



HAL
open science

Space-time species distribution modeling for opportunistic presence-only data: a case study of passerines in a protected area

Florian Lasgorceux, Julien Papaix, Yoann Bunz, Damien Combrisson,
Thomas Opitz

► To cite this version:

Florian Lasgorceux, Julien Papaix, Yoann Bunz, Damien Combrisson, Thomas Opitz. Space-time species distribution modeling for opportunistic presence-only data: a case study of passerines in a protected area. 2024. hal-04616332v2

HAL Id: hal-04616332

<https://hal.science/hal-04616332v2>

Preprint submitted on 1 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Space-time species distribution modeling with opportunistic presence-only data: a case study of passerines in a protected area

Florian Lasgorceux¹, Julien Papaïx¹, Yoann Bunz²,
Damien Combrisson², and Thomas Opitz¹

<https://doi.org/10.17180/RAM8-MV76>

Abstract

Over the recent decades, Europe has experienced a significant decline in common bird species, particularly farmland species, due to anthropic pressures like agricultural intensification. Protected areas, such as the Écrins National Park in France, play a crucial role in mitigating these impacts. In this study, we utilized an opportunistic presence-only dataset of passerine occurrences to evaluate the impact of protected areas on passerine trends. The data, gathered by trained rangers from Écrins National Park, were used for spatio-temporal modeling to assess the relative abundance of 76 passerine species on a regular grid, with aggregated occurrences per spatio-temporal cell serving as a proxy for sampling effort. The model demonstrated good calibration for most species, with AUC values from different cross-validation schemes exceeding 0.8 in most cases. The model also effectively distinguished habitat preferences and migratory status. Additionally, we compared trends in relative abundance over 2001–2019 in the Écrins National Park with the trends in abundance across France using the French Breeding Bird Survey (STOC program) for 59 common species. Our analysis revealed that forest specialist species benefited the most from protected areas, farmland species decreased more slowly, and high-elevation specialist species generally decreased. Though, some results should be interpreted with care since our assumption of a homogeneous reporting rate across target group species may be too strong for some very common species. Our findings illustrate that the high spatio-temporal resolution provided by opportunistic presence-only data offers valuable insights into biological phenomena and their trends while reducing the need for other data sources to determine sampling effort proxies. Similar datasets from other protected areas could serve as powerful tools for assessing the effectiveness of conservation policies.

Keywords: Écrins National Park, INLA, habitat preference, migratory status, Poisson process, opportunistic presence-only data, relative abundance, sampling effort, spatio-temporal model, STOC program, target group

¹Biostatistics and Spatial Processes, INRAE, Avignon, France, ²Parc national des Écrins, Domaine de Charance, Gap, France

Correspondence

florian.lasgorceux@proton.me

Introduction

2

3 Over the recent decades, Europe has experienced a significant decline in the populations of
4 common bird species (Inger et al., 2015), particularly those inhabiting agricultural landscapes
5 (Donald et al., 2001). This sharp decline in farmland bird species has led to extensive research
6 across various countries (Heldbjerg et al., 2018; Kamp et al., 2021; Newton, 2004; Reif et al.,
7 2008b; Sanderson et al., 2013; Traba and Morales, 2019; Wretenberg et al., 2006). The conclu-
8 sions regarding the health status of forest species in Europe are however less consistent. Gre-
9 gory et al., 2007 estimated a decline of 13% in common forest birds and 18% in specialists
10 from 1980 to 2003. In contrast, by considering a different indicator over the period 1980–2015,
11 Gregory et al., 2019 found a relative stability in common European forest bird populations. Mi-
12 gratory status may explain these differences, with non-migratory birds benefiting from improved
13 forest management and migratory species declining (Schulze et al., 2019). Habitat preferences
14 also play a role; species in lowland broad-leaved forests fare better than those in montane and
15 coniferous forests (Reif et al., 2008a). This has been confirmed by the study of Lehtikoinen et
16 al., 2019 affirming that mountain bird species in Europe have declined by 10% due to climate
17 and land use changes, pushing distributions towards mountaintops (Flousek et al., 2015; Zamora
18 and Barea-Azcón, 2015). A recent study by Rigal et al., 2023 identified agricultural intensifica-
19 tion as the primary factor in the general decline of common bird abundance in Europe, especially
20 among invertebrate feeders, with varied responses to forest cover, urbanization, and tempera-
21 ture changes.

22 Protected areas and conservation efforts therefore play a crucial role in mitigating the de-
23 cline of bird populations across Europe. According to the IPBES, 2019, conservation investments
24 from 1996 to 2008 have resulted in an average reduction of 29% in the extinction risk for mam-
25 mals and birds across 109 countries. Focusing on birds across Europe, the EU's Natura 2000
26 protected area network, while not preventing country-wide population declines in some threat-
27 ened grassland bird species (Silva et al., 2018), has been effective for bird conservation, notably
28 for threatened species (Duckworth and Altwegg, 2018). Studies by Barnes et al., 2023 also indi-
29 cate positive associations between protected areas and bird abundance, particularly benefiting
30 rare and declining species. Timmers et al., 2022 conducted a comprehensive meta-analysis and
31 demonstrated a strong association between strict protection measures (International Union for
32 Conservation of Nature [IUCN] categories I–IV) and higher bird occurrence in larger forest frag-
33 ments.

34 National Parks in France are one of those protected areas. A National Park is formed of two
35 geographic zones with different regulatory statuses: the core (regulated area) – the National Park
36 guarantees the protection of this area, as per its decrees, and manages all human activities car-
37 ried out there, in line with its management objectives – and the surrounding area (projects) – a
38 space designed to achieve ecological consistency and solidarity with the National Park core. In
39 this area, the National Park plays an advisory role and can act as a partner in projects to preserve
40 and promote the natural, historic, cultural and landscape heritage. The Écrins National Park is
41 also characterized by significant elevation changes, including multiple peaks exceeding 4000 me-
42 ters, extensive coniferous forests (particularly larch), and numerous alpine pastures. The park is
43 responsible for multiple biodiversity monitoring programs within its territory (Bunz, 2022), which
44 have led to extensive synthesis and research works (Noël et al., 2023).

45 In addition to data from monitoring programs, the Écrins National Park maintains a large

46 dataset of presence-only records collected opportunistically by rangers. The primary use of these
47 data is to ensure ecological surveillance at high spatio-temporal resolution and to serve as a basis
48 for knowledge-sharing within the context of development projects. Using these data in statis-
49 tical modeling presents a challenge due to sampling variability, which concerns the spatial and
50 temporal dimensions, and between species and observers because of their opportunistic nature
51 (Van Strien et al., 2013). This sampling variability translates into biases in species distribution
52 models that do not account for it. In this study, we propose to harness this extensive dataset
53 for modeling species habitat preferences, classifying species migratory status, and studying time
54 variations in relative abundance, by which we mean the ratio of the abundance of a species di-
55 vided by the cumulated abundance of a pool of species constituting the so-called target group
56 of species.

57 Various estimation frameworks and algorithms exist for constructing species distribution
58 models and calibrating their parameters from presence-only data. Widely used methods include
59 Maxent, based on the maximum-entropy principle (Phillips et al., 2006), and various statistical
60 regression modeling frameworks (Valavi et al., 2022). These methods are often based on point-
61 process representations of data (Renner et al., 2015; Warton and Shepherd, 2010), where each
62 observed occurrence of a species is viewed as a point in a cloud of points located in space and
63 time. In this context, models such as Maxent and logistic regression are equivalent to a special
64 case of the point-process setting (Fithian and Hastie, 2013; Renner and Warton, 2013). Due to
65 the uncertainty in reported positions in the Écrins National Park dataset, we rather implement a
66 degraded version of these point process models by aggregating occurrences into spatio-temporal
67 cells, which leads to the use of a generalized additive regression model.

68 Models including random effects are often used to capture the complex spatio-temporal
69 structures in data and ecological processes, thus allowing for an accurate assessment of vari-
70 ous sources of uncertainty. This has led to the widespread adoption of Bayesian inference tech-
71 niques such as Markov Chain Monte Carlo (MCMC, (Gilks et al., 1995; Link et al., 2002)) or In-
72 tegrated Nested Laplace Approximations (INLA, Illian et al., 2013). We apply the INLA approach
73 to perform complex Bayesian spatio-temporal inferences with relatively large datasets by com-
74 bining generalized additive regression models with the Stochastic Partial Differential Equation
75 (SPDE) approach, the latter used to represent spatial random effects (Lindgren et al., 2011). The
76 construction and estimation of complex spatio-temporal ecological models are still recent devel-
77 opments (Belmont et al., 2024; Soriano-Redondo et al., 2019), and as far as we know, our study
78 represents the first attempt to evaluate management policies of protected areas using solely
79 opportunistic presence-only data for this type of model.

80 **Materials and methods**

81 **Data**

82 *Passerines presence-only data.* The dataset comprises 102,513 opportunistically reported events
83 of passerine detections between 1994 and 2021, recorded by rangers of the Écrins National Park
84 with professional training in naturalistic expertise. We chose to focus our study on the *Passeri-*
85 *formes* because it is a taxon generally well known by all rangers, ensuring consistent reporting
86 rates, unlike more specific taxa. Furthermore, we have excluded three species – Common Raven
87 (*Corvus corax*), Red-billed Chough (*Pyrrhocorax pyrrhocorax*), and Alpine Chough (*Pyrrhocorax grac-*
88 *ulus*) – due to the high uncertainty associated with their reported positions compared to the rest

89 of the passerines. We narrowed the temporal window to the 1994–2021 period to ensure a
90 minimum of 2000 reported events per year (Appendix A, Lasgorceux et al., 2024c).

91 The data were collected opportunistically, by sight and/or ear, without a predefined survey
92 methodology. Each row of the dataset contains a position, a date, the name of the reported
93 species, the ID of the observer(s), and the number of individuals. However, we choose not to
94 consider the number of individuals because it can vary greatly due to species behavior, particu-
95 larly in gregarious species, and differences in observers' methods since some observers report
96 species abundance, while others only note presence (see the eBird example in Horns et al., 2018).
97 Thus, we consider one reported event as the presence of at least one reported individual, result-
98 ing in a total of 102,513 passerine occurrences.

99 Based on interviews with various rangers, we have gained insight that the reported posi-
100 tion could be i) an approximation of the reported individual's position at detection time, ii) the
101 observer's position at detection time, or iii) another, different position where the observer gath-
102 ered several occurrences before adding them jointly to the data records. Though, when multiple
103 occurrences are gathered at a station position, rangers ensure that the environment is consis-
104 tent between the station position and the bird's individual position at detection time and that
105 the distance between these two positions does not exceed 250m. This distance of 250m is ob-
106 tained by considering at most 100m for the difference between the positions of the detected
107 individual and the observer by sound recognition (Hauptert et al., 2023), plus an additional 150m
108 for the distance between the position at detection time and the station position. To account for
109 this uncertainty surrounding the reported position, we aggregated the data by assigning each
110 reported position to the center of a regular spatial grid with 500m resolution covering the Écrins
111 National Park. We excluded boundary cells with an area smaller than 25m², which corresponds
112 to the resolution of the environmental covariates we use for characterizing the spatial habitat
113 preferences. This discretization results in dividing the Écrins National Park domain into 11,260
114 spatial cells.

115 Our objective is not only to infer species habitat preferences but also to analyze temporal
116 trends in relative abundance. Therefore, we divided the 28 years of data into seven periods of
117 4 years each to capture long-term changes between such periods. Additionally, in order to take
118 into account that some studied species are migratory, we further subdivided the data tempo-
119 rally by month to investigate intra-annual variations. This spatio-temporal discretization scheme
120 results in a total of $7 \times 12 \times 11,260 = 945,840$ spatio-temporal cells.

121 *Passerine population trend synthesis in France.* To compare local and national trends in relative
122 abundance variations, we also utilized the French Breeding Bird Survey synthesis, known as the
123 STOC program, for the period 2001–2019 (Fontaine et al., 2020; Jiguet et al., 2012). These data
124 present the long-term abundance variation percentages of common birds in France during this
125 period. Comparing trends from opportunistic data and STOC data serves three main purposes:
126 (i) assessing the consistency of modeling assumptions, (ii) comparing local indicators within a
127 protected area to global indicators across the territory of the entire country, and (iii) providing
128 insights into space-time population trends for species specialized in mountainous habitats, for
129 which surveys such as STOC are quite limited in terms of spatio-temporal coverage.

