Peer Community Journal

Section: Ecology

Research article

Published
2026-06-09

Cite as
Florian Lasgorceux, Julien Papaïx, Yoann Bunz, Damien Combrisson and Thomas Opitz (2026) *Space-time species distribution modeling with opportunistic presence-only data: a case study of passerines in a protected area*, Peer Community Journal, 6: e52.

Correspondence
florian.lasgorceux@proton.me

Peer-review
Peer reviewed and recommended by PCI Ecology,
https://doi.org/10.24072/pci.ecology.100750

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

Florian Lasgorceux[1], Julien Papaïx[1], Yoann Bunz[2], Damien Combrisson[2], and Thomas Opitz[1]

Volume **6** (2026), article e52

https://doi.org/10.24072/pcjournal.731

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 (ENP) in France, can help mitigate these impacts. We evaluated whether an opportunistic presence-only dataset collected by trained ENP rangers contains biological signals strong enough to support robust statistical inference. Using a generalized additive Poisson model with spatial and spatio-temporal covariates, monthly latent spatio-temporal Gaussian random fields, and a non-spatial inter-annual effect, we estimated the relative abundance of 76 passerine species on a regular grid, with occurrences aggregated per spatio-temporal cell used as a proxy for sampling effort. The model showed good calibration for most species (AUC > 0.8) and reliably captured habitat preferences and migratory status. Relative-abundance trends in ENP were compared with relative abundance from three monitoring programs: STOM (ENP), STOC (France), and MHB (Switzerland). For most species with significant trends, model predictions aligned with survey-based trends. Forest specialists benefited most from the protected-area status, and farmland species declined more slowly in ENP than in France. High-elevation specialists generally decreased in both ENP and Switzerland. Discrepancies mostly arose for common species, likely reflecting uncorrected declines in ranger reporting rates. These results demonstrate that high-resolution opportunistic presence-only data can provide valuable insights into biological patterns and trends while reducing reliance on external data to estimate sampling effort.

[1] Biostatistics and Spatial Processes, INRAE, Avignon, France, [2] Écrins National Park, Gap, France

Peer Community Journal is a member of the Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/

e-ISSN 2804-3871

Introduction

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

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

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

In addition to data from monitoring programs, the Écrins National Park maintains a large dataset of presence-only records collected opportunistically by rangers. The primary use of these data is to ensure ecological surveillance at high spatio-temporal resolution and to serve as a basis for knowledge-sharing within the context of development projects. Using these data in statistical modeling presents a challenge due to sampling heterogeneity, which concerns the spatial and temporal dimensions, and between species and observers because of their opportunistic nature (Van Strien et al., 2013). This sampling heterogeneity translates into biases in species distribution models that do not account for it. In this study, we propose to harness this extensive dataset for modeling species habitat preferences, classifying species migratory status, and studying time variations in relative abundance. Here, relative abundance is defined as the ratio of the abundance of a species divided by the cumulated abundance of a pool of species constituting the so-called target group of species. Under reasonable assumptions on the sampling behavior of observers, such relative quantities can be estimated without bias.

Various estimation frameworks and algorithms exist for constructing species distribution models and calibrating their parameters from presence-only data. Widely used methods include Maxent, based on the maximum-entropy principle (Phillips et al., 2006), and various statistical regression modeling frameworks (Valavi et al., 2022). These methods are often based on point-process representations of data (Renner et al., 2015; Warton and Shepherd, 2010), where each observed occurrence of a species is viewed as a point in a cloud of points located in space and time. In this context, models such as Maxent and logistic regression are equivalent to special cases of the point-process setting (Fithian and Hastie, 2013; Renner and Warton, 2013). Other methods have been developed to estimate temporal trends from presence-only data. An early approach by Telfer et al., 2002 uses logit-transformed proportions of occupied grid cells across two time periods to calculate indices of relative range change. More recently, Kéry et al., 2010 and Van Strien et al., 2013 apply site-occupancy models on presence-only data: the former necessitates that all observed species are recorded, while the latter assumes to correct for both observation and reporting biases by correcting detection bias (Van Strien et al., 2013). Several studies have also focused on correcting sampling effort from opportunistic data, such as the Frescalo method, which uses the proportion of recorded common species (Hill, 2012), or approaches based on counts of selected species groups (Botella et al., 2020). While these methods may still be influenced by variation in recorder activity, Isaac et al., 2014 recommends them over simpler filtering-based approaches, emphasizing the importance of explicitly accounting for sampling heterogeneity when estimating trends from presence-only data. Due to uncertainty in the reported positions within the Écrins National Park dataset, we implement a simplified version of point-process models by aggregating occurrences into spatio-temporal cells. This leads to the use of a generalized additive regression model designed to simultaneously infer species' ecological niches and temporal trends using opportunistic presence-only data.

