating a species-accumulation curve by plotting the record number against the cumulative number of species; 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.

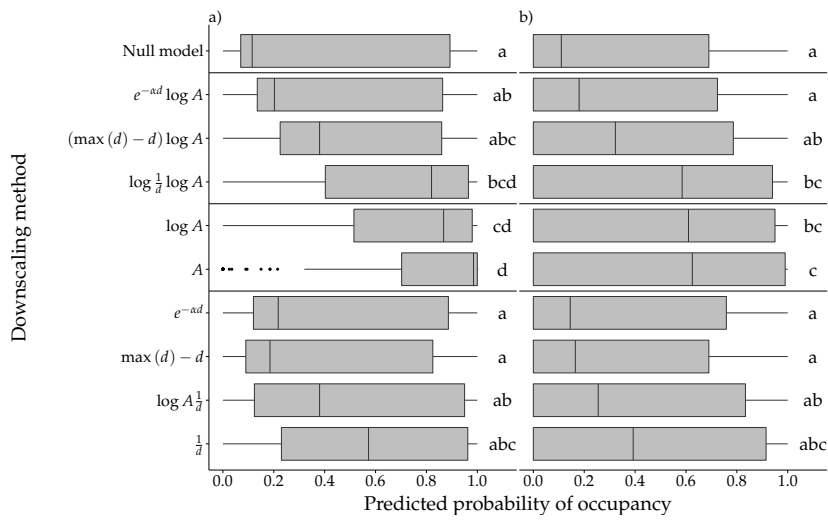


Figure 2: Boxplots showing the probability of occupancy for all species after the burn-in period ($t = 175$ time steps) for all grid cells recorded as a) present and b) absent in the evaluation data. For each species, there are $n = 200$ replicates of IFM simulations. Horizontal lines separate the downscaling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

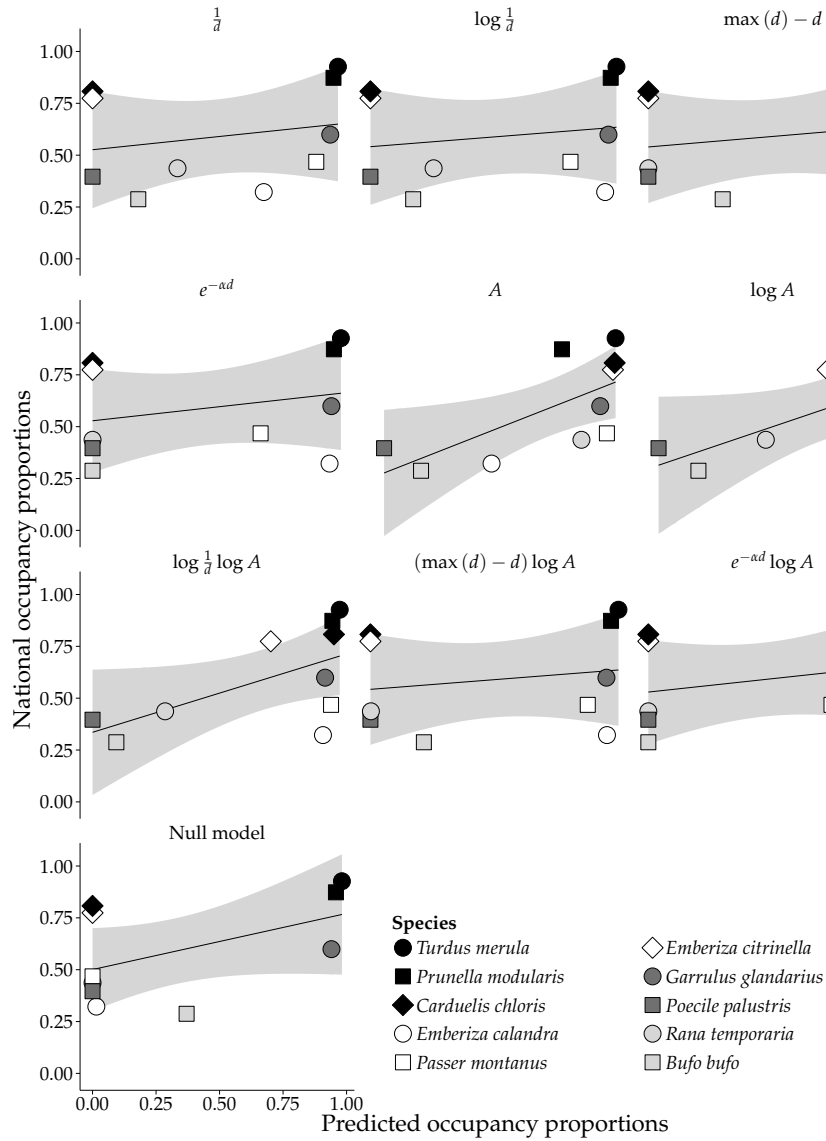


Figure 3: Comparison of mean (of $n = 200$ replicates of IFM simulations) predicted proportion of occupied habitat after the burn-in period ($t = 175$ time steps) against the national proportion of $10 \text{ km} \times 10 \text{ km}$ grid cells occupied. National data are from the 2nd BTO Atlas (Gibbons 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.

783 **Appendix A - Additional study site and data information**

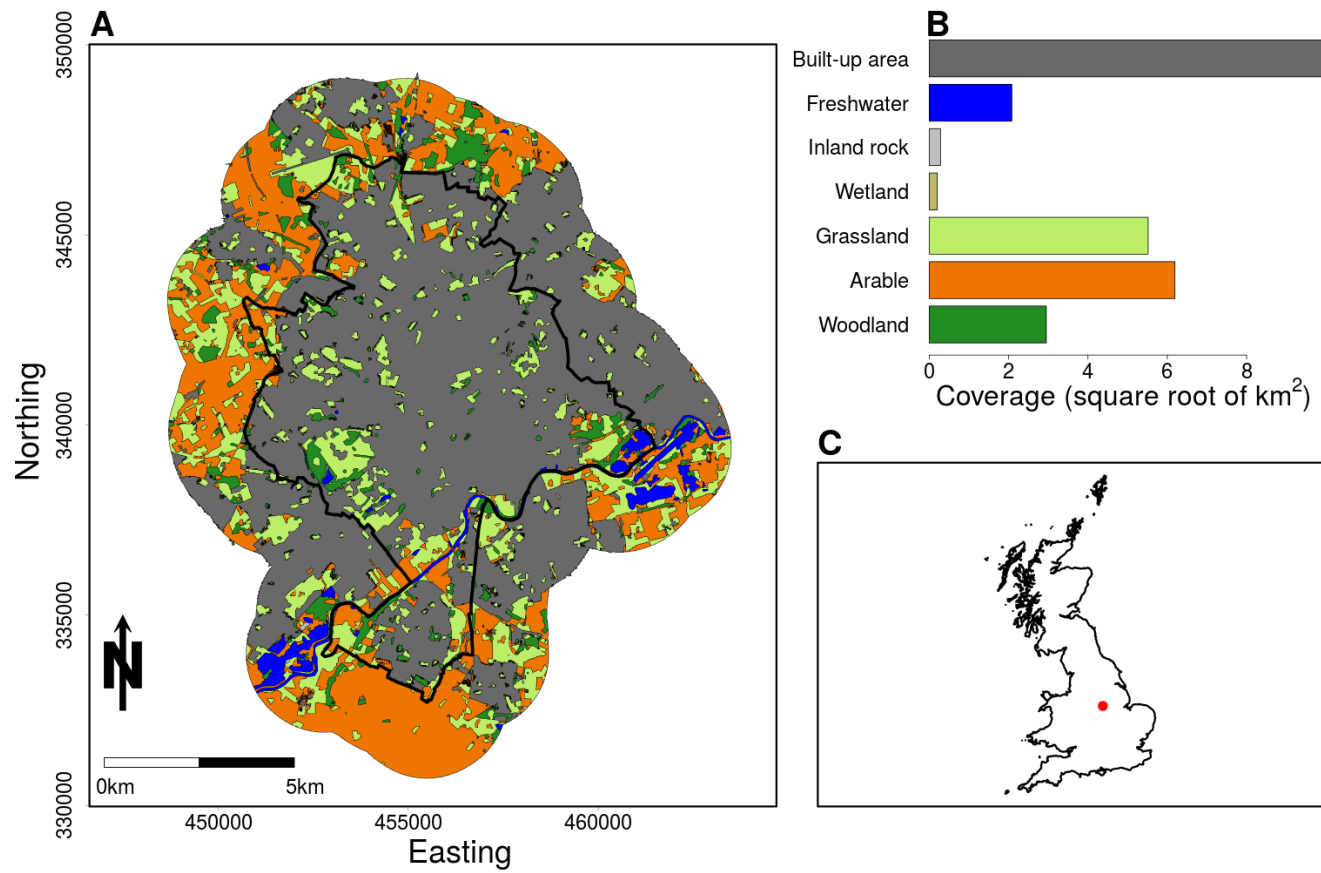


Figure A1: Study site of Nottingham, UK. Nottingham City Administrative Boundary (bold lines) with 2km buffer showing coarse land-cover classes. Inset map shows the location of Nottingham within Great Britain. Land-cover classes based on LCM 2007 (Morton et al., 2011)

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

784 **Appendix B - Additional methods information**

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

$$J_i = \frac{C_i}{C_i + E_i - C_i E_i} \quad (\text{B1})$$

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

$$E_i = \min \left[\frac{u}{A_i^x}, 1 \right] \quad (\text{B2})$$

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

801 The colonisation probability (Equation B3) is a function of patch connectiv-
 802 ity S_i (Equation B4). Species-specific parameter y represents the level of con-
 803 nectivity required by a species to achieve colonisation.

$$C_i = \frac{S_i^2}{S_i^2 + y^2} \quad (\text{B3})$$

804 Finally, connectivity (Equation B4) is a function of the distance from patch
 805 i to patch j (d_{ij}), the occupancy and area of patch j (p_j and A_j respectively) and

806 the species' mean natal dispersal ability, $\frac{1}{\alpha}$. Mean natal dispersal distance is
807 derived from the literature.

$$S_i = \sum_{j=1}^n e^{-\alpha d_{ij}} p_j A_j \quad (\text{B4})$$

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

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

813 Here $\beta_0 = -\log(uy)$ and $\beta_1 = x$. Parameter u is estimated by assuming
814 that for the smallest of all occupied patches, $E_i = 1$, then solving equation B2
815 for u . This value is then substituted into β_0 to solve for y . This follows the
816 method outlined by Oksanen (2004).

817 **Appendix C - IFM outputs**

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

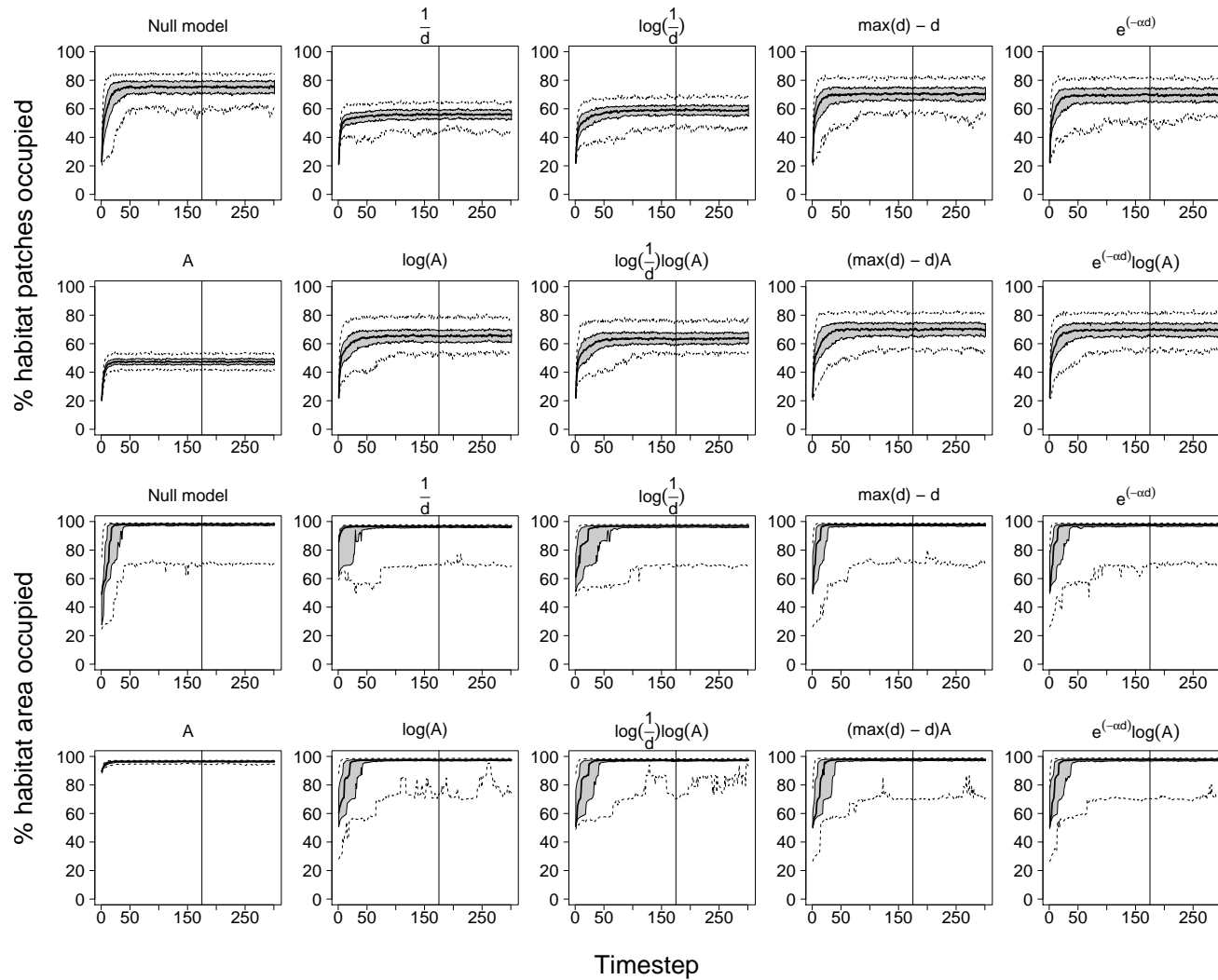


Figure C1: IFM output trajectories for *Turdus merula* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of the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