130 *Environmental Data.* The Écrins National Park is a high-mountain park where the terrain struc-
131 ture plays a predominant role in the ecosystems. We utilized the Digital Elevation Model data

132 from RGE (IGN, 2018) to extract the elevation (McVicar and Körner, 2013) on a 5m regular grid.
133 Subsequently, we calculated the slope at the same resolution using the ‘terrain’ function from
134 the R package *terra* (Hijmans, 2024). Ultimately, we derived two spatial covariates: the mean
135 elevation and the mean slope values per spatial cell at 500m resolution.

136 Regarding land cover and land use data, we make use of the OSO dataset (Inglada et al., 2018).
137 Originally at a resolution of 10/20m, we applied the nearest neighbor strategy to downscale the
138 values to a 5m resolution, for alignment with the RGE data resolution. To avoid statistical issues
139 due to uncommon land cover and land use types in the Écrins National Park, we aggregated
140 data towards 10 representative categories: Urban, Crops, Meadows, Deciduous forest, Conif-
141 erous forest, Grasslands, Woody Heaths, Mineral Surfaces, Glaciers and snow, and Water (see
142 Appendix B in Lasgorceux et al., 2024c for further details). We then calculated the percentage
143 of coverage of each category in each spatial cell at 500m resolution.

144 We also used the Historical Monthly Weather Data from WorldClim (CRU-TS 4.06 (Harris
145 et al., 2020) downscaled with WorldClim 2.1 (Fick and Hijmans, 2017)) at 2.5 minutes spatial
146 resolution (roughly 5km) between the years 1994 and 2021. The available variables were aver-
147 age maximum temperature (°C) and total precipitation (mm). We extracted the corresponding
148 values in each spatial cell, for each month, and each year. Since the inter-annual resolution of
149 our spatio-temporal cells consists of periods of 4 years, we computed the mean of the extracted
150 values of the 4 corresponding years for a given spatial cell and a given month.

151 *Preprocessing of model covariates.* Elevation shows strong variability in the Écrins National Park
152 with certain climate and land cover configurations arising predominantly within specific eleva-
153 tion ranges. As a result, both land cover and climatic variables are correlated.

154 Hence, we began by decomposing the climatic variables into their spatial means and spatio-
155 temporal anomalies with respect to those means. We then performed a Principal Component
156 Analysis (PCA, Janžekovič and Novak, 2012) using the R package *ade4* (Dray and Dufour, 2007).
157 The input variables were: the mean elevation, the mean slope, the percentages of coverage com-
158 puted for each of the ten representative categories derived from the OSO classification, as well
159 as the spatial means of average maximum temperature and total precipitation. The first 6 axes
160 of the PCA, accounting for 75% of the information, allow for useful interpretation and have
161 therefore been incorporated into the species distribution models (see Appendix C in Lasgorceux
162 et al., 2024c for details). The first principal component axis explains a significant 31.1% of the
163 variation in habitat types, primarily driven by high-elevation environments. The second axis is
164 characterized by vegetation at medium to high altitudes, particularly in coniferous forests, espe-
165 cially in steep areas. The third axis is influenced by grasslands, and the fourth axis is primarily
166 driven by wetlands and glaciers. However, since we observe a significant predominance of bird
167 occurrences in “water” habitats within urbanized valleys (Figure 14 in Appendix C, Lasgorceux et
168 al., 2024c) compared to “glaciers and snow” habitats (751 occurrences versus 80), we interpret
169 the fourth axis as “urban wetlands”. The interpretation of the fifth axis is also associated with
170 high-elevation vegetation, as it is primarily driven by dense, low-growing shrubs and small trees;
171 henceforth, we will refer to these environments as “woody heaths”. The sixth axis is driven by
172 low-elevation deciduous forests.

173 This approach results in the implementation of eight covariates in our statistical model: six
174 spatial covariates from the PCA analysis, and two spatio-temporal covariates defined as the
175 anomalies of average maximum temperature and anomalies of total precipitation.

176 Statistical modeling

177 We applied the same Bayesian hierarchical model (Wikle, 2003) to each species with more than
 178 100 occurrences over the period from 1994 to 2021, excluding *Bombycilla garrulus* for which
 179 90% of its 146 occurrences were reported in 2005, due to a brief invasion event of this species
 180 within the Écrins National Park that year. This resulted in a total of 76 species being analyzed.

181 *Data model.* We denote by s a spatial cell, by m a month, and by p a temporal period of 4 years,
 182 with the triplet (s, m, p) representing the corresponding space-time cell. The response variable,
 183 denoted as $Y_i(s, m, p)$, is the number of occurrences of species i reported for that cell in the op-
 184 portunistic presence-only dataset. We model $Y_i(s, m, p)$ with a hierarchical generalized additive
 185 mixed model (Knappe, 2016; Pedersen et al., 2019) using a Poisson response distribution:

$$(1) \quad Y_i(s, m, p) \mid \mu_i(s, m, p) \sim \text{Poisson}(\mu_i(s, m, p)),$$

186 where $\mu_i(s, m, p)$ is the average number of reported occurrences of species i in the cell (s, m, p) .

187 We assume that $\mu_i(s, m, p)$ is obtained through the combination of two factors: (i) the abun-
 188 dant of species i in the cell (s, m, p) , denoted by $\Lambda_i(s, m, p)$, and (ii) the sampling effort in cell
 189 (s, m, p) for species i , denoted by $E_i(s, m, p)$, which we assume to both act in a multiplicative
 190 way on $\mu_i(s, m, p)$ (Giraud et al., 2016). These relationships lead us to formulate the following
 191 equation:

$$(2) \quad \mu_i(s, m, p) = \Lambda_i(s, m, p) \times E_i(s, m, p)$$

192 *Sampling effort.* A major challenge is now to quantify or eliminate the term representing the
 193 sampling effort. Within our spatio-temporal discretized framework, the first step is to obtain in-
 194 formation about where and when no observations occurred. This helps distinguish between cells
 195 with zero occurrences due to the absence of observer presence and those with zero occurrences
 196 due to the absence of the species. The strategy involves utilizing the extensive opportunistic
 197 presence-only dataset, using reported occurrences of certain species as indicators of observer
 198 presence, as proposed by Botella et al., 2020. However, it is necessary to assume that these
 199 species are uniformly sampled by all the observers, corresponding to homogeneity in reporting
 200 rates across the species pool, as well as in the probability of detection. This explains one of our
 201 reasons to primarily restrict our study to passerines (see the subsection *Passerine data*). Following
 202 existing literature, we refer to this set of chosen species with homogeneous sampling conditions
 203 as the target group (TG) of species. Therefore, we denote the sampling effort term as E instead
 204 of E_i , and we write $\Lambda_{TG}(s, m, p)$ for the cumulative abundance of all species in the target group
 205 in cell (s, m, p) . We can then reformulate the average number of occurrences in terms of the
 206 ratio of the abundance of species i to the cumulative abundance of the target group, and the
 207 product of the sampling effort with the cumulative abundance of the target group.(Equation 4):

$$(3) \quad \mu_i(s, m, p) = \Lambda_i(s, m, p) \times E(s, m, p)$$

$$(4) \quad = \frac{\Lambda_i(s, m, p)}{\Lambda_{TG}(s, m, p)} \times \Lambda_{TG}(s, m, p) E(s, m, p).$$

208 To statistically identify the first term that indicates the abundance of species i relative to
 209 the cumulative abundance of the target group of species –hereafter referred to as relative
 210 abundance– we need a proxy for the second term $\Lambda_{TG}(s, m, p) E(s, m, p)$. The most natural ap-
 211 proach is to use the number of target-group reported occurrences, denoted as $y_{TG}(s, m, p)$. We

212 chose to use the combined occurrences of all passerine species, which was consistent with al-
 213 gorithms similar to that of Botella et al., 2020 since they were all selected except for 4 species
 214 observed only once. We kept only the cells with at least one target-group occurrence, reducing
 215 the number of studied cells to 32,486.

216 *Relative abundance.* Employing a spatio-temporal distribution model enables the implementa-
 217 tion of various spatio-temporally structured latent fixed or random effects within the linear struc-
 218 ture that we will assume for the predictor $\log(\mu_i(s, m, p))$. We opted to use only one generic
 219 model for each of the 76 species to automate model fitting and facilitate inter-species compar-
 220 isons. In Poisson models, it is common practice to represent the Poisson intensity, denoted here
 221 as $\mu_i(s, m, p)$, using a log-link function to ensure that $\mu_i(s, m, p)$ remains positive. Focusing on
 222 the parameter of interest, the relative abundance $\frac{\Lambda_i(s, m, p)}{\Lambda_{TG}(s, m, p)}$, we constructed our linear regressor
 223 by incorporating the proxy of the sampling effort, $y_{TG}(s, m, p)$, as a fixed offset (Equation 5).

224 By studying the abundance of a passerine species compared to all others, we compare species
 225 with different biological characteristics, such as habitats and migratory status, which need to be
 226 taken into account in the linear regression. We therefore allowed for a linear effect of the spatial
 227 covariates $\mathbf{X}^{(PCA)}(s)$ (PCA components) and the spatio-temporal covariates $\mathbf{X}^{(anomalies)}(s, m, p)$
 228 (anomalies of climatic covariates) to estimate habitat preferences of species i , relative to other
 229 passerines. We denoted α_i as the intercept. The term $W_i^{(m)}(s)$, $m = 1, \dots, 12$, represents a
 230 monthly latent spatio-temporal Gaussian field. Concretely, this field aims to model intra-annual
 231 spatial variations not captured by the covariates, typically variations in relative abundance due
 232 to the migration of certain passerines within the territory of the Écrins National Park. Lastly, we
 233 denote by $f_i(p)$ a non-spatialized inter-annual effect, aiming to capture the variations in relative
 234 abundance across each period p of 4 years.

$$(5) \quad \log(\mu_i(s, m, p)) = \log(y_{TG}(s, m, p)) + \log\left(\frac{\Lambda_i(s, m, p)}{\Lambda_{TG}(s, m, p)}\right)$$

$$(6) \quad = \log(y_{TG}(s, m, p)) + \alpha_i + \mathbf{X}^{(PCA)}(s)\beta_i + \mathbf{X}^{(anomalies)}(s, m, p)\gamma_i + W_i^{(m)}(s) + f_i(p).$$

235 *Statistical inference.* We employed the *R-INLA* package for the statistical inference of our models
 236 (Rue et al., 2009). INLA is a fast but accurate estimation method based on deterministic Laplace
 237 approximations –by contrast with approaches based on stochastic simulation, such as MCMC–
 238 and requires that the latent layer ($\log(\mu_i(s, m, p))$ in our model) follows a multivariate Gaussian
 239 distribution, possibly in very high dimension, as with our space-time fields. In this setting, INLA
 240 enables the implementation of a wide range of Gaussian random effects, including spatial fields
 241 implemented through the SPDE approach, which provides a numerically convenient representa-
 242 tion of the Matérn covariance function used routinely in spatial statistics (Lindgren et al., 2011),
 243 and of priors for hyperparameters such as variances or correlation ranges. In Appendix D, we
 244 precisely describe the prior distributions of all parts of $\log(\mu_i(s, m, p))$, i.e., the structure of the
 245 model components before updating them with information from observation data to obtain their
 246 posterior distributions (Lasgorceux et al., 2024c).

247 Validation and post-processing

248 The model presented previously is quite comprehensive, allowing for a thorough examination
 249 of information on a species-by-species basis. In the following sections, we delineate the post-
 250 processing phase undertaken to synthesize this information. We address questions pertaining

251 to the model's validation and also delve into the three components of the linear predictor (Equa-
252 tion 6): those concerning the fixed effects – specifically, habitat preferences compared to the
253 target group denoted by β_i and γ_i ; the spatial intra-annual component, characterized by the
254 month-dependent Gaussian field W_i^m , which is associated with migratory status; and the inter-
255 annual component, represented by the random i.i.d. Gaussian effect $f_i(p)$, which captures long-
256 term trends in relative abundance.

257 *Computational cost.* For each species, the model comprises 32,486 space-time cells with obser-
258 vations and a total of 7,676 latent variables to estimate across the prediction cells. We conducted
259 analyses using the R programming language (R Core Team, 2024) and utilized version 24.02.09
260 of *R-INLA* (Rue et al., 2009), leveraging the PARDISO 8.2 library to accelerate computations
261 (Eftekhari et al., 2021; Gaedke-Merzhäuser et al., 2023; Pasadakis et al., 2023). This configu-
262 ration typically yields computation times averaging around one minute for each model fit on a
263 standard workstation with 2.80GHz processor.