Models including random effects are often used to capture the complex spatio-temporal structures in data and ecological processes, thus allowing for an accurate assessment of various sources of uncertainty. This has led to the widespread adoption of Bayesian inference techniques such as Markov Chain Monte Carlo (MCMC, (Gilks et al., 1995; Link et al., 2002)) or Integrated Nested Laplace Approximations (INLA, Illian et al., 2013). We apply the INLA approach to perform complex Bayesian spatio-temporal inferences with relatively large datasets by combining generalized additive regression models with the Stochastic Partial Differential Equation

(SPDE) approach, the latter used to represent spatial random effects (Lindgren et al., 2011). The construction and estimation of complex spatio-temporal ecological models are still recent developments (Belmont et al., 2024; Soriano-Redondo et al., 2019), and as far as we know, our study represents the first attempt to evaluate management policies of protected areas using solely opportunistic presence-only data for this type of model.

Materials and methods

Data

Passerines presence-only data. The dataset comprises 102,513 opportunistically reported events of passerine detections between 1994 and 2021, recorded by rangers of the Écrins National Park with professional training in naturalistic expertise. We chose to focus our study on the *Passeriformes* because it is a taxon generally well known by all rangers, ensuring consistent reporting rates, unlike more specific taxa. Furthermore, we have excluded three species – Common Raven (*Corvus corax*), Red-billed Chough (*Pyrrhocorax pyrrhocorax*), and Alpine Chough (*Pyrrhocorax graculus*) – due to the high uncertainty associated with their reported positions compared to the rest of the passerines. We narrowed the temporal window to the 1994–2021 period to ensure a minimum of 2000 reported events per year (Appendix A, Lasgorceux et al., 2026c).

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

Based on interviews with various rangers, we have gained insight that the reported position could be i) an approximation of the reported individuals' position at detection time, ii) the observer's position at detection time, or iii) another, different position where the observer gathered several occurrences before adding them jointly to the data records. Though, when multiple occurrences are gathered at a station position, rangers ensure that the environment is consistent between the station position and the birds' individual position at detection time and that the distance between these two positions does not exceed 250m. This distance of 250m is obtained by considering at most 100m for the difference between the positions of the detected individual(s) and the observer by sound recognition (Haupert et al., 2023), plus an additional 150m for the distance between the position at detection time and the station position. To account for this uncertainty surrounding the reported position, we aggregated the data by assigning each reported position to the center of a regular spatial grid with 500m resolution covering the Écrins National Park. We excluded boundary cells with an area smaller than 25m^2, which corresponds to the resolution of the environmental covariates we use for characterizing the spatial habitat preferences. This discretization results in dividing the Écrins National Park domain into 11,260 spatial cells.

Our objective is not only to infer species habitat preferences but also to analyze temporal trends in relative abundance. Therefore, we divided the 28 years of data into seven periods of 4 years each to capture long-term changes between such periods. Additionally, in order to take into account that some studied species are migratory, we further subdivided the data temporally by month to investigate intra-annual variations. This spatio-temporal discretization scheme results in a total of $7\times12\times11{,}260 = 945{,}840$ spatio-temporal cells.