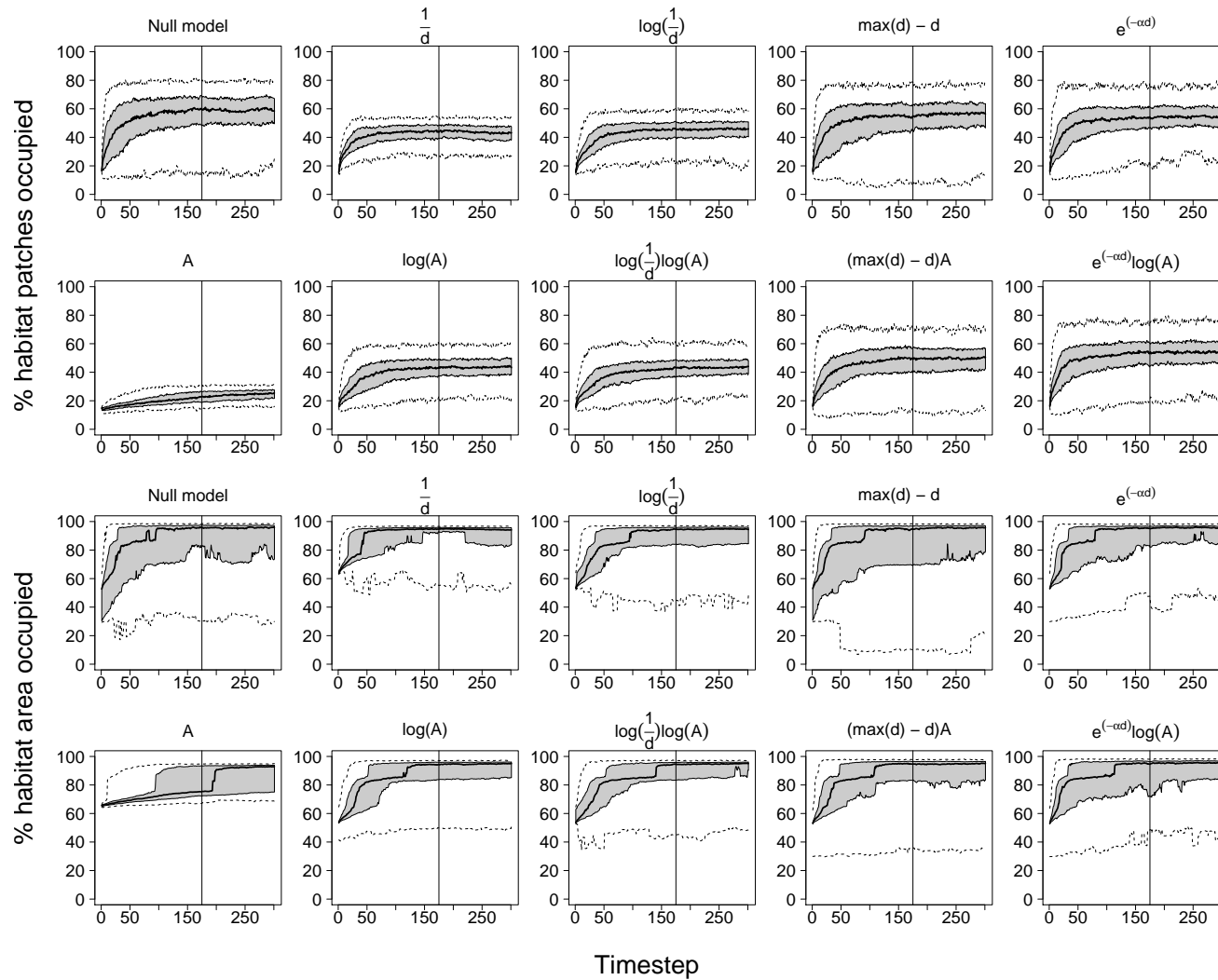


Figure C2: IFM output trajectories for *Prunella modularis* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

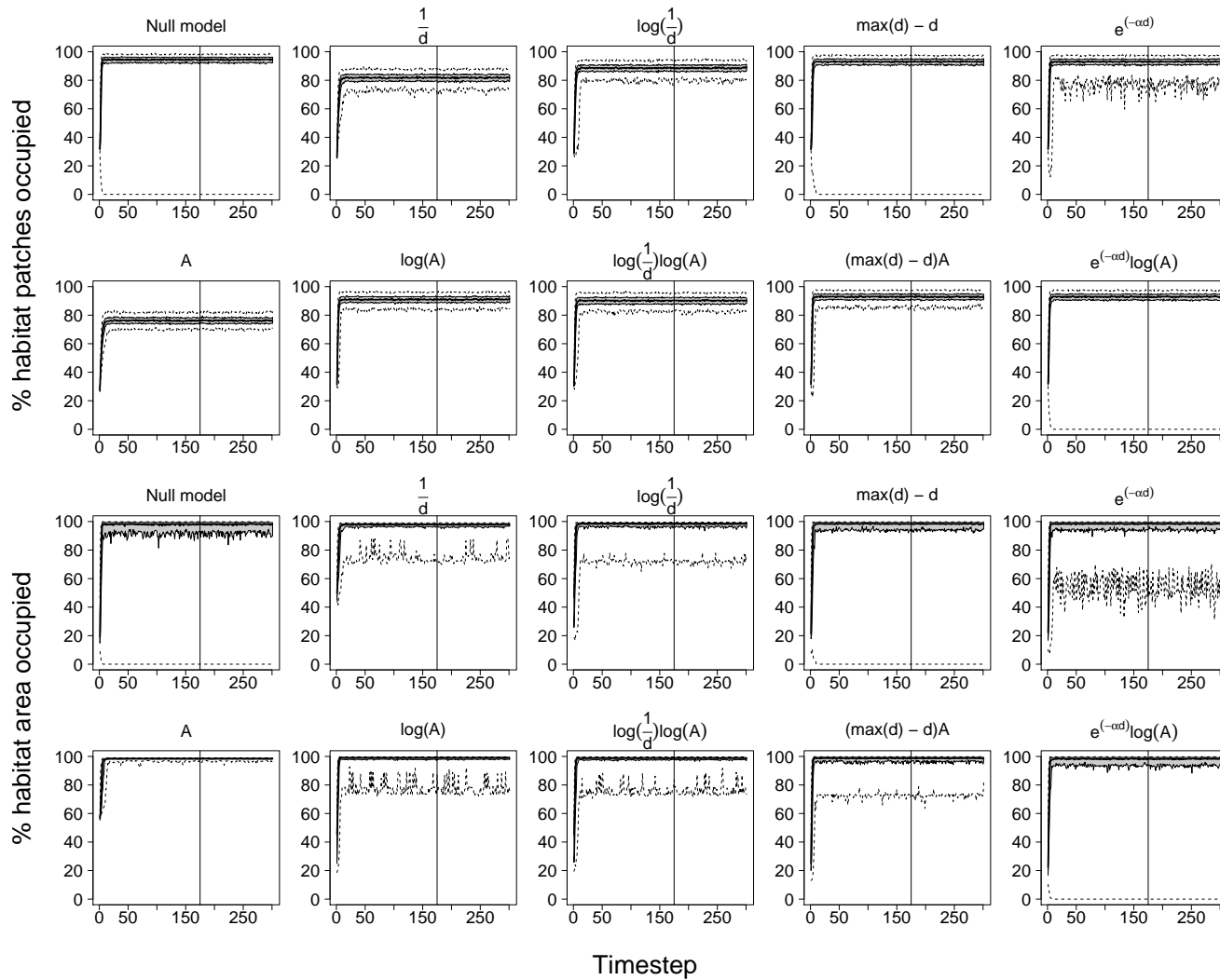


Figure C3: IFM output trajectories for *Carduelis carduelis* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

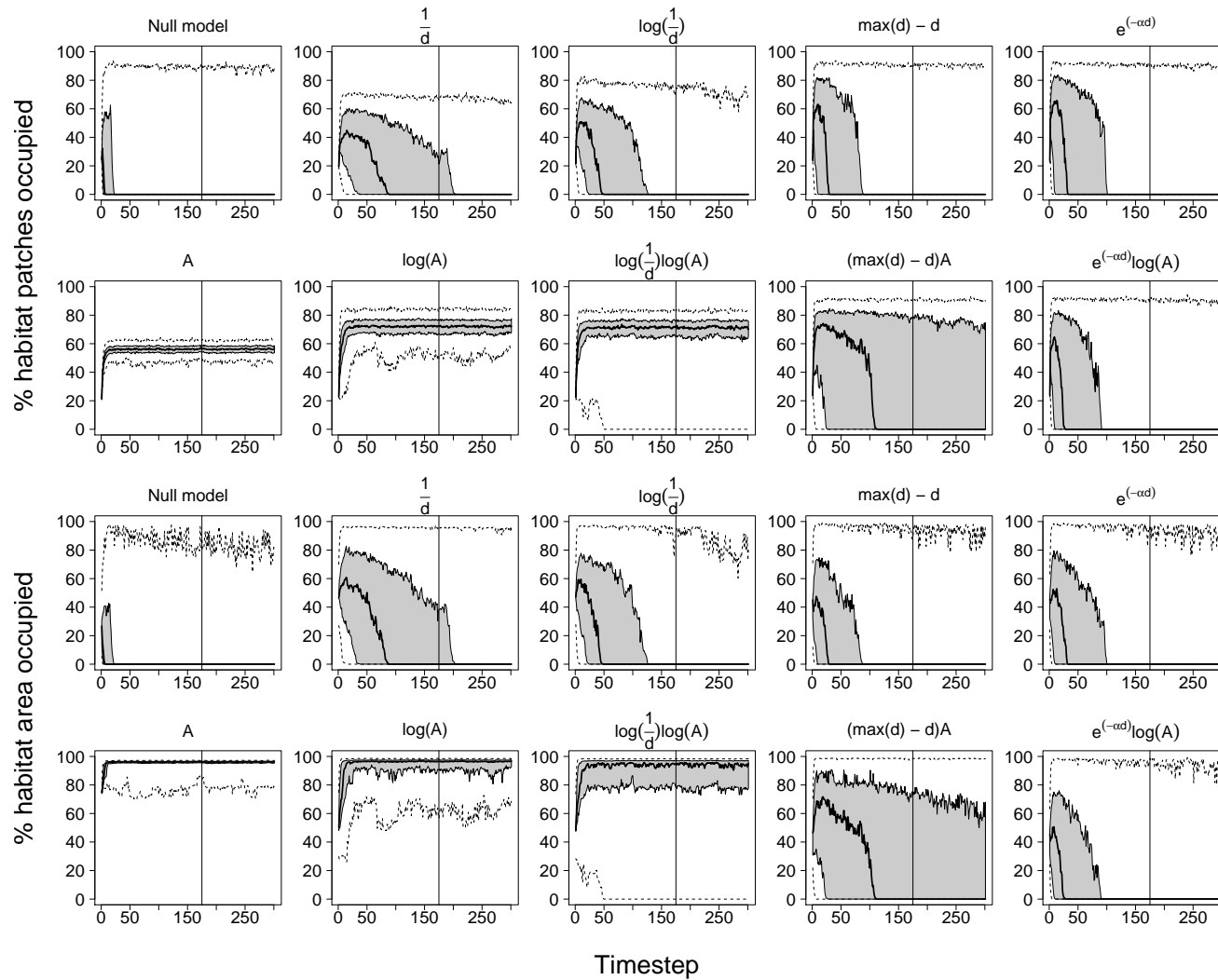


Figure C4: IFM output trajectories for *Carduelis chloris* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