264 *Validation.* We evaluated whether our model effectively extracted meaningful insights from the
265 opportunistic presence-only data by comparing its predictive performance against a baseline
266 that considered only the sampling effort proxy, *i.e.* the target-group occurrences as an offset. We
267 focus on the prediction of the binary event of the reported presence or non-reported presence of
268 a species in a given cell and month and compute the Area-Under-the-Curve (AUC, Fawcett, 2006;
269 Jin Huang and Ling, 2005) (where the curve is the so-called Receiver Operating Characteristic
270 curve) to obtain a simple scalar prediction score for each species, where we compare results for
271 two predictors: (i) the sampling effort proxy offset and (ii) the full linear predictor (Equation 6).

272 Next, we assessed the model's generalization skill for predicting new data using k-fold, spatial,
273 and temporal cross-validation scenarios. In k-fold cross-validation, the dataset is divided into k
274 equal parts, with one part used for validation while the rest is used for training (Jung, 2018). This
275 approach yields $k + 1$ AUC scores per species: k scores from predicting each part separately,
276 and one from predicting the entire dataset. We selected $k = 10$ to ensure comparability with
277 temporal and spatial scenarios, where the training data represented between 3.9% and 13.7% of
278 the total volume. This choice aimed to maintain consistency in the quantity of test data across
279 scenarios, ensuring equitable evaluation conditions. In the spatial scenario, we assessed model
280 robustness by alternately removing all spatio-temporal cells within a given municipality of the
281 Écrins National Park. Seven municipalities were selected based on geographic zones and data
282 availability, resulting in seven AUC scores per species. For the temporal scenario, we alternately
283 randomly removed half of the data for a given period, resulting in seven AUC scores per species.
284 Further details on the size of the training datasets are given in Appendix E in Lasgorceux et al.,
285 2024c.

286 *Fixed effects.* We estimated the *a posteriori* distribution of coefficients β_i and γ_i separately for
287 each species i (Equation 6). The posterior mean of each parameter and its sign quantify the
288 importance of the covariates in explaining occurrences of species i relative to the occurrences
289 of all the passerines taken together. The credibility interval of each parameter helps determine
290 how strongly the corresponding covariate influences the species; if the interval does not contain
291 zero, we can consider the covariate as being “significant” (by analogy with the frequentist notion
292 of significance). To synthesize the estimated mean coefficients and assess parameter similarities
293 across species occupying comparable ecological habitats, we conducted Principal Component

294 Analysis (PCA) on estimated parameters followed by species clustering based on the estimated
 295 mean effects. This clustering was performed using a hierarchical agglomerative clustering (HAC)
 296 algorithm implemented through the *hclust* function, employing the *ward.D2* method (Murtagh
 297 and Legendre, 2014).

298 *Month-based spatio-temporal Gaussian field.* The month-based spatio-temporal Gaussian field
 299 $W_i^{(m)}$ adds flexibility to the model but is challenging to summarize graphically since we have
 300 12 maps per species. Therefore, we opted to reduce dimensionality by spatially averaging each
 301 field. For each species, this yields a curve of 12 points depicting the intra-annual variations in oc-
 302 currences relative to the target group. Instead of plotting all 76 curves on a single graph, we con-
 303 ducted a Functional Principal Component Analysis (FPCA, Ramsay and Silverman, 2005). In our
 304 case, FPCA decomposes the intra-annual mean effect of a given species into (i) the mean intra-
 305 annual effect across all species plus (ii) the remaining signal represented by scores associated
 306 with empirically identified harmonics. In a functional principal component analysis, harmonics
 307 represent the most significant modes of variation in the observed functions. These harmonics
 308 are akin to principal components in traditional PCA but are applied to functions rather than
 309 conventional variables, where the difference is that functions are represented as vectors with a
 310 very large number of components (larger than the sample size of functions), and typically show
 311 correlation for components at near positions within the vector. We utilized the *fda* package to
 312 conduct these analyses (Ramsay et al., 2009).

313 *Inter-annual effects.* Our motivation for introducing the inter-annual effect in the linear predictor
 314 (Equation 6) lies in comparing the relative abundance trends of passerines in the Écrins National
 315 Park with the abundance trends of common birds in the STOC. If we denote by $\lambda_i^{(STOC)}(y)$ the
 316 estimated abundance of species i in year y from the STOC, then the percentage variation in
 317 abundance computed by the STOC (Fontaine et al., 2020) can be expressed as

$$\left(\hat{\lambda}_i^{(STOC)}(2019) - \hat{\lambda}_i^{(STOC)}(2001) \right) \times \frac{100}{\hat{\lambda}_i^{(STOC)}(2001)}.$$

318 In the model presented in Equation 6, the term capturing the effect of the period p on the
 319 relative abundance of species i is $\exp(f_i(p))$. The INLA method allows us to estimate a posterior
 320 distribution of each $\exp(f_i(p))$. To mimic the STOC period of study, we assume the relative abun-
 321 dance to be constant in each period p . This allows us to attribute an estimated relative effect
 322 $f_i(y)$ to each year y between 2001 and 2019. We then calculated a regression line using a linear
 323 model with the $f_i(y)$ as the response variable and y the explanatory variable. If we denote by
 324 $\hat{f}_i^{(ENP)}(y)$ the fitted value of this linear regression, we could calculate the percentage variations
 325 in relative abundance of species i as

$$\left(\hat{f}_i^{(ENP)}(2019) - \hat{f}_i^{(ENP)}(2001) \right) \times \frac{100}{\hat{f}_i^{(ENP)}(2001)}.$$

326 To enhance robustness and take into account statistical uncertainty, we generated 1000 pos-
 327 terior distributions for each relative abundance trend parameter and repeated this procedure.
 328 Then, we used the median percentage variation in relative abundance as a reference for each
 329 species and compared it to the STOC.

Results

330

331 Validation

332 Figure 1 displays the boxplots of AUC values computed species-by-species as described in sec-
333 tion *Validation and post-processing*. There is a clear improvement in AUC when using the full
334 linear predictor compared to using only the sampling-effort proxy as offset. This confirms that
335 sampling methods do not solely constrain opportunistic presence-only data and that valuable
336 information can be extracted from them.

337 The AUC values of models with cross-validation (with one value for each hold-out dataset)
338 are typically slightly lower than those without cross-validation. This is expected due to the re-
339 duction in training data and the more challenging prediction setting with new data not used
340 during training. Despite this slight decrease, the AUC values generally stay high, with approxi-
341 mately 75% of values above 0.8 across all scenarios. Consequently, the model demonstrates a
342 very good fit to the data, except for two species: *Cinclus cinclus* and *Motacilla cinerea*, which are
343 consistent outliers in all scenarios. These two species with poor AUC values relative to others
344 are stream specialist species, for which covariates do not capture the habitat well so we have
345 excluded them from further analysis.

346 Fixed Effects

347 The mean estimated effects of each covariate and their significance are detailed in Tables 5-7
348 in Appendix F (Lasgorceux et al., 2024c). To summarize the information on habitat spatial co-
349 variates, we provide a classification into three clusters based on Principal Component Analysis
350 conducted on the estimated mean coefficients for each habitat spatial covariate (see subsection
351 'Materials and Methods – Validation and Post-processing – Fixed Effects'). We did not include
352 any species with less than two significant covariates since we consider that the model does not
353 provide enough information on them. These species are generally those with very few occur-
354 rences, as 10 out of these 11 species have fewer than 551 occurrences over the 28 years of
355 data, where the value 551 is below the 25th-percentile of species occurrences. The list of re-
356 moved species is given in Table 8 in Appendix F (Lasgorceux et al., 2024c). The PCA reveals two
357 principal axes explaining 62.2% of the classification: the first axis distinguishes vegetated habi-
358 tat (coniferous and deciduous forests VS urban wetlands), while the second axis highlights an
359 altitudinal gradient.

360 The three clusters are presented in Figure 2. We interpret them as follows. Cluster 1: Species
361 mostly found in closed habitats such as coniferous and deciduous forests (green); Cluster 2:
362 Species mostly found in high-elevation environments (orange); Cluster 3: Species mostly found
363 in open habitats and valleys (violet). These results are coherent when comparing a map of eleva-
364 tion with a map of clusters with the highest occurrences on a 500m×500m regular grid in the
365 Écrins National Park (see Figure 15 in Appendix F in Lasgorceux et al., 2024c).

366 The ellipse associated with forest species represents an altitudinal gradient, with species
367 found in low-elevation forests (e.g. *Pyrrhula pyrrhula*), to medium (e.g. *Anthus trivialis*) and high-
368 elevation forests (e.g. *Turdus torquatus*). We also observe this altitudinal gradient in the cluster
369 associated with species from open habitats and valleys, with rocky habitat species such as *Ti-*
370 *chodroma muraria* and *Phoenicurus ochruros*, at the boundary of the high-elevation species clus-
371 ter. The cluster associated with species specialized to high-elevation habitats is smaller but well-
372 defined, with typical species such as *Prunella collaris* and *Montifringilla nivalis*.

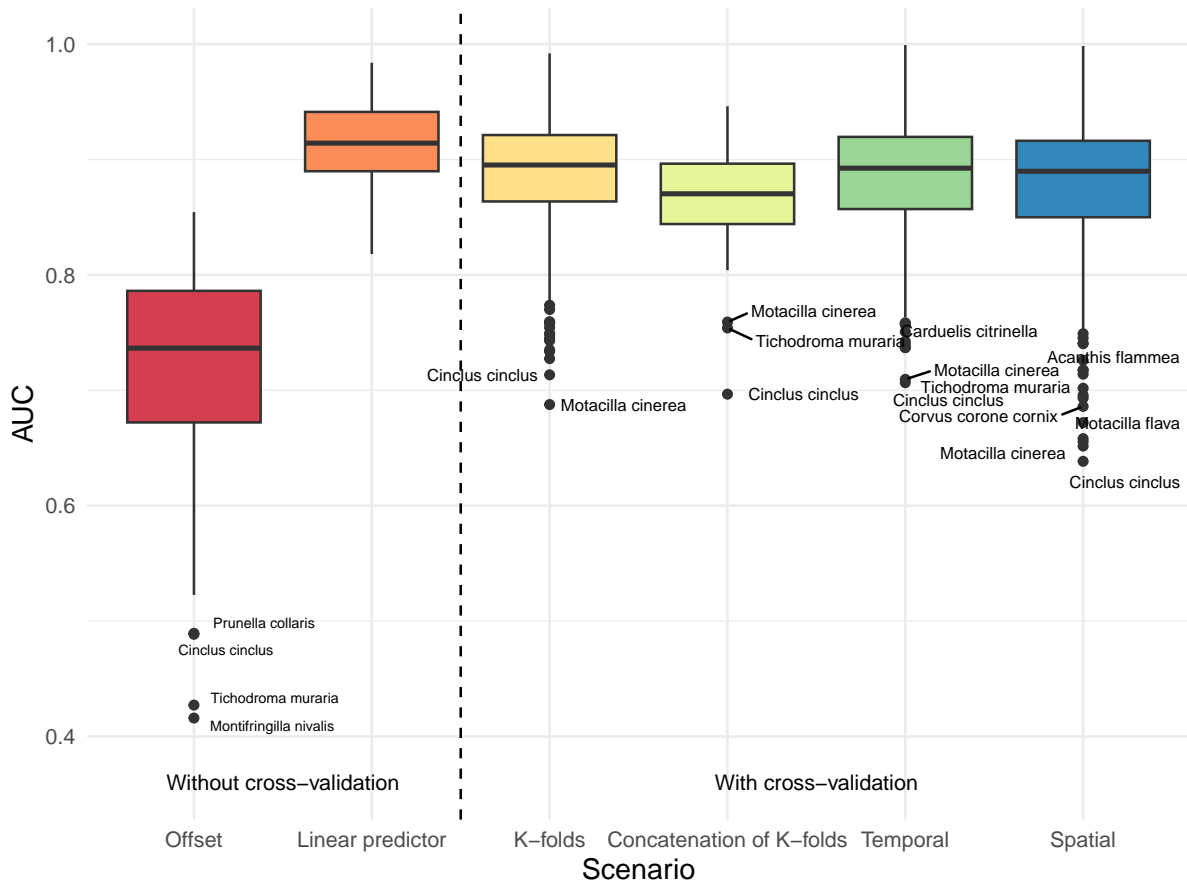


Figure 1 – AUC values calculated for various linear-predictor and cross-validation scenarios. Boxplots summarize the set of AUC values for all species, and in the case of cross-validation for each combination of species and hold-out dataset. Notations are as follows: “Offset” – model where linear predictor contains only the offset; “Linear predictor” – model with full linear predictor; “K-folds” – model fitted by randomly removing 10% of the cells, ten times; “Concatenation of K-folds” – AUC values computed by concatenating the predictions of the K-folds scenarios; “Temporal” – model fitted by randomly removing half of the data in a given time period, for each time period; “Spatial” – model fitted by removing all the data which occurred in a given city, for seven cities. The species with multiple outlier AUC values in “K-folds”, “Temporal” and “Spatial” scenarios are only noted once. Details on the four cross-validation scenarios (where models always have the full linear predictor) can be found in the ‘Materials and Methods – Post-processing and Validation – Validation’ section.