Passerine population trends. To compare our results with local and national trends in relative abundance variations from other sources, we used three additional datasets: the French Breeding Bird Survey (STOC program, 2001–2019) (Fontaine et al., 2020; Jiguet et al., 2012); the Mountain Bird Monitoring Survey in Écrins National Park (STOM program, 2015–2024) (Chiffard, 2024); the Swiss Monitoring of Common Breeding Birds (MHB program, 1990–2024) (Strebel et al., 2024). The STOC and MHB datasets provide long-term percentage changes in the abundance of common bird species in France and Switzerland, respectively. For the STOM dataset in Écrins National Park, due to the limited sample size, we estimated long-term abundance trends using a Poisson generalized linear model with a log link and year as the covariate. Comparing trends derived from opportunistic data with standardized survey data can serve various purposes, such as assessing the consistency of model assumptions, comparing local indicators within a protected area to regional and national indicators, and providing insights into space-time population trends for species specialized in mountainous habitats, for which surveys are quite limited in terms of spatio-temporal coverage.

Environmental Data. The Écrins National Park is a high-mountain park where the terrain structure plays a predominant role in the ecosystems. We utilized the Digital Elevation Model data from RGE (IGN, 2018) to extract the elevation (McVicar and Körner, 2013) on a 5m regular grid. Subsequently, we calculated the slope at the same resolution using the ‘terrain’ function from the R package *terra* (Hijmans, 2024). Ultimately, we derived two spatial covariates: the mean elevation and the mean slope values per spatial cell at 500m resolution.

Regarding land cover and land use data, we make use of the OSO dataset (Inglada et al., 2018). Originally at a resolution of 10/20m, we applied the nearest neighbor strategy to downscale the values to a 5m resolution, for alignment with the RGE data resolution. To avoid statistical issues due to uncommon land cover and land use types in the Écrins National Park, we aggregated data towards 10 representative categories: Urban, Crops, Meadows, Deciduous forest, Coniferous forest, Grasslands, Woody Heaths, Mineral Surfaces, Glaciers and snow, and Water (see Appendix B in Lasgorceux et al., 2026c for further details). We then calculated the percentage of coverage of each category in each spatial cell at 500m resolution.

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

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

Hence, we began by decomposing the climatic variables into their spatial means and spatio-temporal anomalies with respect to those means. We then performed a Principal Component

Analysis (PCA, Janžekovič and Novak, 2012) using the R package *ade4* (Dray and Dufour, 2007). The input variables were: the mean elevation, the mean slope, the percentages of coverage computed for each of the ten representative categories derived from the OSO classification, as well as the spatial means of average maximum temperature and total precipitation. The first six axes of the PCA, accounting for 75% of the information, allow for useful interpretation and have therefore been incorporated into the species distribution models (see Appendix C in Lasgorceux et al., 2026c for details). The first principal component axis explains a significant 31.1% of the variation in habitat types, primarily driven by high-elevation environments. The second axis is characterized by vegetation at medium to high elevations, particularly in coniferous forests, especially in steep areas. The third axis is influenced by grasslands, and the fourth axis is primarily driven by wetlands and glaciers. However, since we observe a significant predominance of bird occurrences in “water” habitats within urbanized valleys (Figure 14 in Appendix C, Lasgorceux et al., 2026c) compared to “glaciers and snow” habitats (751 occurrences versus 80), we interpret the fourth axis as “urban wetlands”. The interpretation of the fifth axis is also associated with high-elevation vegetation, as it is primarily driven by dense, low-growing shrubs and small trees; henceforth, we will refer to these environments as “woody heaths”. The sixth axis is driven by low-elevation deciduous forests.