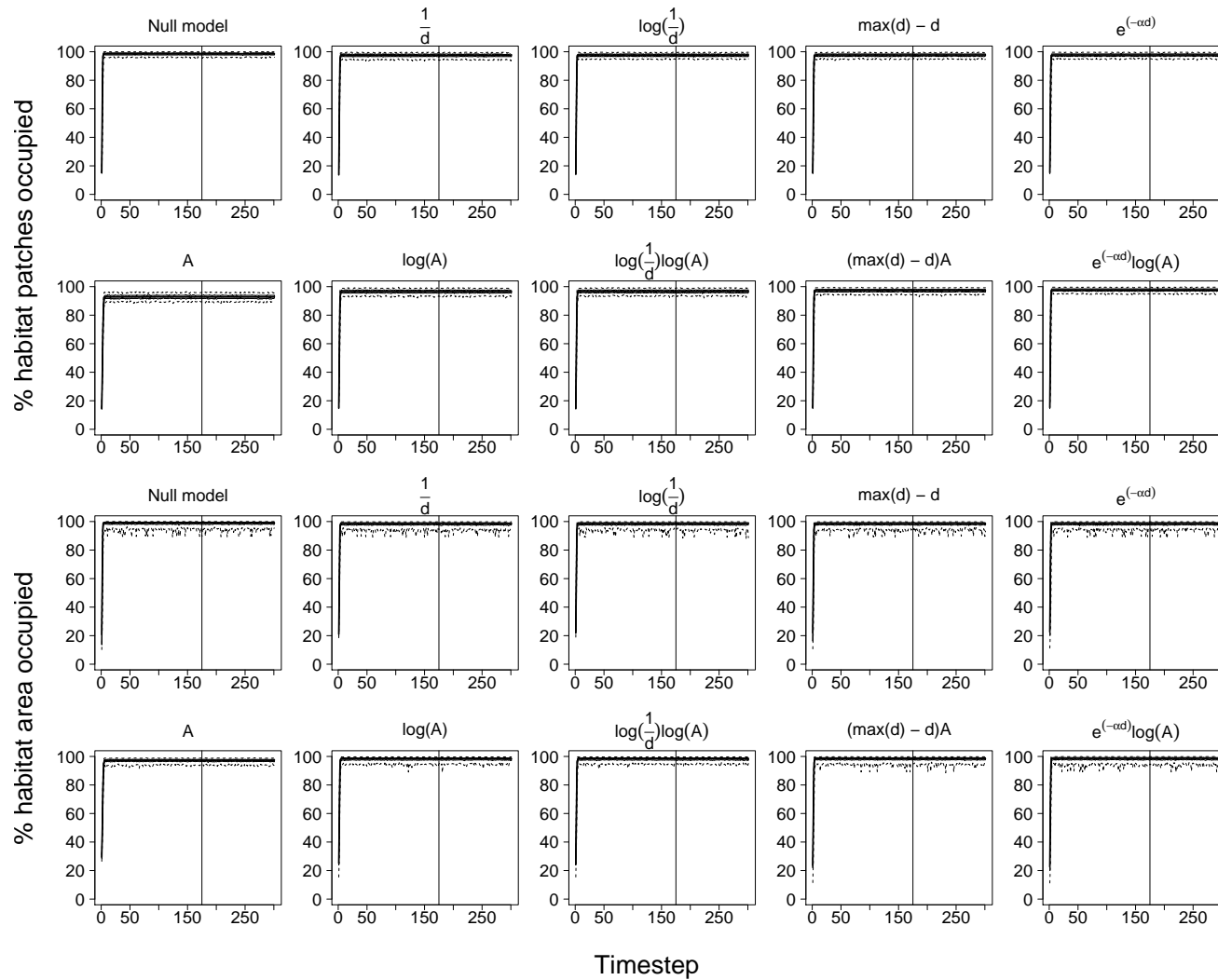


Figure C5: IFM output trajectories for *Muscicapa striata* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

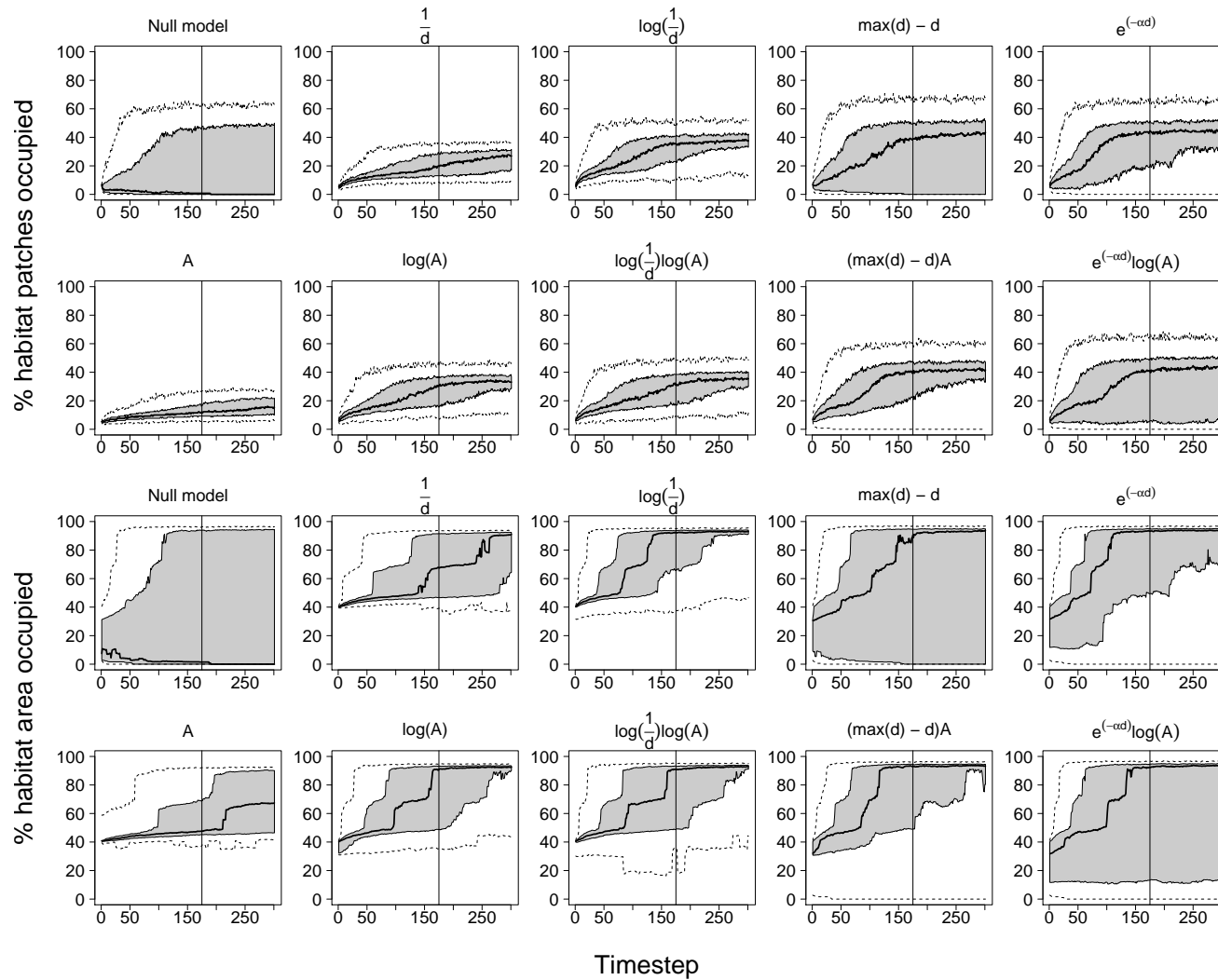


Figure C6: IFM output trajectories for *Emberiza calandra* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

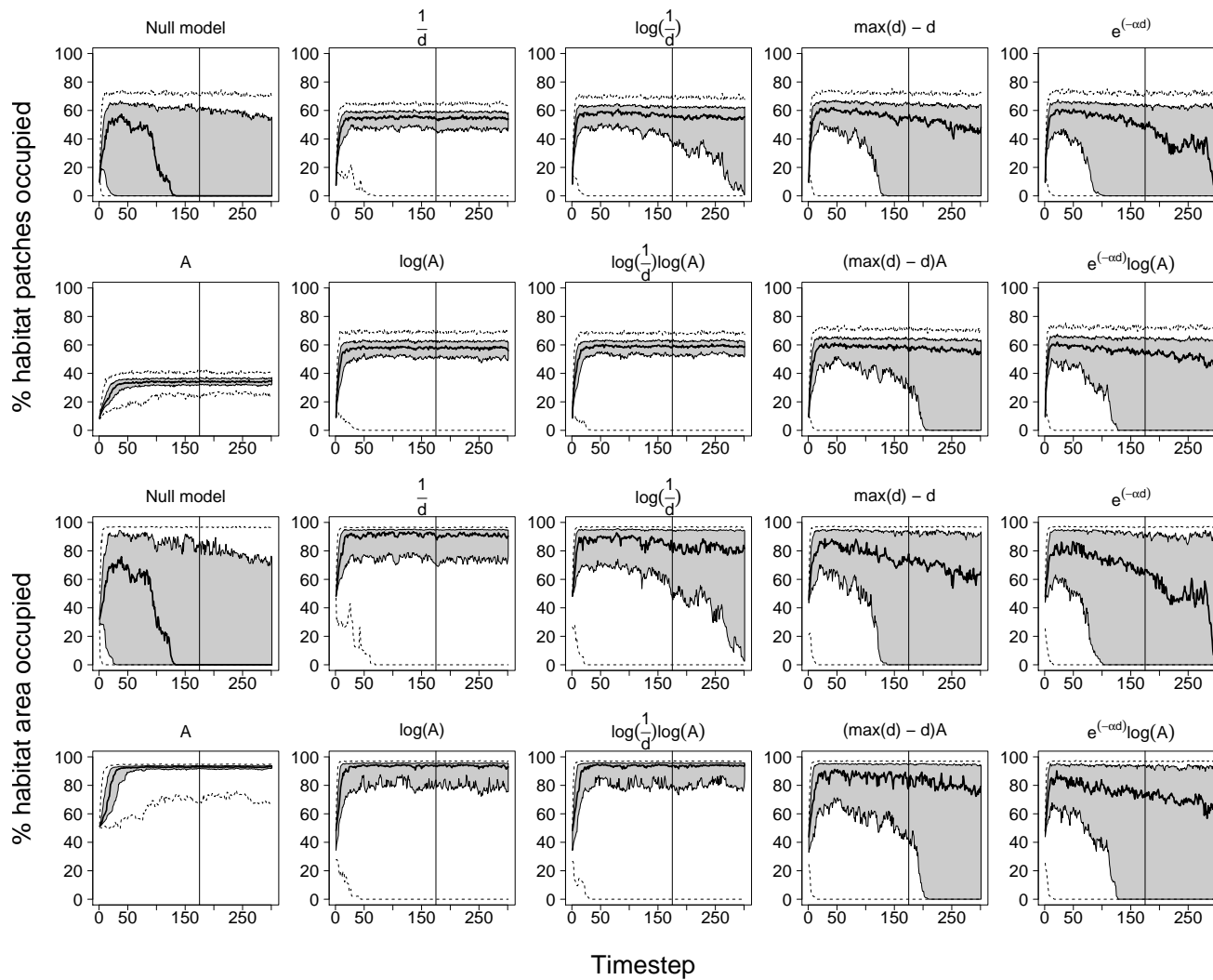


Figure C7: IFM output trajectories for *Passer montanus* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of the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