373 To compare the contribution of spatio-temporal covariates with spatial ones in the linear pre-
 374 dictor, we computed the percentage of explained variance for each covariate with respect to
 375 the overall variance of the linear part $(\alpha_i + \mathbf{X}^{(PCA)}(s)\beta_i + \mathbf{X}^{(anomalies)}(s, m, p)\gamma_i)$ for each species.
 376 The results are presented in Figure 3. Our analysis revealed that, for most species, PCA Axis 1,
 377 interpreted as being related to high-elevation environments, explains the majority of the vari-
 378 ance (Median=54%). Axis 2 and 3, interpreted as environments with medium to high altitude
 379 vegetation and grasslands, respectively, explain a substantial portion of the variance for about
 380 twenty species (Q3=27% and 15%).

381 Regarding spatio-temporal covariates, precipitation anomalies had the smallest percentage
 382 of explained variance, with a median below $10^{-5}\%$. This is further supported by only 12 out of 76

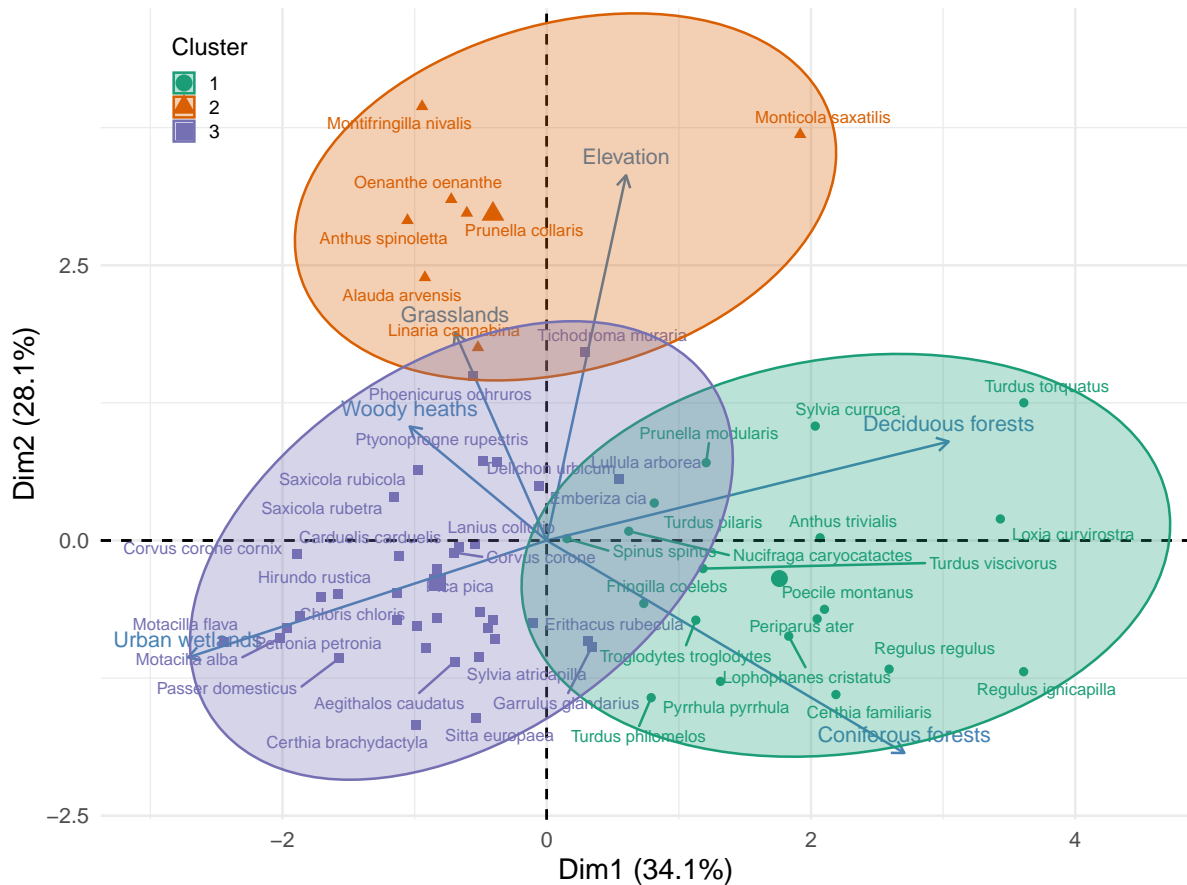


Figure 2 – Classification of 63 species in three clusters based on the estimated mean effects of spatial covariates. A Principal Component Analysis was conducted on the species with at least two credibility intervals that do not include 0. The two main axes explain 62.2% of the variance. Cluster 1: Species mostly found in closed habitats such as coniferous and deciduous forests (green); Cluster 2: Species mostly found in high-elevation environments (orange); Cluster 3: Species mostly found in open habitats and valleys (violet).

383 species showing a significant parameter associated with precipitation anomalies. Similarly, the ex-
 384 plained variance for maximum temperature anomalies is low for most species (Median 10^{-3});
 385 however, 45 species showed a significant parameter. It notably explained a significant amount
 386 of information for five species: *Coccothraustes coccothraustes*, *Fringilla montifringilla*, *Lanius col-*
 387 *lurio*, *Spinus spinus*, and *Turdus pilaris*. These results align with the ecological characteristics of
 388 these species, confirming that the model effectively correlates temperature anomalies with the
 389 presence of thermophilic and cold-resistant species.

390 Intra-annual effect

391 The first two extracted harmonics capture a significant portion of the intra-annual variability,
 392 explaining 83% of the variance. The primary harmonic shows a stronger positive effect in
 393 winter months compared to summer months, while the second harmonic has a positive effect in
 394 spring and summer and a negative effect in winter (see Figure 16 in Appendix G in Lasgorceux
 395 et al., 2024c). Thus, species with positive scores on the first harmonic, referred to as the winter
 396 harmonic, are predominantly overrepresented in winter. Conversely, species with positive
 397 scores on the second harmonic, now termed the warm months harmonic, are more likely to be

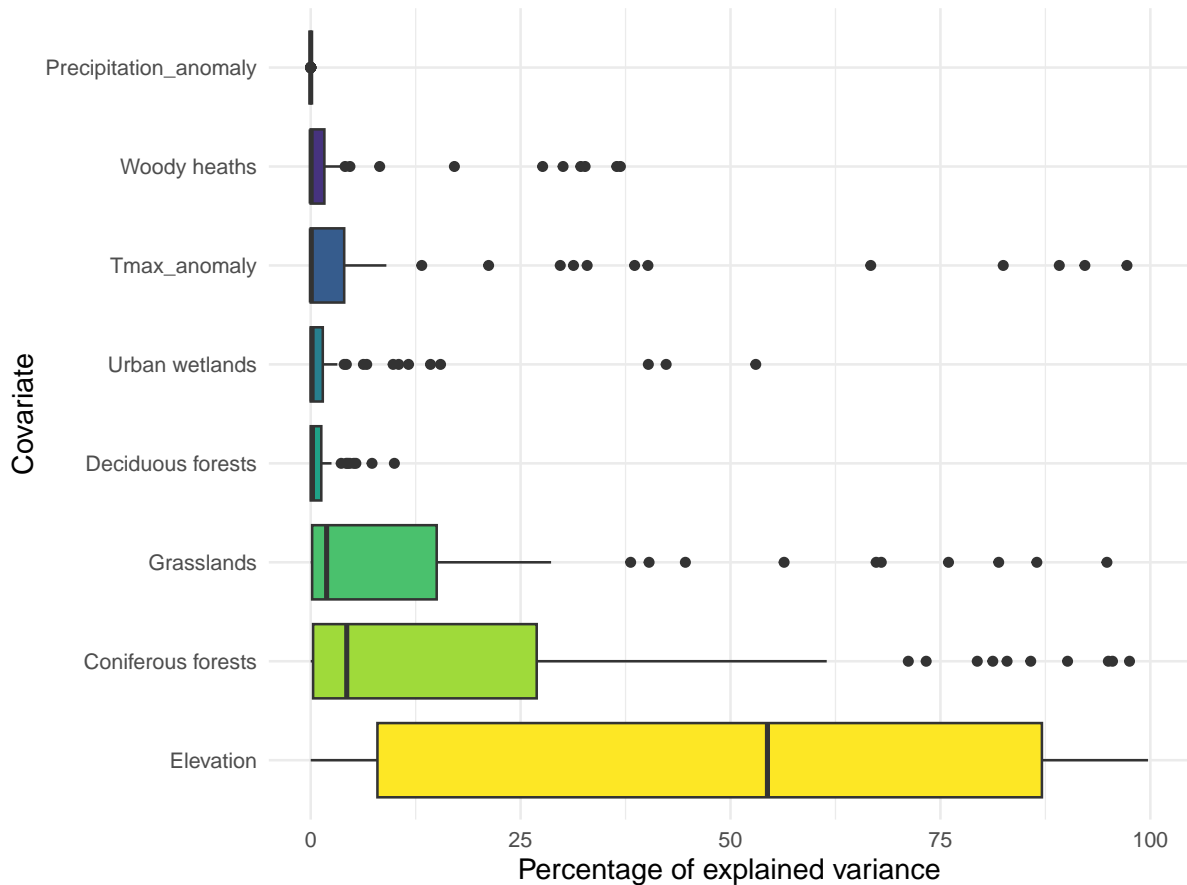


Figure 3 – Percentage of explained variance of each covariate for 74 species. The covariates include six spatial factors derived from PCA analysis and two spatio-temporal climatic factors. The interpretation of the PCA axes is detailed in the ‘Materials and Methods – Data – Preprocessing of model covariates’. Covariate boxplots are ordered by their median values.

398 overrepresented in spring and summer, suggesting a potential link with migratory status.

399

400 Figure 4 displays the harmonic scores for each species, categorized by migratory status within
 401 the territory of the Écrins National Park, as provided by experts from the park. This analysis con-
 402 firms that species overrepresented in winter are mainly sedentary, while those overrepresented
 403 in warm months are migratory. Detailed scores for each species are provided in Table 9 in Appen-
 404 dix G (Lasgorceux et al., 2024c). These harmonics effectively distinguish between sedentary and
 405 migratory species in the Écrins National Park, affirming that the month-based spatio-temporal
 406 Gaussian field $W_i^{(m)}$ accurately accounts for spatial and intra-annual species variations not ex-
 407 plained by physical covariates across the territory.