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

Statistical modeling

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

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

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

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

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

$$\mu_i(s, m, p) = \Lambda_i(s, m, p) \times E_i(s, m, p) \tag{2}$$

Sampling effort. A major challenge is now to quantify or eliminate the term representing the sampling effort. Within our spatio-temporal discretized framework, the first step is to obtain information about where and when no observations occurred. This helps distinguish between cells with zero occurrences due to the absence of observer presence and those with zero occurrences due to the absence of the species. The strategy involves utilizing the extensive opportunistic presence-only dataset, using reported occurrences of certain species as indicators of observer presence, as proposed by Botella et al., 2020. However, we have to assume that, for any given species, the reporting rate (i.e. the probability of detection and reporting) is constant across observers and does not vary between species. This explains one of our reasons to primarily restrict our study to passerines (see the subsection *Passerine data*). Following existing literature, we refer to this set of chosen species with homogeneous sampling conditions as the target group (TG) of species. Therefore, we denote the sampling effort term as E instead of E_i, and we write $\Lambda_{TG}(s, m, p)$ for the cumulative abundance of all species in the target group in cell (s, m, p). We can then reformulate the average number of occurrences in terms of the ratio of the abundance of species i to the cumulative abundance of the target group, and the product of the sampling effort with the cumulative abundance of the target group (Equation 4):

$$\mu_i(s, m, p) = \Lambda_i(s, m, p) \times E(s, m, p) \tag{3}$$

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

To statistically identify the first factor that indicates the abundance of species i relative to the cumulative abundance of the target group of species – hereafter referred to as relative abundance – we need a proxy for the second term $\Lambda_{TG}(s, m, p)E(s, m, p)$. The most natural approach is to use the number of target-group reported occurrences, denoted as $y_{TG}(s, m, p)$. We chose to include all passerine species in our target group of species. This choice is supported by the results of the algorithm adapted from Botella et al., 2020, which excluded only four species that were observed just once. We kept only the cells with at least one target-group occurrence, reducing the number of studied cells to 32,486.

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

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

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

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

Statistical inference. We employed the *R-INLA* package for the statistical inference of our models (Rue et al., 2009). INLA is a fast but accurate estimation method based on deterministic Laplace approximations – by contrast with approaches based on stochastic simulation, such as MCMC – and requires that the latent layer ($\log(\mu_i(s,m,p))$ in our model) follows a multivariate Gaussian distribution, possibly in very high dimension, as with our space-time fields. In this setting, INLA enables the implementation of a wide range of Gaussian random effects, including spatial fields implemented through the SPDE approach, which provides a numerically convenient representation of the Matérn covariance function used routinely in spatial statistics (Lindgren et al., 2011), and of priors for hyperparameters such as variances or correlation ranges. In Appendix D, we precisely describe the prior distributions of all parts of $\log(\mu_i(s,m,p))$, i.e., the structure of the model components before updating them with information from observation data to obtain their posterior distributions (Lasgorceux et al., 2026c).

Model evaluation and post-processing

The model introduced above is quite comprehensive, allowing for a thorough examination and comparison of its results on a species-by-species basis. In the following sections, we describe the post-processing approach to synthesizing this information. We address questions pertaining to the model's evaluation and also delve into the three components of the linear predictor (Equation 6): those concerning the fixed effects – specifically, habitat preferences compared to the target group denoted by β_i and γ_i; the spatial intra-annual component, characterized by the month-dependent Gaussian field $W_i^{(m)}$, which is associated with migratory status; and the inter-annual component, represented by the random i.i.d. Gaussian effect $f_i(p)$, which captures long-term trends in relative abundance.

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

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

Next, we assessed the model's generalization skill for predicting new data using either fully random k-fold, or spatial, or temporal cross-validation scenarios. In k-fold cross-validation, the dataset is divided into k equal parts, with one part used for evaluation while the rest is used for training (Jung, 2018). This approach yields $k+1$ AUC scores per species: k scores from predicting each part separately, and one from predicting the entire dataset. We selected $k = 10$ to ensure comparability with temporal and spatial scenarios, where the training data represented between 3.9% and 13.7% of the total volume. This choice aimed to maintain consistency in the quantity of test data across scenarios, ensuring equitable evaluation conditions. In the spatial scenario, we assessed model robustness by alternately removing all spatio-temporal cells within a given municipality of the Écrins National Park. Seven municipalities were selected based on geographic zones and data availability, resulting in seven AUC scores per species. For the temporal scenario, we alternately randomly removed half of the data for a given period, resulting in seven AUC scores per species. Further details on the size of the training datasets are given in Appendix E in Lasgorceux et al., 2026c.