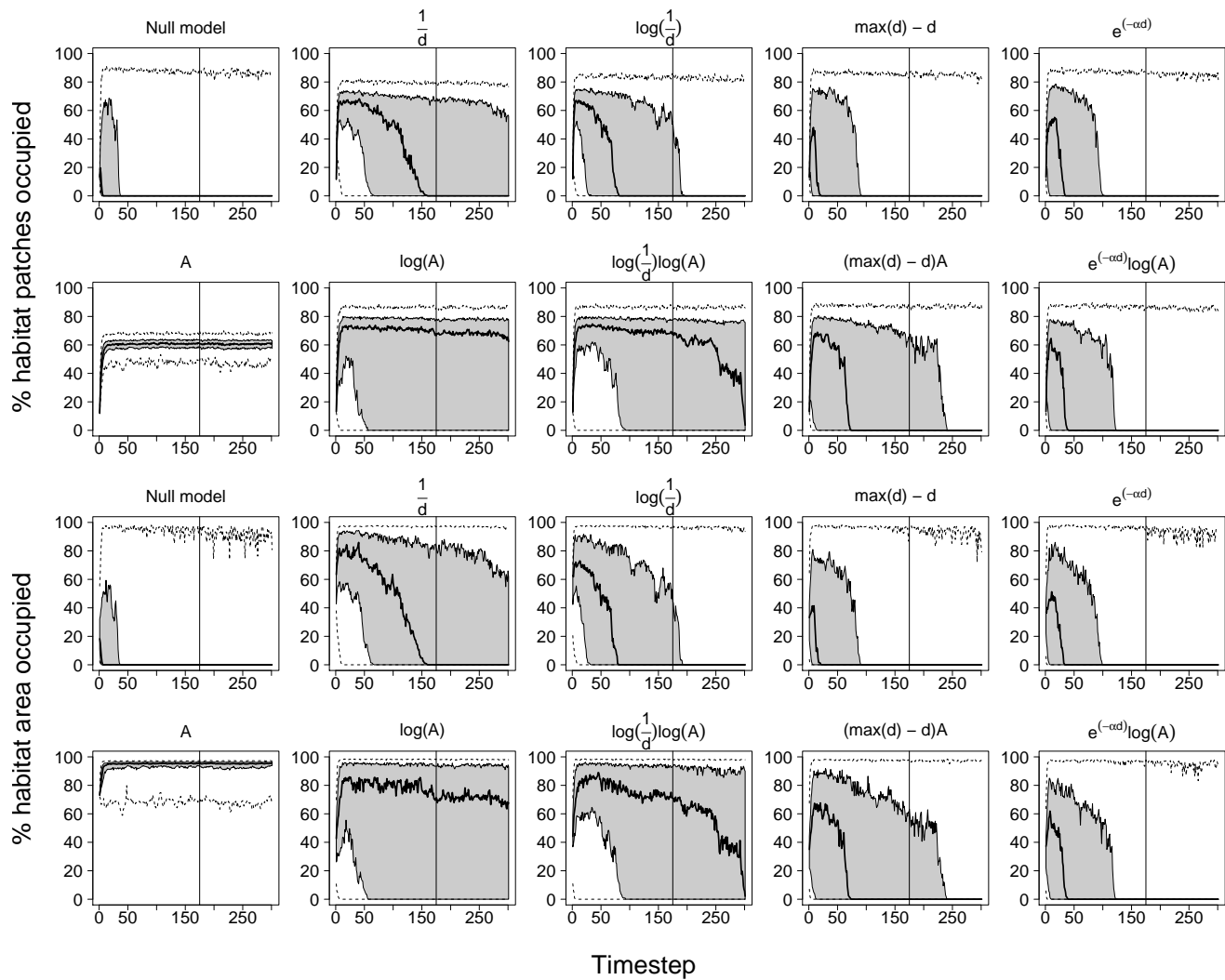


Figure C8: IFM output trajectories for *Emberiza citrinella* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

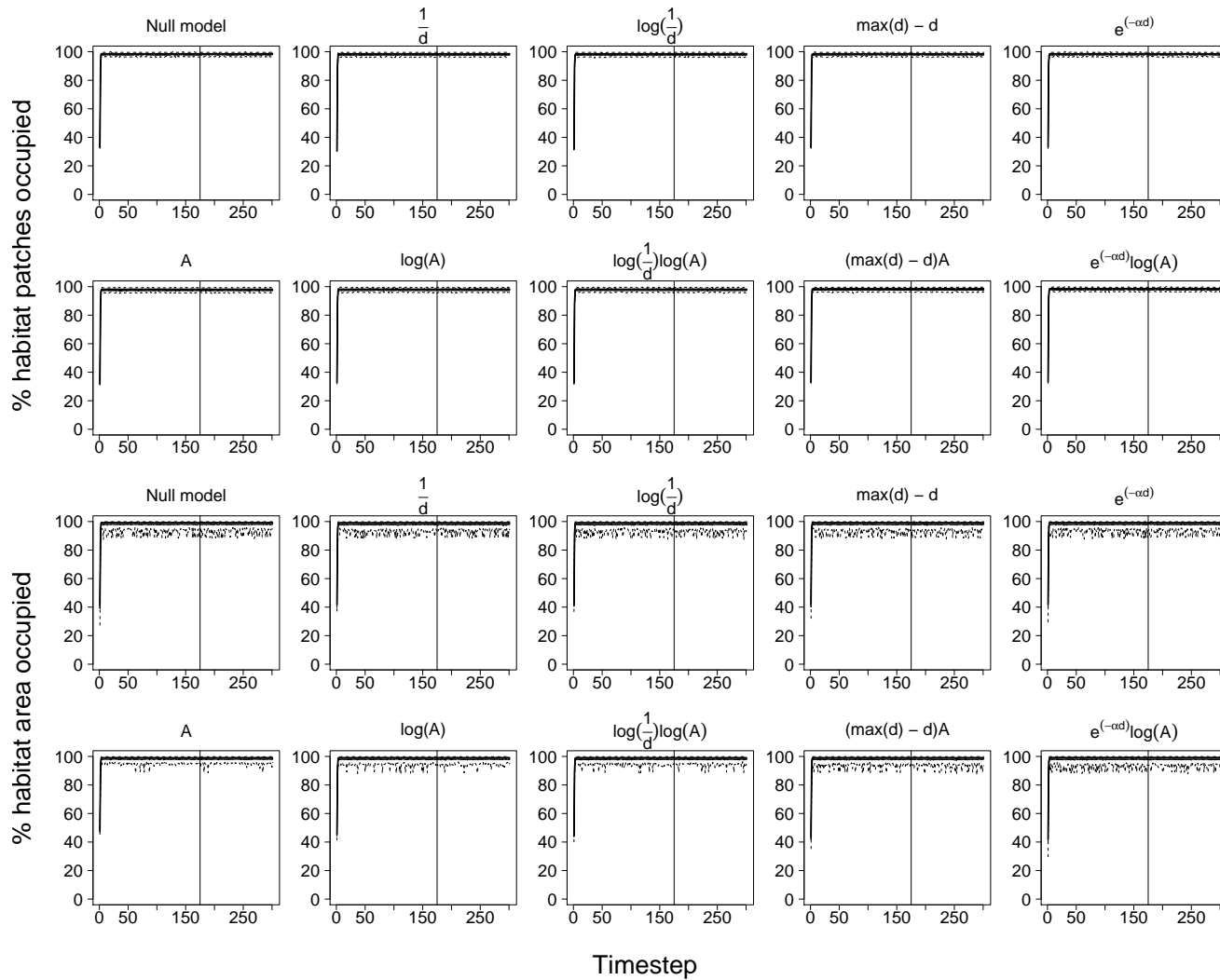


Figure C9: IFM output trajectories for *Sylvia atricapilla* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

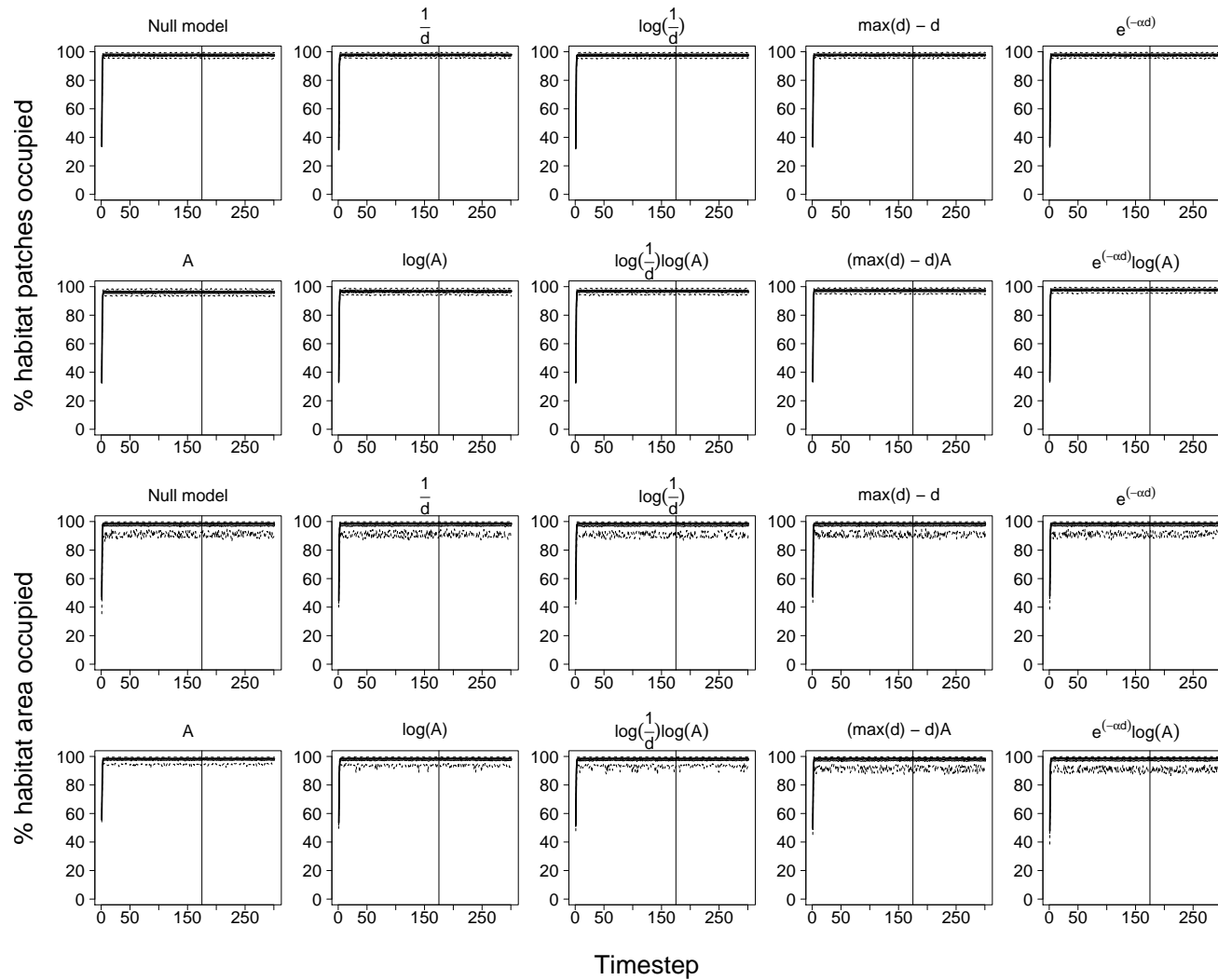


Figure C10: IFM output trajectories for *Dendrocopos major* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