408 **Inter-annual variability and STOC comparison**

409 The estimated percentage variations in relative abundance in the Écrins National Park between
 410 2001 and 2019, based on 1000 *a posteriori* samples, are illustrated in Figure 5. Species associ-
 411 ated with the high-elevation habitats cluster (Figure 2) exhibit a mean decline of –16.5% in rel-
 412 ative abundance during 2001–2019. This is mainly due to the strong decrease of three species:

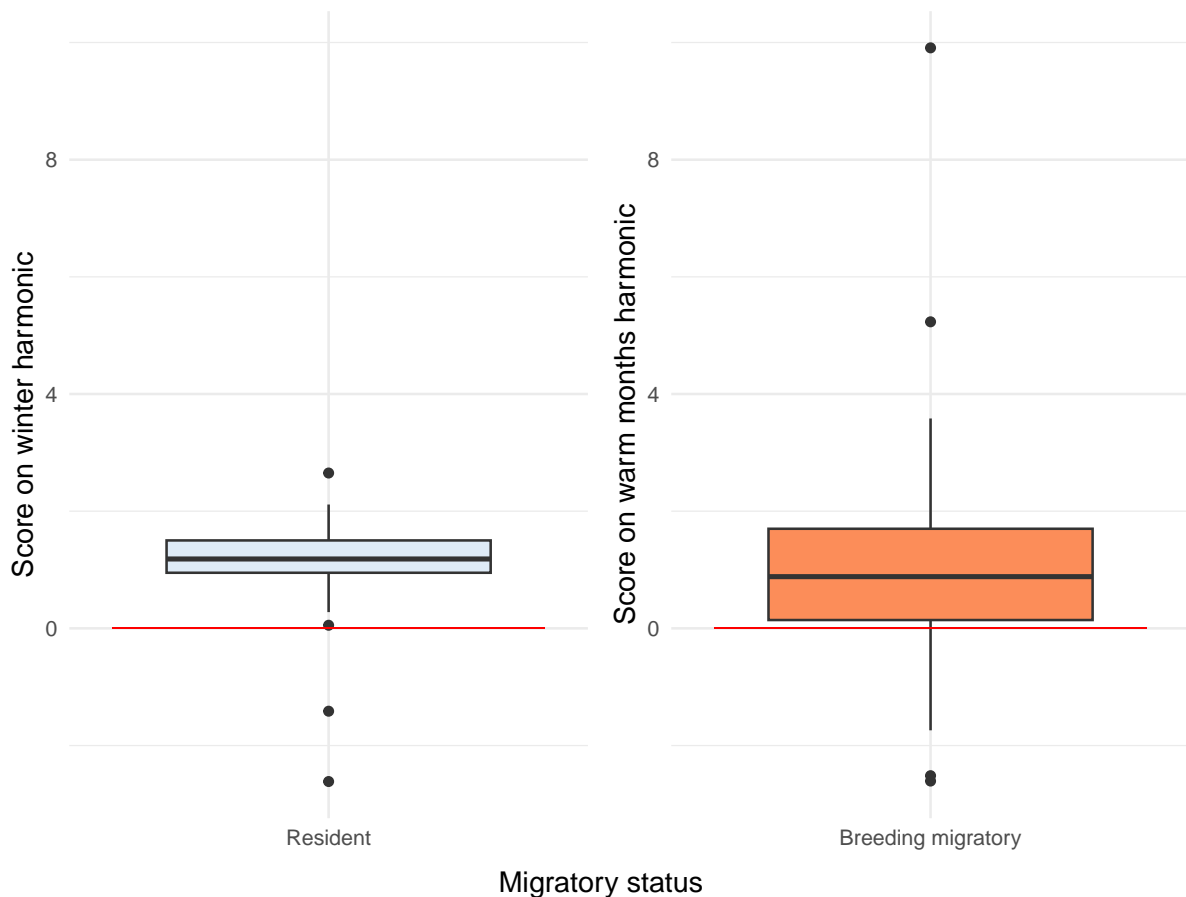


Figure 4 – Point plot of scores associated with the first two harmonics of the FPCA computed on the month-based spatio-temporal Gaussian field $W_i^{(m)}$ estimated for each species. The two first harmonics explain 83% of the variance. The species are categorized by their main migratory status within the territory of the Écrins National Park (ENP), as provided by experts from the park: breeding migratory (red); resident (green); wintering (blue).

413 *Monticola saxatilis* (–53.5%), *Prunella collaris* (–26.4%) and *Montifringilla nivalis* (–23.4%). Con-
 414 versely, species affiliated with the forest cluster experienced a growth of 12.4% over the same
 415 period. Figure 5 shows that, out of the 19 forest species, 6 exhibited a positive trend while only 2
 416 showed a negative trend based on the 95% credibility interval. For comparison, it is noteworthy
 417 that the STOC reported a decrease of 9.7% in abundance for specialist forest species in France.
 418 When focusing specifically on the passerines considered as forest specialists by the STOC, the
 419 difference becomes even more pronounced, with a growth of 23.2% in the Écrins National Park,
 420 while the STOC observed a growth of 2% for these species during the same period in France.
 421 It is important to note that this percentage is nevertheless skewed by *Regulus ignicapilla*, which
 422 shows a growth of 168% in relative abundance over the period.

423 Species linked to the valleys cluster experienced a mean decrease of -3.2% in relative abun-
 424 dance. However, comparing this figure with STOC trends is challenging due to the mix of gener-
 425 alist species with specialists of built and agricultural environments within this cluster. To address
 426 this, we compare below the trends of passerines in the Écrins National Park, classified by the
 427 STOC as generalist species or species specialized in built and agricultural environments, with
 428 our estimated trends in relative abundance. Built-environment passerines demonstrate stability

429 in the Écrins National Park (0.26%) compared to a slight decrease in France (-5.68%). General-
 430 ist passerines show an increase of 7.4% in relative abundance in the park, aligning with the 7%
 431 growth reported by the STOC for these species (and 19.4% for generalist common birds, not
 432 only passerines). For farmland species, both show a decrease, but at a slower rate of -16.5%
 433 compared to -22.6% (and -29.5% for farmland common birds, not only passerines).

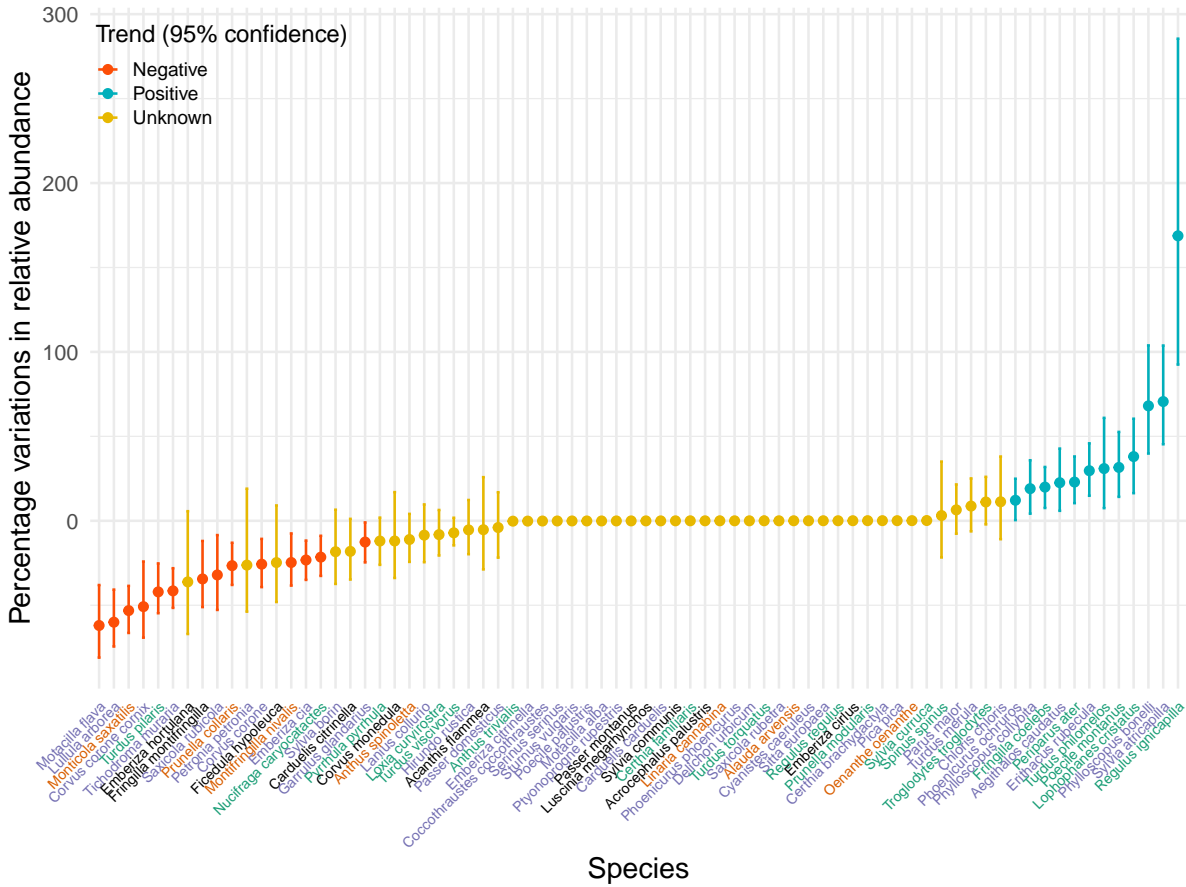


Figure 5 – Illustration of the estimated median percentage variations in relative abundance and credibility intervals for 74 retained species in the Écrins National Park spanning the period 2001–2019. These trends are derived from 1000 samples of the *a posteriori* distribution of the inter-annual effect $\exp(f_i(p))$. Details can be found in the ‘Materials and Methods – Post-processing and Validation – Inter-annual effects’ section. Credibility intervals shown in red (respectively blue) correspond to intervals below (respectively above) zero, while yellow intervals include zero. Species names are color-coded according to their cluster affiliation in Figure 2. Species with fewer than two significant covariates were not classified and are shown in black.

434 Figure 6 displays more precisely the percentage variations of relative abundance in the Écrins
 435 National Park against the percentage variations of abundance in France computed by the STOC,
 436 between 2001 and 2019. More details on the trends are given in Table 10, Appendix H (Las-
 437 gorceux et al., 2024c). Each species lies in one of the four quadrants. Quadrants (ii) and (iii)
 438 correspond to species whose trends locally within the Écrins National Park align with those in
 439 France: increasing for quadrant (ii) and decreasing for quadrant (iii). Quadrant (i) corresponds to
 440 species trending upward nationally but downward in the Écrins National Park, whereas quad-
 441 rant (iv) represents species increasing in the Écrins National Park but decreasing nationally. The
 442 number of species in each quadrant is as follows: 8 for (i), 14 for (ii), 24 for (iii) and 13 for (iv).

443 Quadrant (iv) is primarily composed of forest species, while quadrant (i) consists mainly of gen-
 444 eralist species, particularly common ones.

445 Furthermore, while the national park status does not prevent the decline of certain species,
 446 it is noteworthy that for 38 species out of 59, trends are more favorable within the national park
 447 compared to national trends for all of France. The names of these species are represented under
 448 the diagonal in Figure 6. However, some species for which this is not the case are very common
 449 (quadrant (ii)) and may be subject to reporting bias (see Discussion).

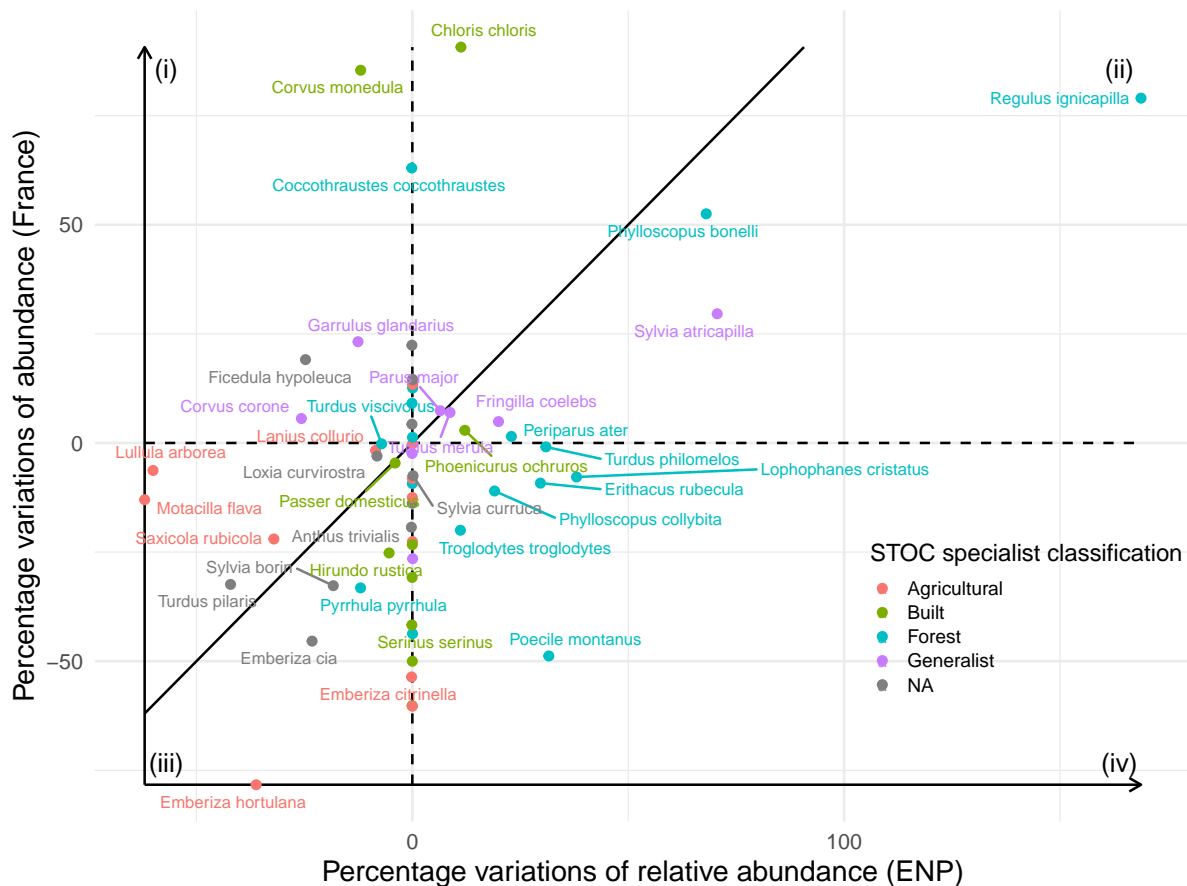


Figure 6 – Point plot of the estimated median percentage variations of relative abundance in the Écrins National Park between 2001 and 2019, against the percentage variations of abundance in France between 2001 and 2019 computed by the STOC for 59 species. The names of species with less than 0.1% variation in relative abundance are not plotted for readability.