Model evaluation using reliability plots. While AUC values computed under cross-validation provide an assessment of model generalization and potential overfitting by comparing only the ranks of prediction values in comparison to observations, we also compute reliability plots to obtain a more informative evaluation of predictive performance and potential prediction biases (Bröcker and Smith, 2007). Reliability plots compare the predicted mean intensity with the observed mean occurrence per cell by using appropriate aggregation schemes to form groups of cells. Ideally, points should align along the diagonal line in such plots, indicating good calibration. To further investigate model performance, we stratified the cells according to three factors: (i) elevation, to assess performance across the altitudinal gradient; (ii) predicted intensity, to evaluate accuracy across the full range of low to high predictions; and (iii) the number of species records in the target group, to examine the influence of sampling effort on predictive accuracy. Cells were stratified into 10 groups for each factor (elevation, intensity, and sampling effort), using quantiles or iso-intensity bins as appropriate.

Fixed effects. We estimated the *a posteriori* distribution of coefficients β_i and γ_i separately for each species i (Equation 6). The posterior mean of each parameter and its sign quantify the importance of the covariates in explaining occurrences of species i relative to the occurrences of all the passerines taken together. The credibility interval of each parameter helps determine how strongly the corresponding covariate influences the species in comparison to the whole pool of species; if the interval does not contain zero, we can consider the covariate as being "significant" (by analogy with the frequentist notion of significance). To synthesize the estimated mean coefficients and assess parameter similarities across species occupying comparable ecological habitats, we conducted Principal Component Analysis (PCA) on estimated parameters, followed by species clustering based on the estimated mean effects. This clustering was performed using a hierarchical agglomerative clustering (HAC) algorithm implemented through the *hclust* function, employing the *ward.D2* method (Murtagh and Legendre, 2014).

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

Inter-annual effects. We introduced an inter-annual effect $f_i(p)$ for each 4-year period p in the linear predictor (Equation 6) to describe how the relative abundance of species i changes over time, and to compare these changes with standardized survey trends (STOC, STOM, MHB). Using INLA, we obtained a posterior distribution for $f_i(p)$ in each 4-year period and assumed that the relative abundance of species i is constant within that period. To recover a temporal trend rate at the yearly scale from these coarse 4-year estimates, we fitted a linear regression to obtain predictions $\hat{f}_i(y)$ of $f_i(y)$, the inter-annual effect for the given year y and species i. Note that we generated 1000 posterior realizations of $f_i(p)$ to propagate uncertainty towards $\hat{f}_i(y)$. Because our temporal effect estimates $\hat{f}_i(y)$ measure changes in each species relative to the target group rather than in absolute terms, they must be interpreted with caution, as they reflect deviations from the overall community trend and not the species' standalone population change.

To compare our model-based trends with standardized surveys, we used the species-specific percentage changes reported by each survey $k \in \{\text{STOC}, \text{MHB}\}$ between years y_1 and y_2, denoted $t_i^{(k)}(y_1, y_2)$ (in %). On the log scale used in our model, this percentage change corresponds to $\log\left(\frac{t_i^{(k)}(y_1,y_2)}{100} + 1\right)$, which is the change in the log abundance index for species i over the period $[y_1, y_2]$ in survey k. Survey reports provide these trends for individual species but not for the target group as a whole. To obtain an approximate survey-based target-group trend, we therefore computed a weighted average of the species-specific percentage changes:

$$t_{TG}^{(k)}(y_1, y_2) = \sum_j \omega_j \, t_j^{(k)}(y_1, y_2), \qquad \omega_j = \frac{n_j^{(\text{ENP})}}{n_{TG}^{(\text{ENP})}},$$