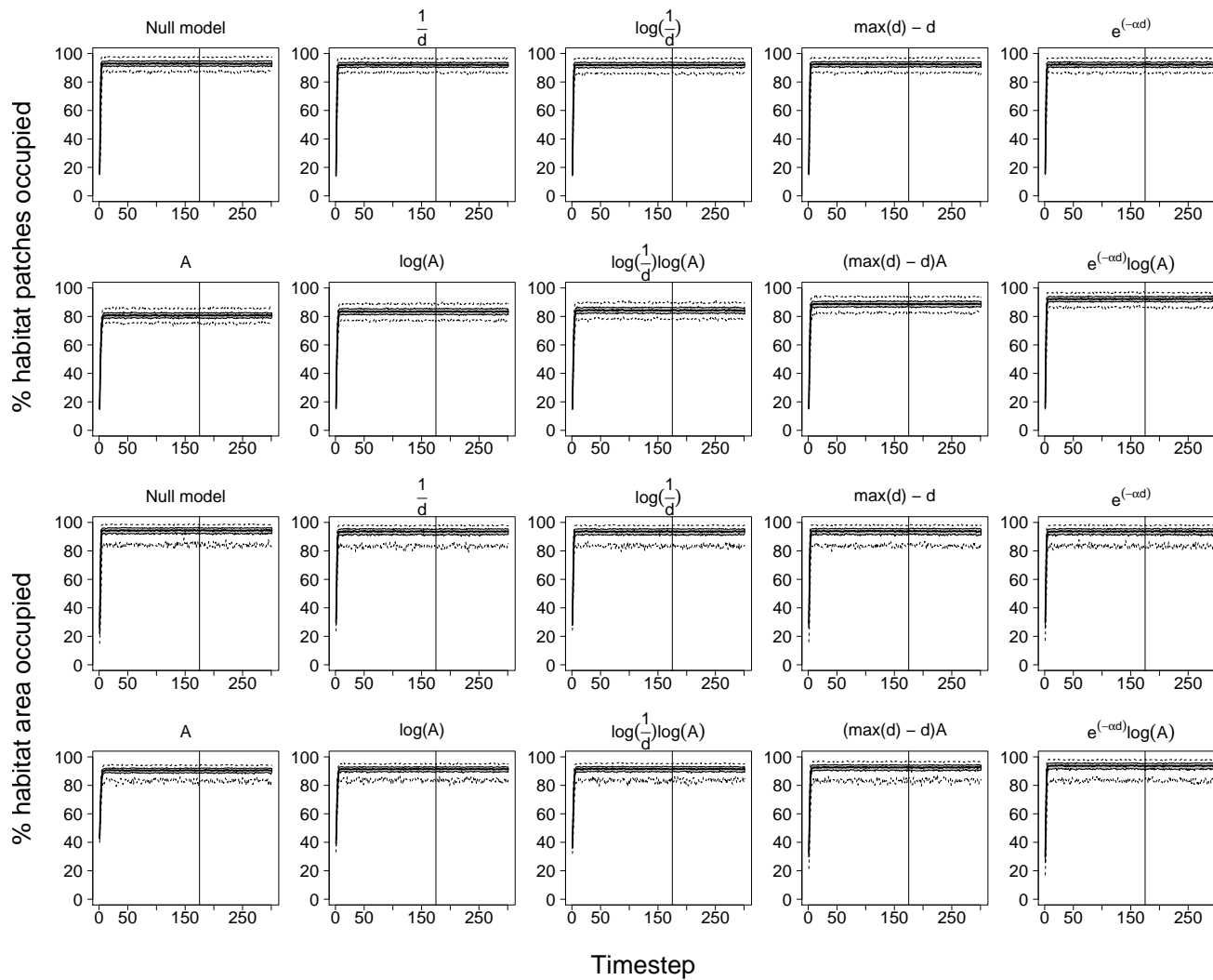


Figure C11: IFM output trajectories for *Garrulus glandarius* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations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

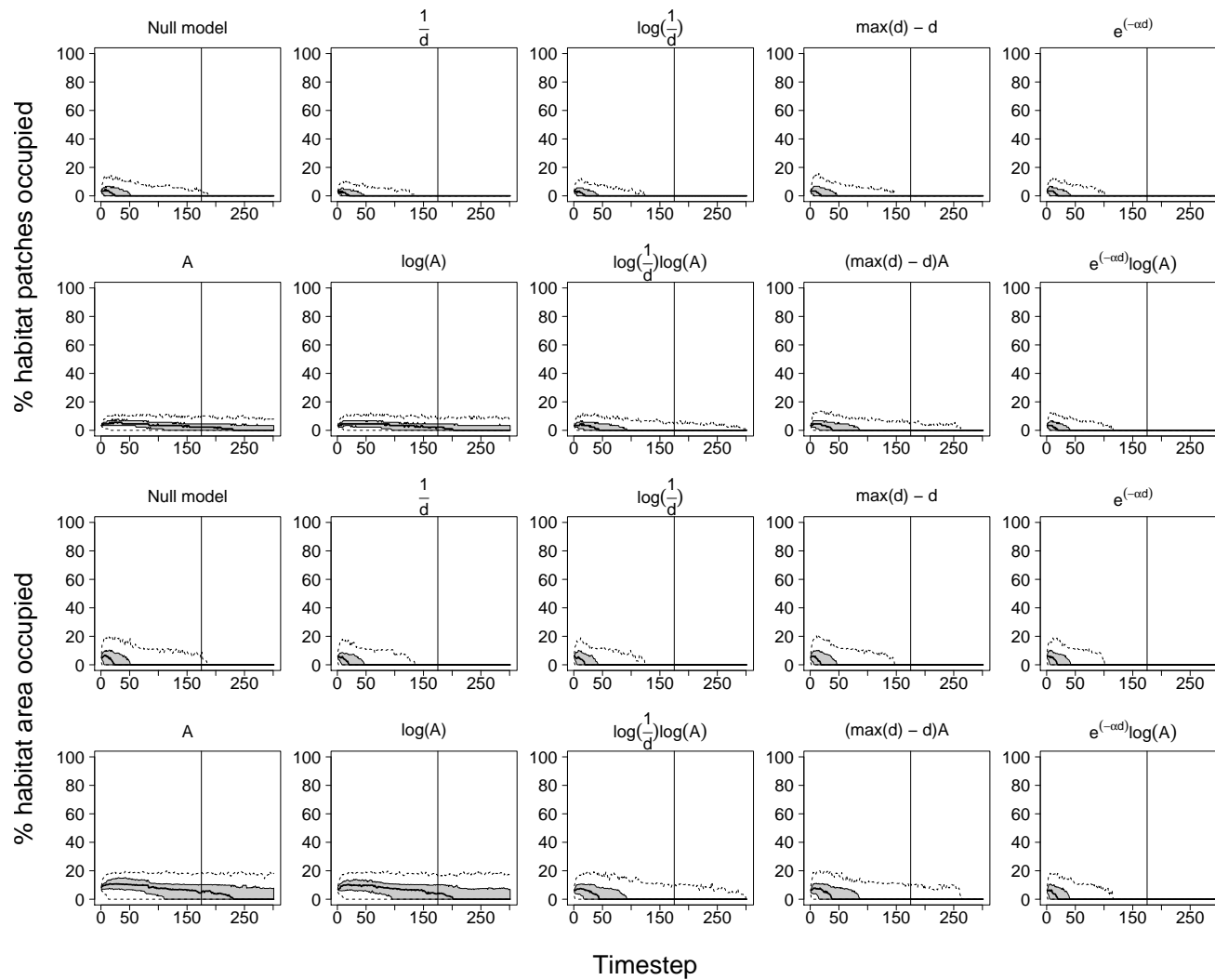


Figure C12: IFM output trajectories for *Poecile palustris* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of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

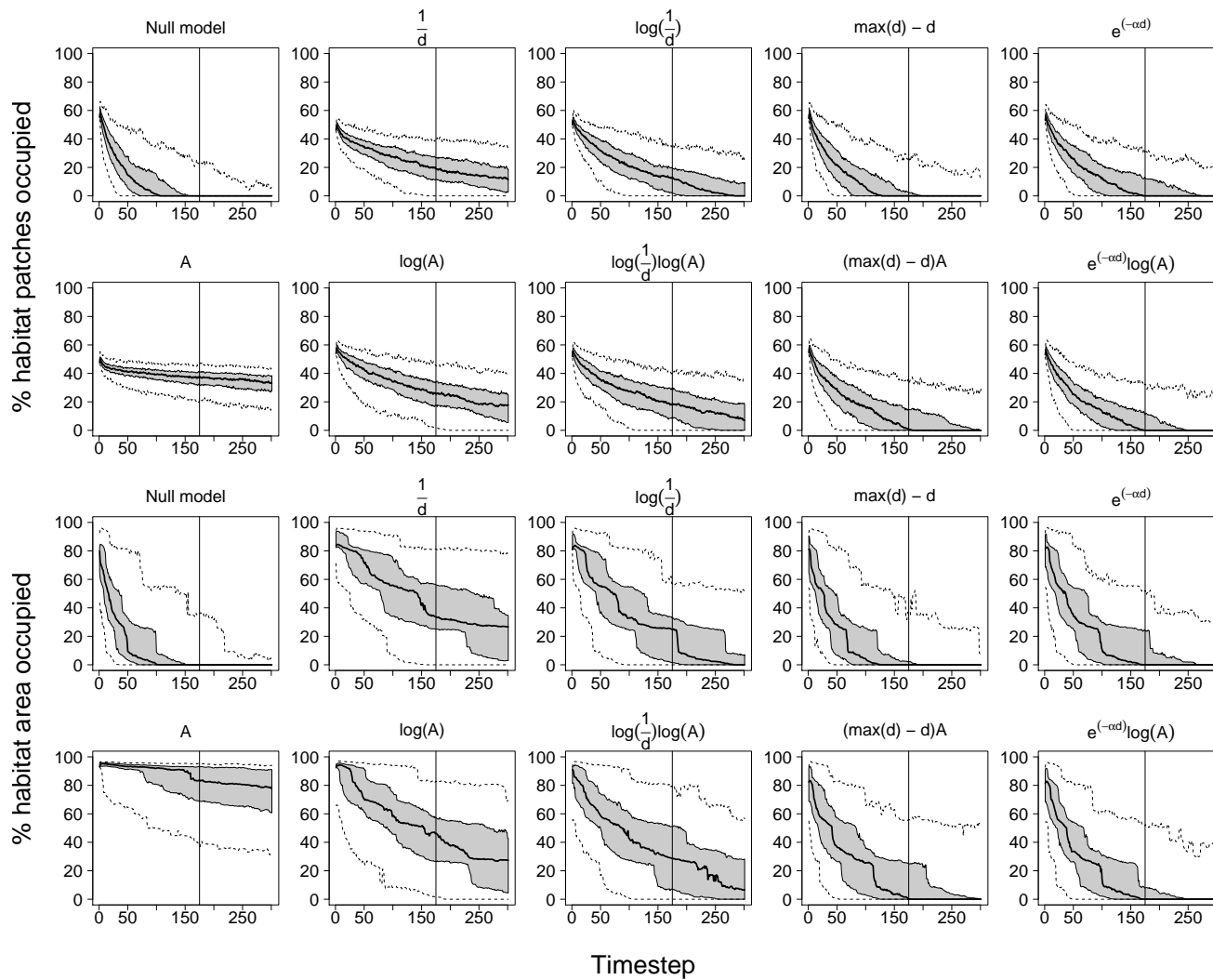


Figure C13: IFM output trajectories for *Rana temporaria* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of the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

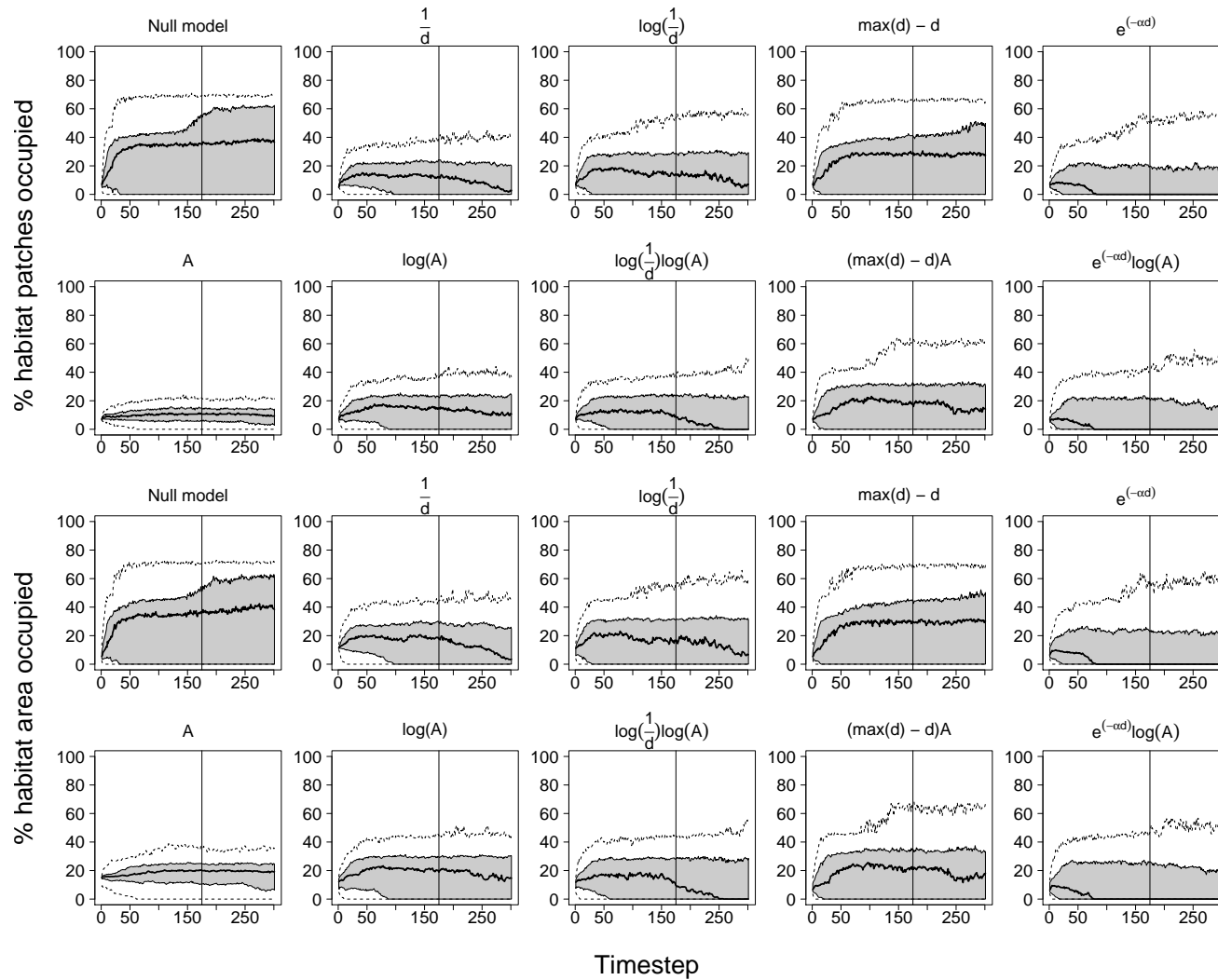


Figure C14: IFM output trajectories for *Bufo bufo* based on all ten downscaling methods. Results given show % of habitat patches occupied and % of total area occupied. Solid line shows median, shaded area the inter-quartile range, dotted lines the 5 and 95 percentile, $n = 200$ iterations of the IFM.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