450

Discussion

451 In our study, we demonstrated that under the assumption of sampling homogeneity across
 452 species (but where sampling effort is still allowed to vary in space and time), opportunistic
 453 presence-only data could be leveraged to estimate habitat preferences, migratory status, and
 454 relative abundance trends of passerine species. This finding is not novel *per se*, as numerous
 455 studies have utilized such data to model species distributions or abundance trends (Botella
 456 et al., 2020; Bradter et al., 2018; Kéry et al., 2010; Phillips et al., 2009; Valavi et al., 2022;
 457 Van Strien et al., 2013). However, it is one of only a few approaches achieving this within a
 458 unique spatio-temporal model. To our knowledge, the most similar work has been done by

459 Giraud et al., 2016 and Coron et al., 2018 modeling spatial relative abundance (abundance
460 divided by the abundance of an arbitrarily chosen site) of passerines using a Poisson distribution.
461 Their work provides theoretical results regarding parameter identifiability, as well as scenarios
462 for data simulation and parameter estimation. However, the use of their models requires
463 access to appropriate survey data based on a standardized sampling protocol, whereas our
464 study relied solely on a single set of opportunistic data. We also found the inclusion of the
465 month-based spatio-temporal latent Gaussian field $W_i^{(m)}$ and the use of Functional Principal
466 Component Analysis to be pertinent for assessing phenological effects when modeling relative
467 abundances of species with different migratory statuses. Another potential application of such
468 Gaussian fields and FPCA could involve implementing a nonlinear additive predictor component
469 for altitude to study species' phenological changes as responses to climate: altitude ascents,
470 migrations, etc.

471 The STOC has demonstrated the effectiveness of public policies on protected areas. Outside
472 of reserves, the average decline of 56 species between 2004 and 2018 was -6.6% , whereas
473 within reserves, an increase of 12.5% was observed (Fontaine et al., 2020). Additionally, our
474 study offers new insights into the role of protected areas in mitigating the decline of passerine
475 populations, though our trends should be interpreted with caution. Indeed, several very
476 common species show a negative estimated relative abundance in the park, while the STOC
477 indicates positive trends for these same species. Thus, we believe that our assumption of
478 homogeneous reporting is overly optimistic and partially inaccurate. Despite this issue affecting
479 8 out of 59 species, we can still draw meaningful conclusions.

480 Notably, we demonstrate that forest passerines benefit the most in the Écrins National Park.
481 While not reversing the trend, the decline of farmland passerines in the park is slower than in
482 the rest of France, consistent with findings from previous studies (Palacín and Alonso, 2018;
483 Silva et al., 2018). Furthermore, our study provides novel insights into high mountain species
484 not previously investigated by the STOC. Our indicator, the relative abundance, shows a global
485 decline for species specialist to high-elevation environments. This trend aligns with the global
486 decrease of mountain specialist species in Europe (-10% during 2002–2014) as reported by
487 Lehikoinen et al., 2019. These findings will be valuable for comparison with upcoming analyses
488 of the STOM, a survey analogous to the STOC for mountain species, which has only been
489 operational since 2012, with its first results yet to be analyzed.

490 We identify three areas for future improvements in our modeling approach using solely
491 opportunistic data:

492 (i) The assumption of homogeneous sampling of different species, which could be questionable
493 due to differences in detectability, species commonness/rarity (Snäll et al., 2011), or agent
494 identification skills and reporting prevalence (Bradter et al., 2018), albeit reasonable in practice
495 and lacking better options. Heterogeneous detectability remains an issue without data based on
496 standardized sampling protocols and seems difficult to correct except for trying to include prior
497 expert knowledge into models. Different reporting behavior due to species commonness/rarity
498 can introduce bias in estimated intercept terms, such as α_i in Equation 6, but would not have
499 any impact on the other terms varying in space and/or time, as long as this bias in reporting
500 behavior is homogeneous across space, time and physical predictors. Effects related to agent
501 identification skills and reporting prevalence could potentially be identified by including random
502 effects for the rangers in the model. As a general recommendation to reduce biases and improve

503 the exploitability of opportunistic presence-only data, we suggest that when observers report
504 a species, all species of the same taxon should also be reported for better use of opportunistic
505 presence-only data in species distribution modeling.

506 (ii) The acquired information is relative to a target group of species since we used the target-
507 group occurrences as a proxy for the sampling effort (Equation 4). This implies that the potential
508 for interpretations of results in terms of absolute abundance and its variation across covariates,
509 space, and time remains relatively limited, as long as we do not have additional external
510 information on the absolute abundance of the target group as a whole.

511 (iii) There are multiple ways to define this target group, which impacts the interpretation of
512 abundances relative to it. The composition of this target group primarily relies on available data
513 sources, ranging from studies covering a single species (Farr et al., 2021) to those encompassing
514 thousands (Botella et al., 2021). Some studies focus exclusively on species within the same
515 taxonomic group (Van Strien et al., 2013), while others consider species across different taxa
516 (Escamilla Molgora et al., 2022). Botella et al., 2020 recommends choosing a set of species that
517 are consistently abundant across a broad range of environmental sub-regions, which is what we
518 did in this study.

519 We have recognized a pressing need for more comprehensive descriptions of what we refer
520 to as “sampling effort” and “observer bias” when dealing with opportunistic presence-only data,
521 as well as for methodologies to effectively incorporate it into modeling endeavors. Various
522 strategies have been proposed to address sampling efforts and observer bias depending on the
523 studied taxa, the available datasets, and model type. These strategies include the use of random
524 or target-group pseudo-absences with Maxent (Phillips et al., 2006, 2009), proxy variables
525 representing site accessibility (Henckel et al., 2020; Moreira et al., 2024; Warton et al., 2013)
526 and refinements (Chauvier et al., 2021), integration of information from detection/nondetection
527 data (Dorazio, 2014; Fithian et al., 2015; Giraud et al., 2016), and questionnaires (Bradter et al.,
528 2018). In our case study, very large occurrence numbers were available for the target group,
529 such that uncertainties about the cumulative abundance of the target group can be expected to
530 be relatively small. For modeling more precisely the cumulative abundance of the target group,
531 for example in cases with smaller occurrence numbers or when large areas of the study domain
532 are only very weakly sampled, we could develop an alternative approach wherein $Y_{TG}(s, m, p)$
533 could be initially modeled with a Poisson distribution using a mean parameter $\mu_{TG}(s, m, p)$,
534 fitted to the data to estimate $\mu_{TG}(s, m, p)$. Subsequently, we would use the estimated Poisson
535 mean as a proxy for $\Lambda_{TG}(s, m, p)E(s, m, p)$, rather than the raw count. This modification could
536 potentially enhance robustness and improve the assessment of uncertainties inherent to our
537 modeling approach.

538 Our findings suggest that, despite the statistical challenges arising from taking sampling
539 biases into account, opportunistic presence-only data provide unparalleled spatio-temporal
540 coverage, allowing for the depiction of consistent biological indicators that would be challenging
541 to obtain using traditional methods. This includes monitoring trends in relative abundance over
542 extended periods and developing species distribution models across vast geographical areas.
543 Such information can serve as valuable quantitative indicators for assessing the effectiveness
544 of conservation policies in protected areas.

545

Acknowledgements

546

547 The authors would like to thank all the staff of the Écrins National Park and its scientific team
548 for contributing to the dataset. They also appreciate advice of Denis Allard on using functional
549 principal component analysis, and Loïc Houde for technical support with the high-performance
550 computing cluster used to implement the models of this case study.

Fundings

551

552 The authors declare that they have received no specific funding for this study.

Conflict of interest disclosure

553

554 The authors declare that they comply with the PCI rule of having no financial conflicts of interest
555 in relation to the content of the article.

Data, script, code, and supplementary information availability

556

557 Data, scripts, and code can be accessed online (<https://doi.org/10.5281/zenodo.12545052>;
558 cite Lasgorceux et al., 2024a).

559 Outputs from fitted models under various cross-validation scenarios are also available online
560 (<https://doi.org/10.5281/zenodo.12544212>; cite Lasgorceux et al., 2024b).

561 Supplementary information is accessible online (<https://doi.org/10.5281/zenodo.12541412>;
562 cite Lasgorceux et al., 2024c).

563

References

564

565 Barnes AE, Davies JG, Martay B, Boersch-Supan PH, Harris SJ, Noble DG, Pearce-Higgins JW,
566 Robinson RA (2023). *Rare and declining bird species benefit most from designating protected
567 areas for conservation in the UK. Nature Ecology & Evolution* **7**, 92–101. [https://doi.org/10.
568 1038/s41559-022-01927-4](https://doi.org/10.1038/s41559-022-01927-4).

569 Belmont J, Martino S, Illian J, Rue H (2024). *Spatio-temporal Occupancy Models with INLA*. [https:
570 //doi.org/10.48550/arXiv.2403.10680](https://doi.org/10.48550/arXiv.2403.10680).

571 Botella C, Joly A, Bonnet P, Munoz F, Monestiez P (2021). *Jointly estimating spatial sampling
572 effort and habitat suitability for multiple species from opportunistic presence-only data. Methods
573 in Ecology and Evolution* **12**, 933–945. <https://doi.org/10.1111/2041-210X.13565>.

574 Botella C, Joly A, Monestiez P, Bonnet P, Munoz F (2020). *Bias in presence-only niche models
575 related to sampling effort and species niches: Lessons for background point selection. PLOS ONE*
576 **15(5)**, e0232078. <https://doi.org/10.1371/journal.pone.0232078>.

577 Bradter U, Mair L, Jönsson M, Knape J, Singer A, Snäll T (2018). *Can opportunistically collected
578 Citizen Science data fill a data gap for habitat suitability models of less common species? Methods
579 in Ecology and Evolution* **9**, 1667–1678. <https://doi.org/10.1111/2041-210X.13012>.

580 Bunz Y (2022). *Les programmes scientifiques de la faune vertebrée*. Parc national des Écrins,
581 Domaine de Charance, Gap, France, p. 423. URL: [https://www.documentation.
582 eauetbiodiversite.fr/notice/les-programmes-scientifiques-de-la-faune-
583 vertebree-edition-20220](https://www.documentation.eauetbiodiversite.fr/notice/les-programmes-scientifiques-de-la-faune-vertebree-edition-20220).