where $n_j^{(\text{ENP})}$ is the number of detections of species j in the Écrins National Park and $n_{TG}^{(\text{ENP})}$ is the total number of detections for the whole target group. On the log scale, the corresponding change for the survey target group is approximated by $\log\left(\frac{t_{TG}^{(k)}(y_1,y_2)}{100} + 1\right)$. We then express survey-based trends relative to the survey target group by subtracting these two log-changes, $\log\left(\frac{t_i^{(k)}(y_1,y_2)}{100} + 1\right) - \log\left(\frac{t_{TG}^{(k)}(y_1,y_2)}{100} + 1\right)$, and compare this quantity to the corresponding model-based difference $\hat{f}_i(y_2) - \hat{f}_i(y_1)$.

For STOM, we have access to the observation data, and temporal trends $t_i^{(STOM)}(2015, 2024)$ were estimated on the log scale using a Generalized Linear Model with Poisson response fitted to annual counts. Then, the target-group trend $t_{TG}^{(STOM)}(2015, 2024)$ was computed using the same weighting scheme, and we compared $t_i^{(STOM)}(2015, 2024) - t_{TG}^{(STOM)}(2015, 2024)$ with $\hat{f}_i(2024) -$

$\hat{f}_i(2015)$. A STOM trend was considered significant when the p-value associated with the year covariate was below 0.05.

Results

Model evaluation using AUC

Figure 1 displays the boxplots of AUC values computed species-by-species as described in subsection *Materials and methods – Model evaluation and post-processing – Model evaluation using AUC*. There is a clear improvement in AUC when using the full linear predictor compared to using only the sampling-effort proxy as offset. This confirms that sampling methods do not solely constrain opportunistic presence-only data, and valuable information can be extracted from data.

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

Model evaluation using reliability plots

Figure 2 presents the reliability plot for cells stratified by elevation. Predicted and observed occurrences align closely with the diagonal axis for most groups, except at very high elevations (group 9), where occurrences are often overestimated. This pattern indicates insufficient penalization of the elevation PCA covariate, suggesting that a random or spline effect may be more appropriate. As group 9 accounts for only 10% of the cells, the model nevertheless remains well calibrated along most of the altitudinal gradient.

Figures 15 and 16 (Appendix F) show similar reliability analyses. In the iso-intensity stratification (Figure 15), points remain near the diagonal axis, but dispersion increases in higher groups due to smaller sample sizes and greater prediction variance. An analogous effect is observed in the target-group stratification (Figure 16), where higher-record groups exhibit greater variability.

Fixed effects

The mean estimated effects of each covariate and their significance are detailed in Tables 6-8 in Appendix G (Lasgorceux et al., 2026c). To summarize the information on spatial covariates, we provide a classification into three clusters based on Principal Component Analysis conducted on the estimated mean coefficients for each spatial covariate (see subsection '*Materials and Methods – Model evaluation and post-processing – Fixed Effects*'). We did not include any species with less than two significant covariates since we consider that the model does not provide enough information on them. These species are generally those with very few occurrences, as 10 out of these 11 species have fewer than 551 occurrences over the 28 years of data, where the value 551 is below the 25th-percentile of per-species occurrence numbers. The list of removed species is given in Table 9 in Appendix G (Lasgorceux et al., 2026c). The PCA reveals two principal axes explaining 62.2% of the classification: the first axis distinguishes vegetated habitat (coniferous and