824 **Appendix D - Summary of model and downscaling performance by species**

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

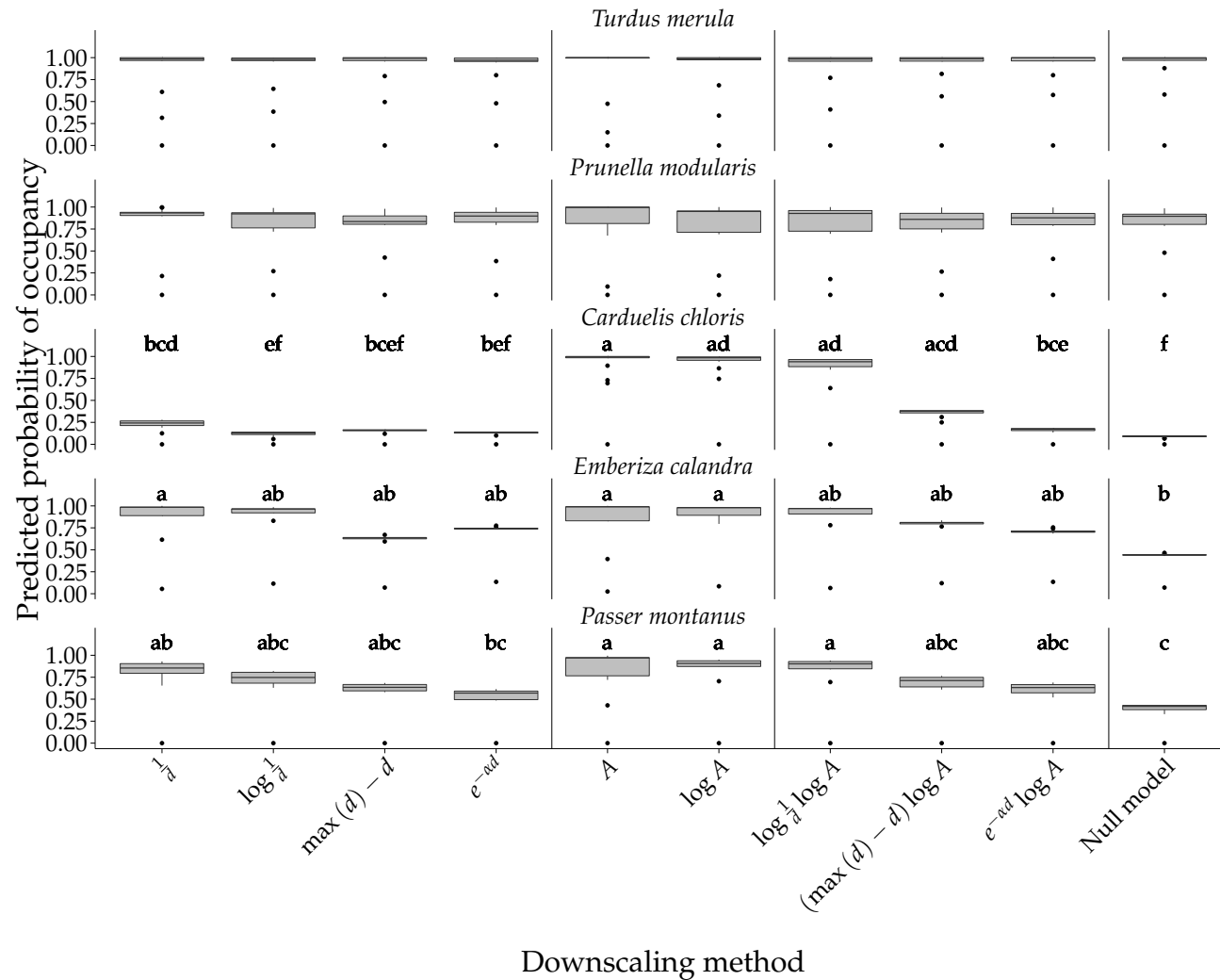


Figure D1: Boxplots showing the probability of occupancy for individual species after the burn-in period ($t = 175$) for all grid cells recorded as present in the evaluation data. Vertical lines separate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling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

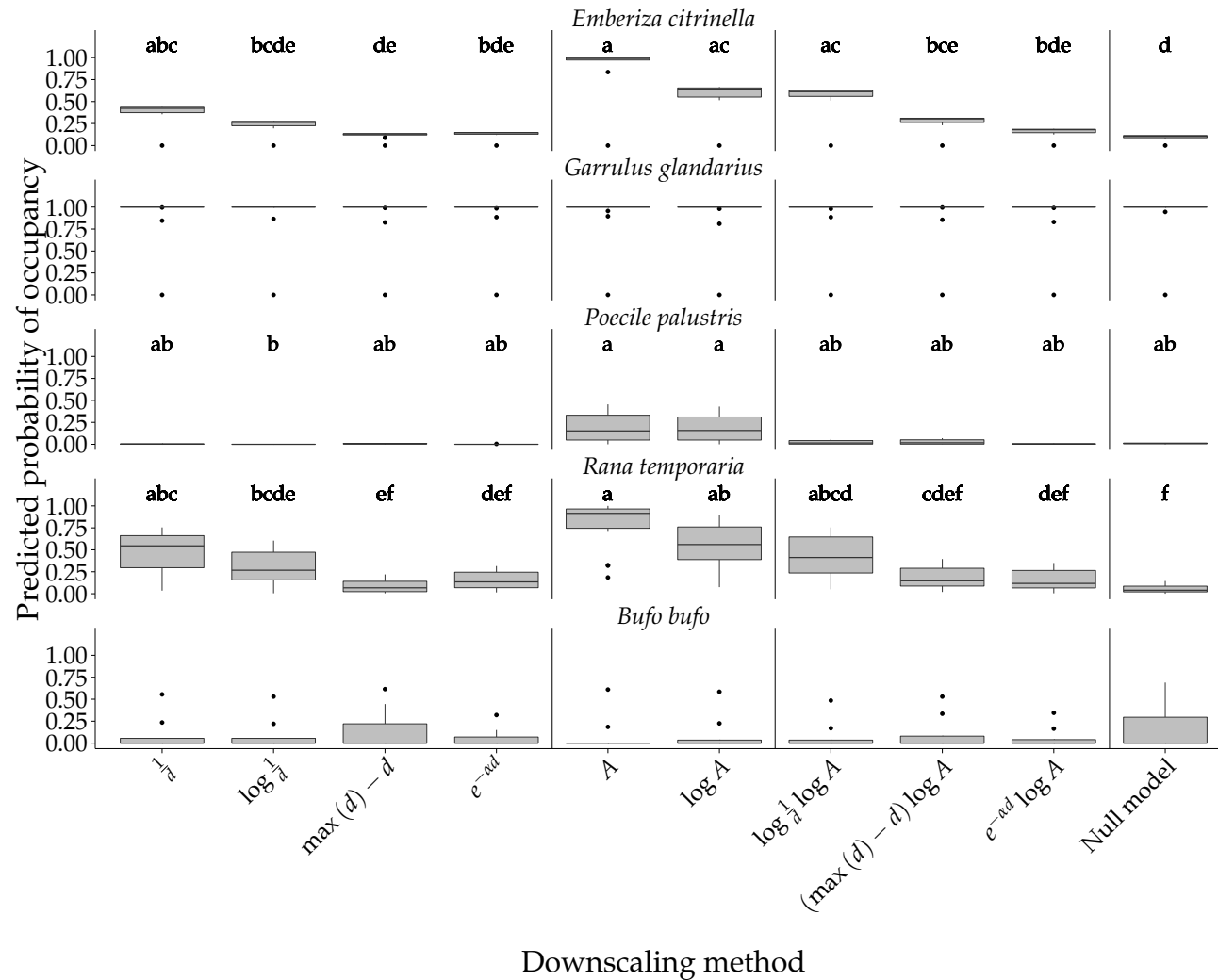


Figure D2: Boxplots showing the probability of occupancy for individual species after the burn-in period ($t = 175$) for all grid cells recorded as present in the evaluation data. Vertical lines separate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling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

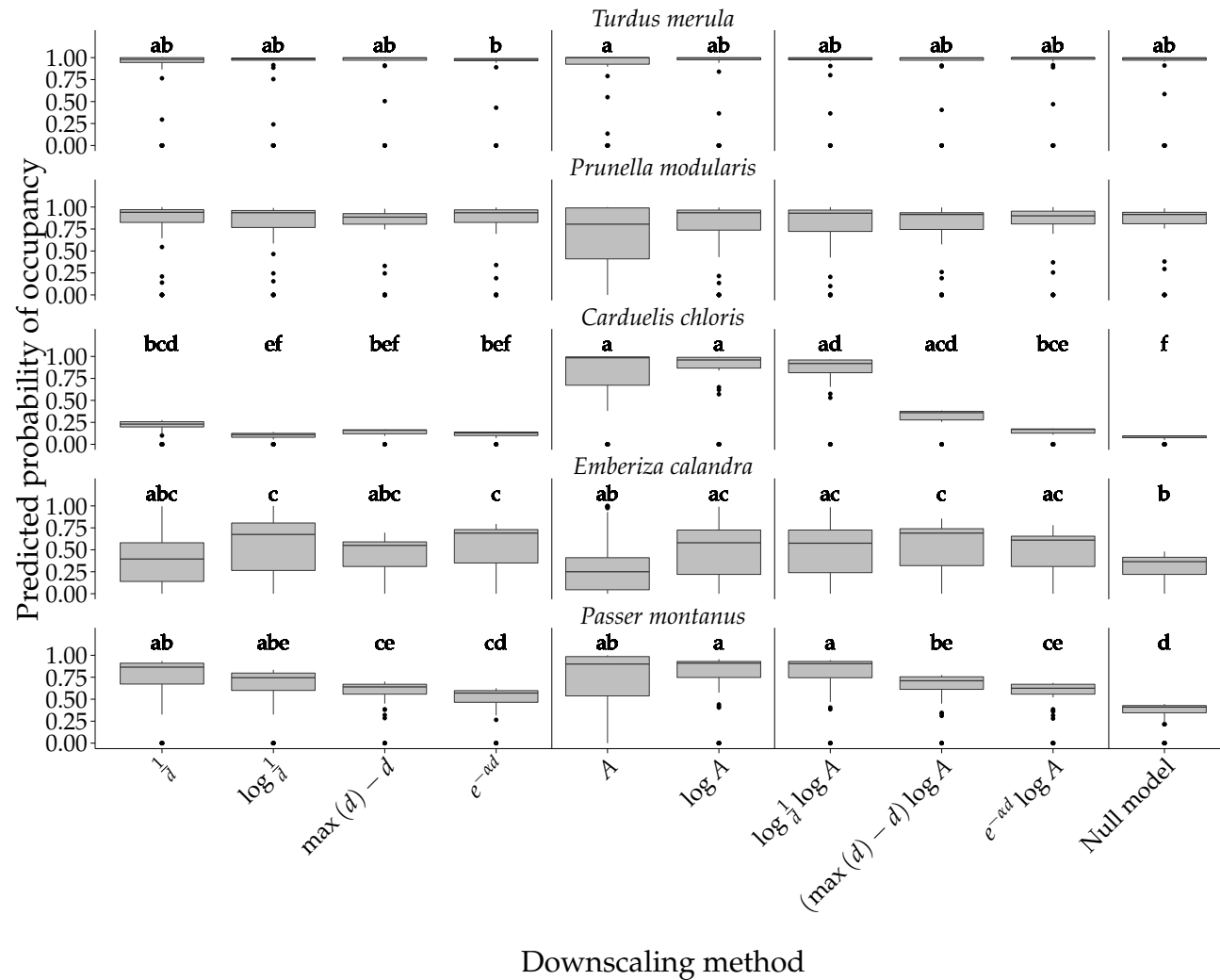


Figure D3: Boxplots showing the probability of occupancy for individual species after the burn-in period ($t = 175$) for all grid cells recorded as present in the evaluation data. Vertical lines separate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling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