- 584 Chauvier Y, Zimmermann NE, Poggiato G, Bystrova D, Brun P, Thuiller W (2021). Novel meth-
585 ods to correct for observer and sampling bias in presence-only species distribution models. *Global*
586 *Ecology and Biogeography* **30**, 2312–2325. <https://doi.org/10.1111/geb.13383>.
- 587 Coron C, Calenge C, Giraud C, Julliard R (2018). Bayesian estimation of species relative abundances
588 and habitat preferences using opportunistic data. *Environmental and Ecological Statistics* **25**, 71–
589 93. <https://doi.org/10.1007/s10651-018-0398-2>.
- 590 Donald PF, Green RE, Heath MF (2001). Agricultural intensification and the collapse of Europe's
591 farmland bird populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences*
592 **268**, 25–29. <https://doi.org/10.1098/rspb.2000.1325>.
- 593 Dorazio RM (2014). Accounting for imperfect detection and survey bias in statistical analysis of
594 presence-only data. *Global Ecology and Biogeography* **23**, 1472–1484. <https://doi.org/10.1111/geb.12216>.
- 596 Dray S, Dufour AB (2007). The *ade4* Package: Implementing the Duality Diagram for Ecologists.
597 *Journal of Statistical Software* **22**, 1–20. <https://doi.org/10.18637/jss.v022.i04>.
- 598 Duckworth GD, Altwegg R (2018). Effectiveness of protected areas for bird conservation depends
599 on guild. *Diversity and Distributions* **24**, 1083–1091. <https://doi.org/10.1111/ddi.12756>.
- 600 Eftekhari A, Pasadakis D, Bollhöfer M, Scheidegger S, Schenk O (2021). Block-enhanced precision
601 matrix estimation for large-scale datasets. *Journal of Computational Science* **53**. <https://doi.org/10.1016/j.jocs.2021.101389>.
- 603 Escamilla Molgora JM, Sedda L, Diggle P, Atkinson PM (2022). A joint distribution framework to
604 improve presence-only species distribution models by exploiting opportunistic surveys. *Journal of*
605 *Biogeography* **49**, 1176–1192. <https://doi.org/10.1111/jbi.14365>.
- 606 Farr MT, Green DS, Holekamp KE, Zipkin EF (2021). Integrating distance sampling and presence-
607 only data to estimate species abundance. *Ecology* **102**, e03204. [https://doi.org/10.1002/](https://doi.org/10.1002/ecy.3204)
608 [ecy.3204](https://doi.org/10.1002/ecy.3204).
- 609 Fawcett T (2006). An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>.
- 611 Fick SE, Hijmans RJ (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global
612 land areas. *International Journal of Climatology* **37**, 4302–4315. [https://doi.org/10.1002/](https://doi.org/10.1002/joc.5086)
613 [joc.5086](https://doi.org/10.1002/joc.5086).
- 614 Fithian W, Elith J, Hastie T, Keith DA (2015). Bias correction in species distribution models: pooling
615 survey and collection data for multiple species. *Methods in Ecology and Evolution* **6**, 424–438.
616 <https://doi.org/10.1111/2041-210X.12242>.
- 617 Fithian W, Hastie T (2013). Finite-sample equivalence in statistical models for presence-only data.
618 *The Annals of Applied Statistics* **7**, 1917–1939. <https://doi.org/10.1214/13-AOAS667>.
- 619 Flousek J, Telenský T, Hanzelka J, Reif J (2015). Population Trends of Central European Montane
620 Birds Provide Evidence for Adverse Impacts of Climate Change on High-Altitude Species. *PLOS*
621 *ONE* **10**, e0139465. <https://doi.org/10.1371/journal.pone.0139465>.
- 622 Fontaine B, Moussy C, Chiffard Carricaburu J, Dupuis J, Schmaltz L, Lorrillière R, Lois G, Gau-
623 dard C, Couzi L (2020). *Suivi des oiseaux communs en France 1989-2019 : 30 ans de suivis*
624 *participatifs*. MNHN- Centre d'Ecologie et des Sciences de la Conservation, LPO BirdLife
625 France - Service Connaissance, Ministère de la Transition écologique et solidaire, p. 46.
626 URL: [https://www.vigienature.fr/sites/vigienature/files/atoms/files/](https://www.vigienature.fr/sites/vigienature/files/atoms/files/syntheseoiseauxcommuns2020_final.pdf)
627 [syntheseoiseauxcommuns2020_final.pdf](https://www.vigienature.fr/sites/vigienature/files/atoms/files/syntheseoiseauxcommuns2020_final.pdf).

- 628 Gaedke-Merzhäuser L, Niekerk J, Schenk O, Rue H (2023). *Parallelized integrated nested Laplace*
629 *approximations for fast Bayesian inference*. *Statistics and Computing* **33**, 25. [https://doi.org/](https://doi.org/10.1007/s11222-022-10192-1)
630 [10.1007/s11222-022-10192-1](https://doi.org/10.1007/s11222-022-10192-1).
- 631 Gilks WR, Richardson S, Spiegelhalter D (1995). *Markov Chain Monte Carlo in Practice*. New York,
632 NY: Chapman and Hall/CRC. 512 pp. <https://doi.org/10.1201/b14835>.
- 633 Giraud C, Calenge C, Coron C, Julliard R (2016). *Capitalizing on opportunistic data for monitoring*
634 *relative abundances of species*. *Biometrics* **72**, 649–658. [https://doi.org/10.1111/biom.](https://doi.org/10.1111/biom.12431)
635 [12431](https://doi.org/10.1111/biom.12431).
- 636 Gregory RD, Skorpilova J, Vorisek P, Butler S (2019). *An analysis of trends, uncertainty and species*
637 *selection shows contrasting trends of widespread forest and farmland birds in Europe*. *Ecological*
638 *Indicators* **103**, 676–687. <https://doi.org/10.1016/j.ecolind.2019.04.064>.
- 639 Gregory RD, Vorisek P, Van Strien A, Gmelig Meyling AW, Jiguet F, Fornasari L, Reif J, Chylarecki
640 P, Burfield IJ (2007). *Population trends of widespread woodland birds in Europe*. *Ibis* **149** (s2),
641 78–97. <https://doi.org/10.1111/j.1474-919X.2007.00698.x>.
- 642 Harris I, Osborn TJ, Jones P, Lister D (2020). *Version 4 of the CRU TS monthly high-resolution*
643 *gridded multivariate climate dataset*. *Scientific Data* **7**, 109. [https://doi.org/10.1038/](https://doi.org/10.1038/s41597-020-0453-3)
644 [s41597-020-0453-3](https://doi.org/10.1038/s41597-020-0453-3).
- 645 Hauptert S, Sèbe F, Sueur J (2023). *Physics-based model to predict the acoustic detection distance*
646 *of terrestrial autonomous recording units over the diel cycle and across seasons: Insights from an*
647 *Alpine and a Neotropical forest*. *Methods in Ecology and Evolution* **14**, 614–630. [https://doi.](https://doi.org/10.1111/2041-210X.14020)
648 [org/10.1111/2041-210X.14020](https://doi.org/10.1111/2041-210X.14020).
- 649 Heldbjerg H, Sunde P, Fox AD (2018). *Continuous population declines for specialist farmland birds*
650 *1987-2014 in Denmark indicates no halt in biodiversity loss in agricultural habitats*. *Bird Conser-*
651 *vation International* **28**, 278–292. <https://doi.org/10.1017/S0959270916000654>.
- 652 Henckel L, Bradter U, Jönsson M, Isaac NJB, Snäll T (2020). *Assessing the usefulness of citizen*
653 *science data for habitat suitability modelling: Opportunistic reporting versus sampling based on a*
654 *systematic protocol*. *Diversity and Distributions* **26**, 1276–1290. [https://doi.org/10.1111/](https://doi.org/10.1111/ddi.13128)
655 [ddi.13128](https://doi.org/10.1111/ddi.13128).
- 656 Hijmans RJ (2024). *terra: Spatial Data Analysis*. R package version 1.7-71. URL: [https://CRAN.R-](https://CRAN.R-project.org/package=terra)
657 [project.org/package=terra](https://CRAN.R-project.org/package=terra).
- 658 Horns JJ, Adler FR, Şekerciöğlü ÇH (2018). *Using opportunistic citizen science data to estimate*
659 *avian population trends*. *Biological Conservation* **221**, 151–159. [https://doi.org/10.1016/](https://doi.org/10.1016/j.biocon.2018.02.027)
660 [j.biocon.2018.02.027](https://doi.org/10.1016/j.biocon.2018.02.027).
- 661 IGN (2018). *RGE ALTI® [Data set]*. URL: <https://geoservices.ign.fr/rgealti>.
- 662 Illian JB, Martino S, Sørbye SH, Gallego-Fernández JB, Zunzunegui M, Esquivias MP, Travis JMJ
663 (2013). *Fitting complex ecological point process models with integrated nested Laplace approxi-*
664 *mation*. *Methods in Ecology and Evolution* **4**, 305–315. [https://doi.org/10.1111/2041-](https://doi.org/10.1111/2041-210x.12017)
665 [210x.12017](https://doi.org/10.1111/2041-210x.12017).
- 666 Inger R, Gregory R, Duffy JP, Stott I, Voříšek P, Gaston KJ (2015). *Common European birds are de-*
667 *clining rapidly while less abundant species' numbers are rising*. *Ecology Letters* **18**, 28–36. [https://](https://doi.org/10.1111/ele.12387)
668 doi.org/10.1111/ele.12387.
- 669 Inglada J, Vincent A, Thierion V (2018). *Theia OSO Land Cover Map 2018 [Data set]*. [https://](https://doi.org/10.5281/zenodo.3613415)
670 doi.org/10.5281/zenodo.3613415.

- 671 IPBES (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosys-*
672 *tem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Ser-*
673 *vices*. Bonn, Germany. 56 pages. URL: <https://doi.org/10.5281/zenodo.3553579>.
- 674 Janžekovič F, Novak T (2012). *PCA – A Powerful Method for Analyze Ecological Niches*. In: *Principal*
675 *Component Analysis*. 8. Rijeka: IntechOpen. <https://doi.org/10.5772/38538>.
- 676 Jiguet F, Devictor V, Julliard R, Couvet D (2012). *French citizens monitoring ordinary birds provide*
677 *tools for conservation and ecological sciences*. *Acta Oecologica* **44**, 58–66. [https://doi.org/](https://doi.org/10.1016/j.actao.2011.05.003)
678 [10.1016/j.actao.2011.05.003](https://doi.org/10.1016/j.actao.2011.05.003).
- 679 Jin Huang, Ling C (2005). *Using AUC and accuracy in evaluating learning algorithms*. *IEEE Transac-*
680 *tions on Knowledge and Data Engineering* **17**, 299–310. [https://doi.org/10.1109/TKDE.](https://doi.org/10.1109/TKDE.2005.50)
681 [2005.50](https://doi.org/10.1109/TKDE.2005.50).
- 682 Jung Y (2018). *Multiple predicting K -fold cross-validation for model selection*. *Journal of Nonpara-*
683 *metric Statistics* **30**, 197–215. <https://doi.org/10.1080/10485252.2017.1404598>.
- 684 Kamp J, Frank C, Trautmann S, Busch M, Dröschmeister R, Flade M, Gerlach B, Karthäuser J,
685 Kunz F, Mitschke A, Schwarz J, Sudfeldt C (2021). *Population trends of common breeding birds*
686 *in Germany 1990–2018*. *Journal of Ornithology* **162**, 1–15. [https://doi.org/10.1007/](https://doi.org/10.1007/s10336-020-01830-4)
687 [s10336-020-01830-4](https://doi.org/10.1007/s10336-020-01830-4).
- 688 Kéry M, Royle JA, Schmid H, Schaub M, Volet B, Häfliger G, Zbinden N (2010). *Site-occupancy*
689 *distribution modeling to correct population-trend estimates derived from opportunistic observa-*
690 *tions*. *Conservation Biology* **24**, 1388–1397. [https://doi.org/10.1111/j.1523-1739.2010.](https://doi.org/10.1111/j.1523-1739.2010.01479.x)
691 [01479.x](https://doi.org/10.1111/j.1523-1739.2010.01479.x).
- 692 Knape J (2016). *Decomposing trends in Swedish bird populations using generalized additive mixed*
693 *models*. *Journal of Applied Ecology* **53**, 1852–1861. [https://doi.org/10.1111/1365-](https://doi.org/10.1111/1365-2664.12720)
694 [2664.12720](https://doi.org/10.1111/1365-2664.12720).
- 695 Lasgorceux F, Papaix J, Bunz Y, Combrisson D, Opitz T (2024a). *Data, scripts and code for 'Space-*
696 *time species distribution modeling with opportunistic presence-only data: a case study of passer-*
697 *ines in a protected area'*. <https://doi.org/10.5281/zenodo.12545052>.
- 698 Lasgorceux F, Papaix J, Bunz Y, Combrisson D, Opitz T (2024b). *Outputs from fitted models across*
699 *the cross-validation scenarios for 'Space-time species distribution modeling with opportunistic*
700 *presence-only data: a case study of passerines in a protected area'*. [https://doi.org/10.](https://doi.org/10.5281/zenodo.12544212)
701 [5281/zenodo.12544212](https://doi.org/10.5281/zenodo.12544212).
- 702 Lasgorceux F, Papaix J, Bunz Y, Combrisson D, Opitz T (2024c). *Supplementary information for*
703 *'Space-time species distribution modeling with opportunistic presence- only data: a case study of*
704 *passerines in a protected area'*. <https://doi.org/10.5281/zenodo.12541412>.
- 705 Lehtikoinen A, Brotons L, Calladine J, Campedelli T, Escandell V, Flousek J, Grueneberg C, Haas
706 F, Harris S, Herrando S, Husby M, Jiguet F, Kålås JA, Lindström Å, Lorrillière R, Molina B,
707 Pladevall C, Calvi G, Sattler T, Schmid H, et al. (2019). *Declining population trends of European*
708 *mountain birds*. *Global Change Biology* **25**, 577–588. <https://doi.org/10.1111/gcb.14522>.
- 709 Lindgren F, Rue H, Lindström J (2011). *An explicit link between Gaussian fields and Gaussian Markov*
710 *random fields: the stochastic partial differential equation approach*. *Journal of the Royal Statistical*
711 *Society Series B: Statistical Methodology* **73**, 423–498. [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-9868.2011.00777.x)
712 [9868.2011.00777.x](https://doi.org/10.1111/j.1467-9868.2011.00777.x).

- 713 Link WA, Cam E, Nichols JD, Cooch EG (2002). *Of Bugs and Birds: Markov Chain Monte Carlo for*
714 *Hierarchical Modeling in Wildlife Research. The Journal of Wildlife Management* **66**, 277. <https://doi.org/10.2307/3803160>.
715
- 716 McVicar TR, Körner C (2013). *On the use of elevation, altitude, and height in the ecological and*
717 *climatological literature. Oecologia* **171**, 335–337. [https://doi.org/10.1007/s00442-012-
718 *2416-7*.](https://doi.org/10.1007/s00442-012-2416-7)
- 719 Moreira GA, Menezes R, Wise L (2024). *Presence-Only for Marked Point Process Under Preferential*
720 *Sampling. Journal of Agricultural, Biological and Environmental Statistics* **29**, 92–109. <https://doi.org/10.1007/s13253-023-00558-x>.
721
- 722 Murtagh F, Legendre P (2014). *Ward's Hierarchical Agglomerative Clustering Method: Which Algo-*
723 *rithms Implement Ward's Criterion? Journal of Classification* **31**, 274–295. [https://doi.org/](https://doi.org/10.1007/s00357-014-9161-z)
724 [10.1007/s00357-014-9161-z](https://doi.org/10.1007/s00357-014-9161-z).
- 725 Newton I (2004). *The recent declines of farmland bird populations in Britain: an appraisal of causal*
726 *factors and conservation actions. Ibis* **146**, 579–600. [https://doi.org/10.1111/j.1474-
727 *919X.2004.00375.x*.](https://doi.org/10.1111/j.1474-919X.2004.00375.x)
- 728 Noël F, Combrisson D, Geoffroy JJ, Nicolas MG (2023). *Synthèse des connaissances sur les Isopodes*
729 *et Diplopodes terrestres du Parc national des Écrins (SE France). Naturae*, 151–170. <https://doi.org/10.5852/naturae2023a9>.
730
- 731 Palacín C, Alonso JC (2018). *Failure of EU Biodiversity Strategy in Mediterranean farmland protected*
732 *areas. Journal for Nature Conservation* **42**, 62–66. [https://doi.org/10.1016/j.jnc.2018.](https://doi.org/10.1016/j.jnc.2018.02.008)
733 [02.008](https://doi.org/10.1016/j.jnc.2018.02.008).
- 734 Pasadakis D, Bollhöfer M, Schenk O (2023). *Sparse Quadratic Approximation for Graph Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 11256–11269. <https://doi.org/10.1109/TPAMI.2023.3263969>.
735
- 736 Pedersen EJ, Miller DL, Simpson GL, Ross N (2019). *Hierarchical generalized additive models in*
737 *ecology: an introduction with mgcv. PeerJ* **7**, e6876. <https://doi.org/10.7717/peerj.6876>.
738
- 739 Phillips SJ, Anderson RP, Schapire RE (2006). *Maximum entropy modeling of species geographic*
740 *distributions. Ecological Modelling* **190**, 231–259. [https://doi.org/10.1016/j.ecolmodel.](https://doi.org/10.1016/j.ecolmodel.2005.03.026)
741 [2005.03.026](https://doi.org/10.1016/j.ecolmodel.2005.03.026).
- 742 Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009). *Sample selec-*
743 *tion bias and presence-only distribution models: implications for background and pseudo-absence*
744 *data. Ecological Applications* **19**, 181–197. <https://doi.org/10.1890/07-2153.1>.
- 745 R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for
746 *Statistical Computing*. Vienna, Austria. URL: <https://www.r-project.org>.
- 747 Ramsay JO, Silverman BW (2005). *Functional Data Analysis*. Springer Series in Statistics. New
748 *York, NY: Springer. https://doi.org/10.1007/b98888*.
- 749 Ramsay J, Hooker G, Graves S (2009). *Functional Data Analysis with R and MATLAB*. New York,
750 *NY: Springer. https://doi.org/10.1007/978-0-387-98185-7*.
- 751 Reif J, Storch D, Voříšek P, Šťastný K, Bejček V (2008a). *Bird-habitat associations predict population*
752 *trends in central European forest and farmland birds. Biodiversity and Conservation* **17**, 3307–
753 *3319. https://doi.org/10.1007/s10531-008-9430-4*.
- 754 Reif J, Voříšek P, Šťastný K, Bejček V, Petr J (2008b). *Agricultural intensification and farmland birds:*
755 *new insights from a central European country. Ibis* **150**, 596–605. [https://doi.org/10.1111/](https://doi.org/10.1111/j.1474-919X.2008.00829.x)
756 [j.1474-919X.2008.00829.x](https://doi.org/10.1111/j.1474-919X.2008.00829.x).

- 757 Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI (2015).
758 *Point process models for presence-only analysis. Methods in Ecology and Evolution* **6**, 366–379.
759 <https://doi.org/10.1111/2041-210X.12352>.
- 760 Renner IW, Warton DI (2013). *Equivalence of MAXENT and Poisson Point Process Models for Species*
761 *Distribution Modeling in Ecology. Biometrics* **69**, 274–281. [https://doi.org/10.1111/j.](https://doi.org/10.1111/j.1541-0420.2012.01824.x)
762 [1541-0420.2012.01824.x](https://doi.org/10.1111/j.1541-0420.2012.01824.x).
- 763 Rigal S, Dakos V, Alonso H, Auniņš A, Benkő Z, Brotons L, Chodkiewicz T, Chylarecki P, Carli
764 E, Moral JC, Domşa C, Escandell V, Fontaine B, Foppen R, Gregory R, Harris S, Herrando S,
765 Husby M, Ieronymidou C, Jiguet F, et al. (2023). *Farmland practices are driving bird population*
766 *decline across Europe. Proceedings of the National Academy of Sciences* **120**, e2216573120.
767 <https://doi.org/10.1073/pnas.2216573120>.
- 768 Rue H, Martino S, Chopin N (2009). *Approximate Bayesian inference for latent Gaussian models by*
769 *using integrated nested Laplace approximations. Journal of the Royal Statistical Society Series B:*
770 *Statistical Methodology* **71**, 319–392. [https://doi.org/10.1111/j.1467-9868.](https://doi.org/10.1111/j.1467-9868.2008.00700.x)
771 [2008.](https://doi.org/10.1111/j.1467-9868.2008.00700.x)
[00700.x](https://doi.org/10.1111/j.1467-9868.2008.00700.x).
- 772 Sanderson FJ, Kucharz M, Jobda M, Donald PF (2013). *Impacts of agricultural intensification and*
773 *abandonment on farmland birds in Poland following EU accession. Agriculture, Ecosystems & En-*
774 *vironment* **168**, 16–24. <https://doi.org/10.1016/j.agee.2013.01.015>.
- 775 Schulze ED, Craven D, Durso AM, Reif J, Guderle M, Kroihner F, Hennig P, Weiserbs A, Schall P,
776 Ammer C, Eisenhauer N (2019). *Positive association between forest management, environmental*
777 *change, and forest bird abundance. Forest Ecosystems* **6**, 3. [https://doi.org/10.1186/](https://doi.org/10.1186/s40663-019-0160-8)
778 [s40663-019-0160-8](https://doi.org/10.1186/s40663-019-0160-8).
- 779 Silva JP, Correia R, Alonso H, Martins RC, D'Amico M, Delgado A, Sampaio H, Godinho C, Mor-
780 eira F (2018). *EU protected area network did not prevent a country wide population decline in a*
781 *threatened grassland bird. PeerJ* **6**, e4284. <https://doi.org/10.7717/peerj.4284>.
- 782 Snäll T, Kindvall O, Nilsson J, Pärt T (2011). *Evaluating citizen-based presence data for bird moni-*
783 *toring. Biological Conservation* **144**, 804–810. [https://doi.org/10.1016/j.biocon.2010.](https://doi.org/10.1016/j.biocon.2010.11.010)
784 [11.010](https://doi.org/10.1016/j.biocon.2010.11.010).
- 785 Soriano-Redondo A, Jones-Todd CM, Bearhop S, Hilton GM, Lock L, Stanbury A, Votier SC, Il-
786 lian JB (2019). *Understanding species distribution in dynamic populations: a new approach using*
787 *spatio-temporal point process models. Ecography* **42**, 1092–1102. [https://doi.org/10.1111/](https://doi.org/10.1111/ecog.03771)
788 [ecog.03771](https://doi.org/10.1111/ecog.03771).
- 789 Timmers R, Kuijk M, Verweij PA, Ghazoul J, Hautier Y, Laurance WF, Arriaga-Weiss SL, Askins RA,
790 Battisti C, Berg Å, Daily GC, Estades CF, Frank B, Kurosawa R, Pojar RA, Woinarski JC, Soons
791 MB (2022). *Conservation of birds in fragmented landscapes requires protected areas. Frontiers in*
792 *Ecology and the Environment* **20**, 361–369. <https://doi.org/10.1002/fee.2485>.
- 793 Traba J, Morales MB (2019). *The decline of farmland birds in Spain is strongly associated to the loss*
794 *of fallowland. Scientific Reports* **9**, 9473. <https://doi.org/10.1038/s41598-019-45854-0>.
- 795 Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J (2022). *Predictive performance of presence-*
796 *only species distribution models: a benchmark study with reproducible code. Ecological Mono-*
797 *graphs* **92**, e01486. <https://doi.org/10.1002/ecm.1486>.
- 798 Van Strien AJ, Van Swaay CA, Termaat T (2013). *Opportunistic citizen science data of animal species*
799 *produce reliable estimates of distribution trends if analysed with occupancy models. Journal of*
800 *Applied Ecology* **50**, 1450–1458. <https://doi.org/10.1111/1365-2664.12158>.

- 801 Warton DI, Renner IW, Ramp D (2013). *Model-Based Control of Observer Bias for the Analysis of*
802 *Presence-Only Data in Ecology*. *PLoS ONE* **8**, e79168. <https://doi.org/10.1371/journal.pone.0079168>.
- 803
804 Warton DI, Shepherd LC (2010). *Poisson point process models solve the “pseudo-absence problem”*
805 *for presence-only data in ecology*. *The Annals of Applied Statistics* **4**. <https://doi.org/10.1214/10-AOAS331>.
- 806
807 Wikle CK (2003). *Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes*.
808 *Ecology* **84**, 1382–1394. [https://doi.org/10.1890/0012-9658\(2003\)084\[1382:HBMFPT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2).
- 809
810 Wretenberg J, Lindström Å, Svensson S, Thierfelder T, Pärt T (2006). *Population trends of farmland*
811 *birds in Sweden and England: similar trends but different patterns of agricultural intensification*.
812 *Journal of Applied Ecology* **43**, 1110–1120. <https://doi.org/10.1111/j.1365-2664.2006.01216.x>.
- 813
814 Zamora R, Barea-Azcón JM (2015). *Long-Term Changes in Mountain Passerine Bird Communities*
815 *in the Sierra Nevada (Southern Spain): A 30-Year Case Study*. *Ardeola* **62**, 3–18. <https://doi.org/10.13157/arla.62.1.2015.3>.
- 816