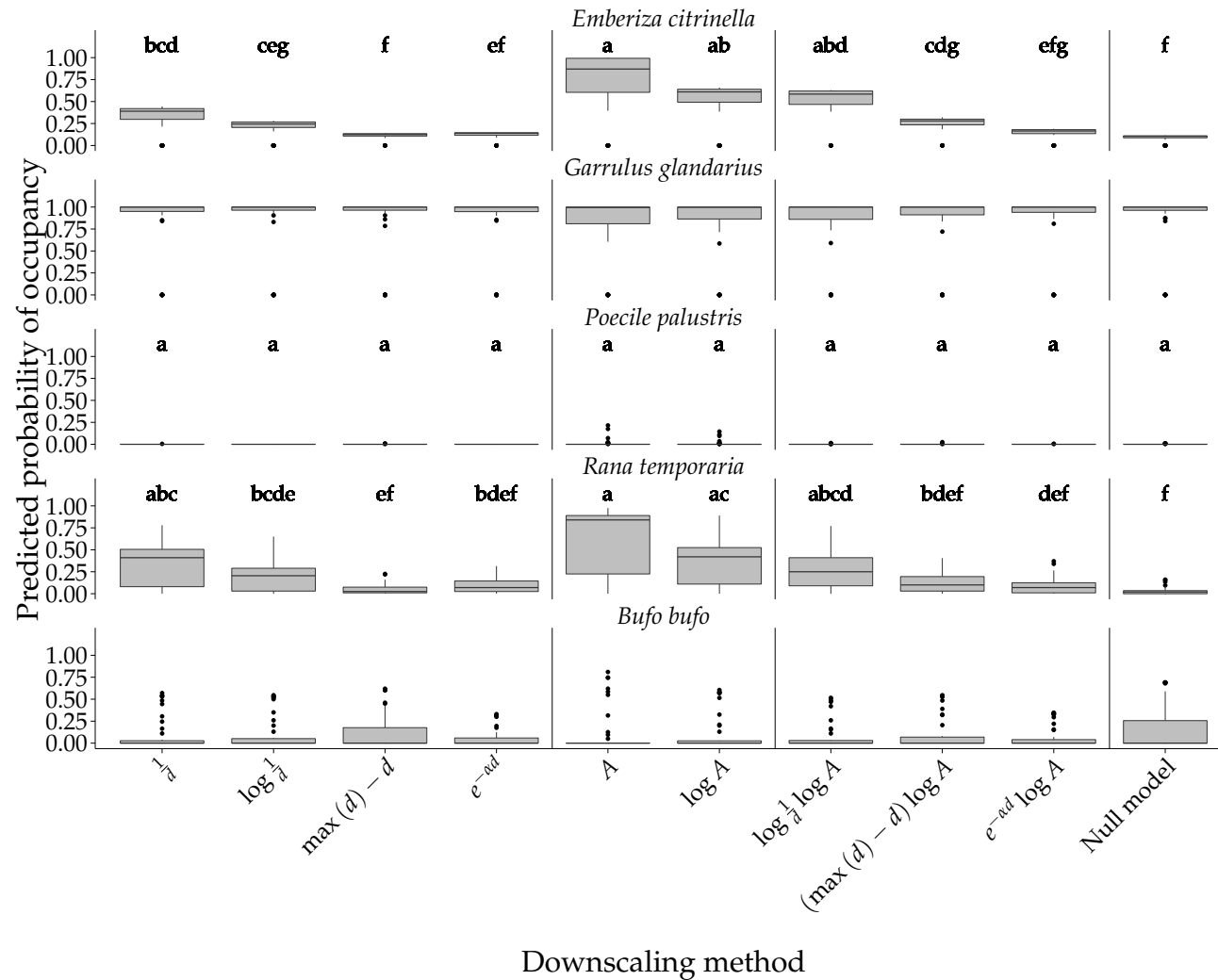


Figure D4: Boxplots showing the probability of occupancy for individual species after the burn-in period ($t = 175$) for all grid cells recorded as present in the evaluation data. Vertical lines separate the downscaling methods into distance only, area only, combined distance and area, null. In the downscaling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

837 **Appendix E - Validation with independent data from the British Trust for**
838 **Ornithology**

839 Validation against an independent data set — the 2 km data from the field-
840 work for the 2nd British Trust for Ornithology Atlas (1988–1991, Gibbons et al.
841 1993) — was performed in the same way as the validation against the Notting-
842 hamshire Birdwatchers dataset, but limited to only grid cells visited in the BTO
843 data.

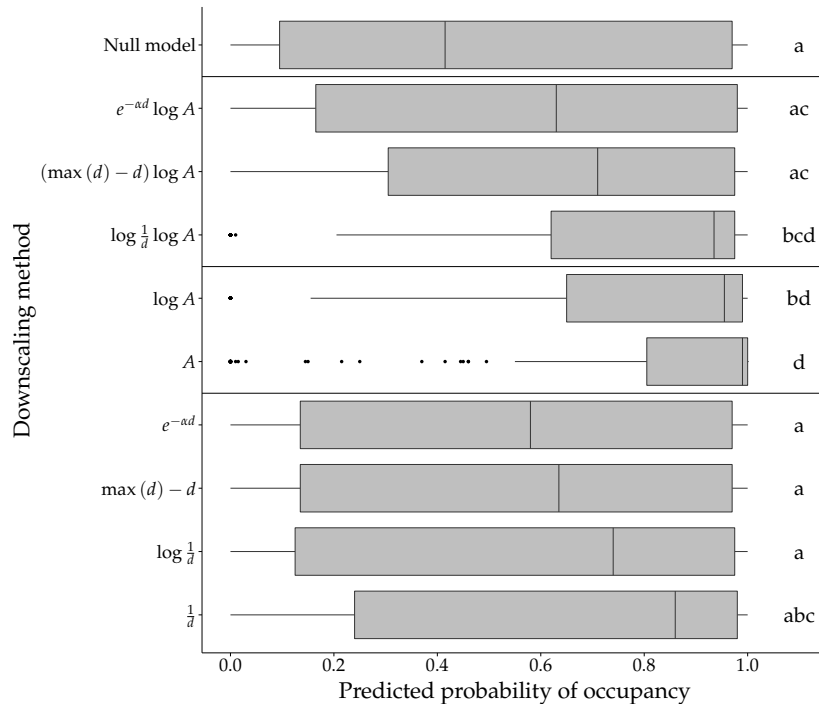


Figure E1: Boxplots showing the probability of occupancy for all species after the burn-in period ($t = 175$ time steps) for all grid cells recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). For each species, there are $n = 200$ replicates of IFM simulations. 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 α 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, $df = 9$, $P < 0.001$.

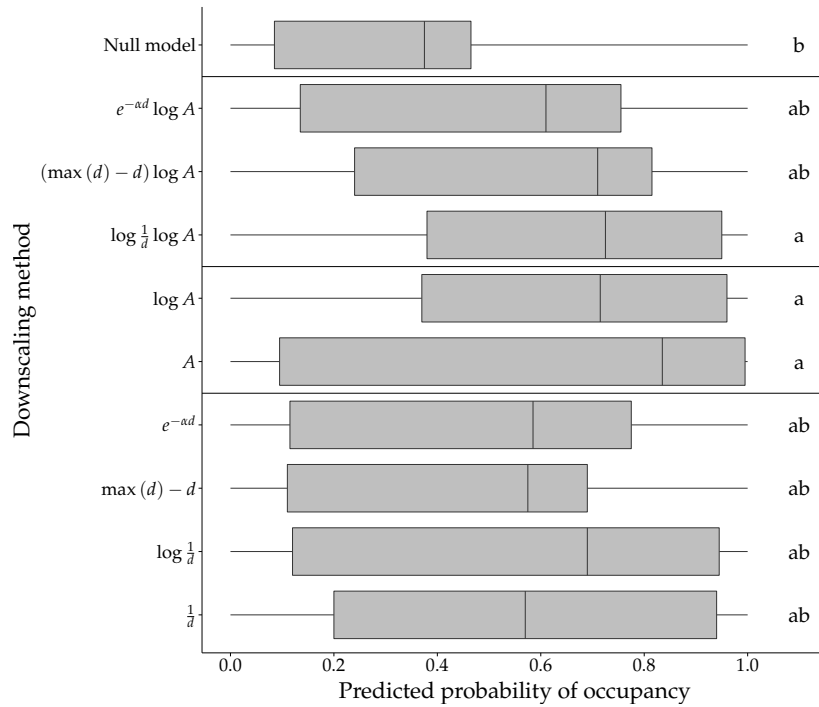


Figure E2: Boxplots showing the probability of occupancy for all species after the burn-in period ($t = 175$ time steps) for all grid cells which were visited but the species not recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). For each species, there 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 α 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, $df = 9$, $P < 0.001$.

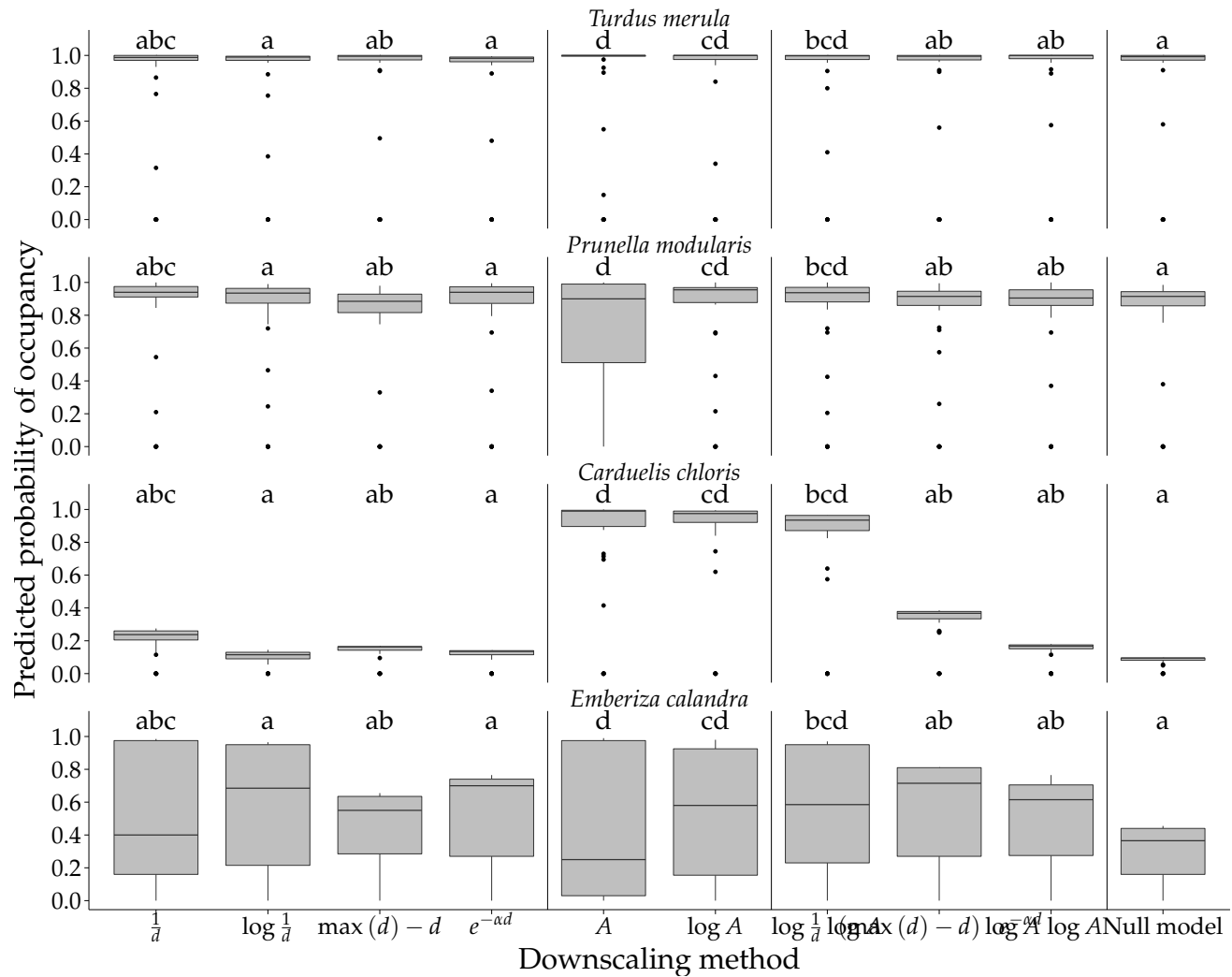


Figure E3: Boxplots showing the probability of occupancy for individual species after the burn-in period ($t = 175$) for all grid cells recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). Vertical lines separate the downscaling methods into distance only,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

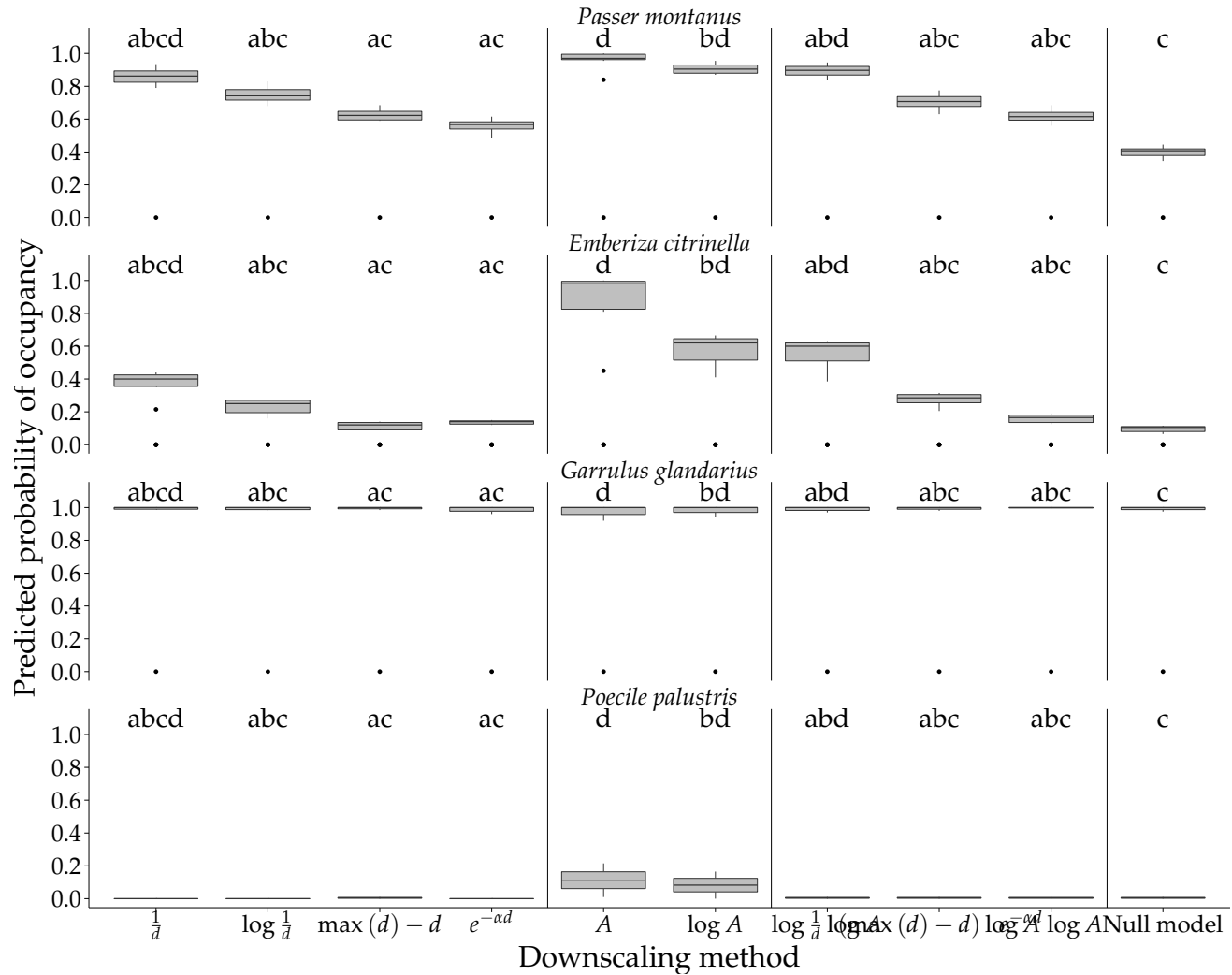


Figure E4: Boxplots showing the probability of occupancy for individual species after the burn-in period ($t = 175$) for all grid cells recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). Vertical lines separate the downscaling methods into distance only,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

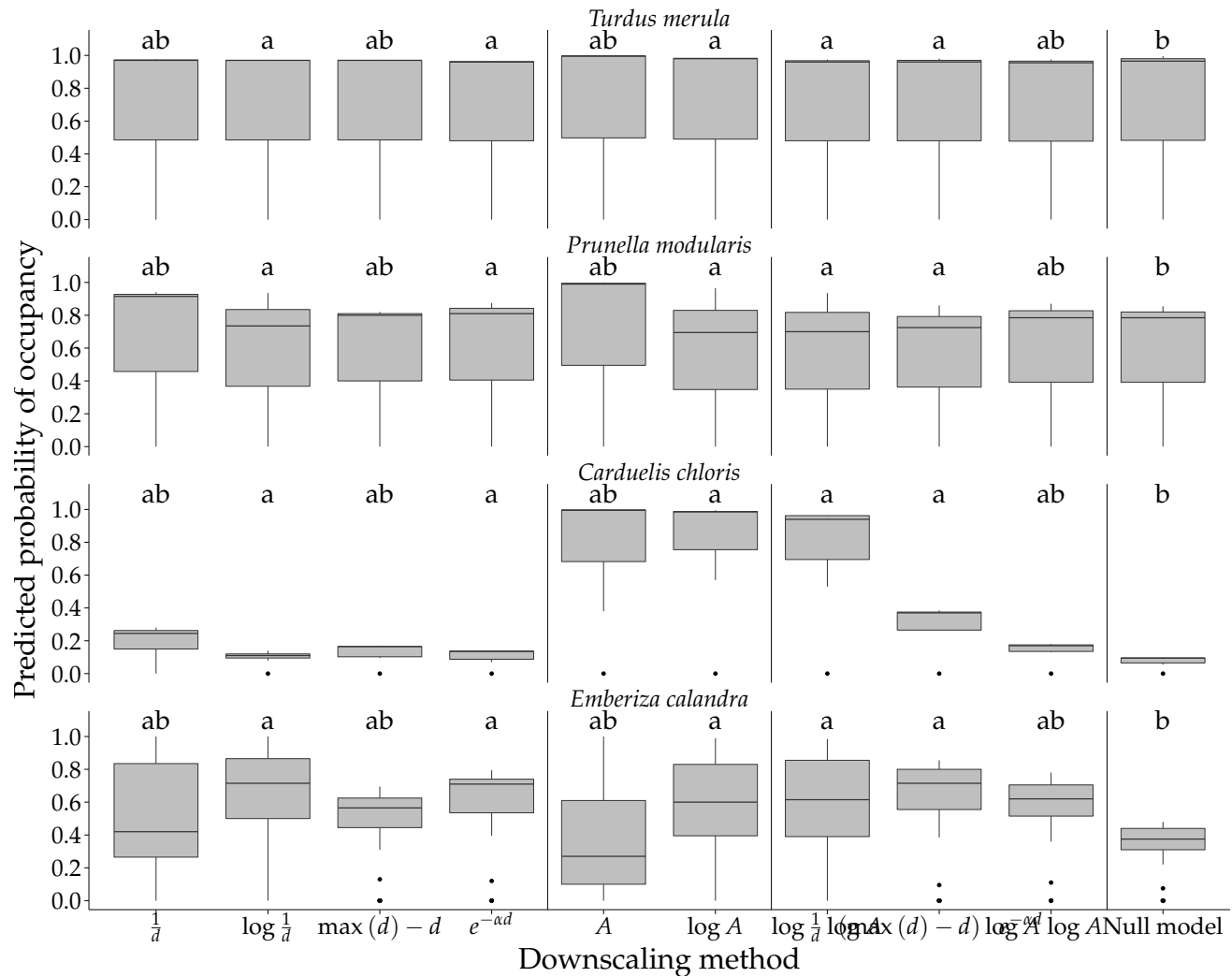


Figure E5: Boxplots showing the probability of occupancy for individual species after the burn-in period ($t = 175$) for all grid cells which were visited but the species not recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). Vertical lines separate 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 α the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).

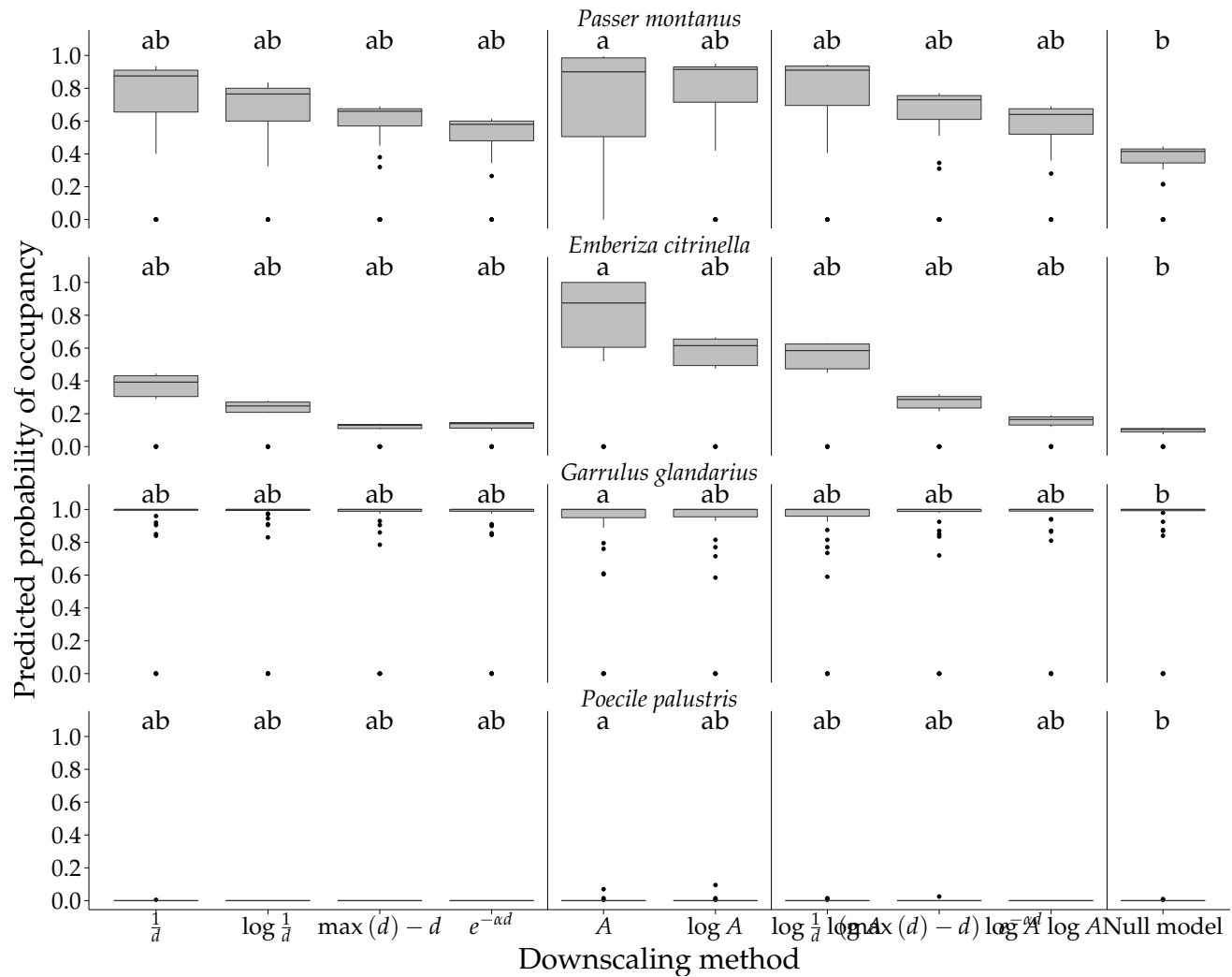


Figure E6: Boxplots showing the probability of occupancy for individual species after the burn-in period ($t = 175$) for all grid cells which were visited but the species not recorded as present in the independent evaluation data from the 2nd BTO Atlas (Gibbons et al., 1993). Vertical lines separate 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 α the species dispersal ability. Different letters indicate that we detected a significant difference between the median occupancy between methods based on post-hoc tests ($P < 0.05$).