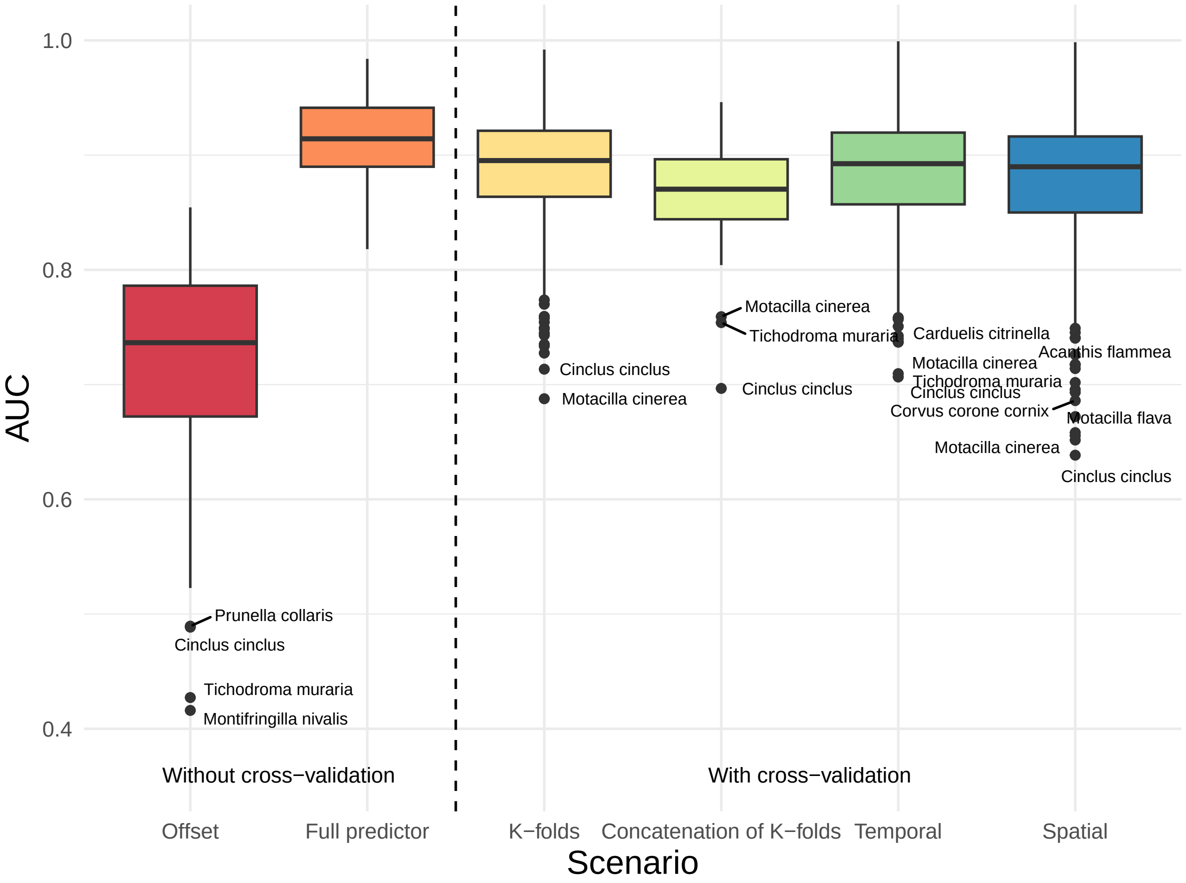

Figure 1 – AUC values calculated for various linear-predictor and cross-validation scenarios. Boxplots summarize the set of AUC values for all species, and in the case of cross-validation for each combination of species and hold-out dataset. Notations are as follows: "Offset" – model where the linear 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 of the data which occurred in a given municipality, for seven municipalities. 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 – Model evaluation and post-processing – Model evaluation using AUC*' subsection.

deciduous forests VS urban wetlands), while the second axis highlights an altitudinal gradient. The three clusters are presented in Figure 3. We interpret them as follows. Cluster 1: Species mostly found in closed habitats such as coniferous and deciduous forests (19 species–green); Cluster 2: Species mostly found in high-elevation environments (7 species–orange); Cluster 3: Species mostly found in open habitats and valleys (37 species–violet). These results are coherent when comparing a map of elevation with a map of clusters with the highest occurrences on a 500m×500m regular grid in the Écrins National Park (see Figure 17 in Appendix G in Lasgorceux et al., 2026c).

The ellipse associated with forest species represents an altitudinal gradient, with species found in low-elevation forests (e.g. *Pyrrhula pyrrhula*), to medium (e.g. *Anthus trivialis*) and high-elevation forests (e.g. *Turdus torquatus*). We also observe this altitudinal gradient in the cluster

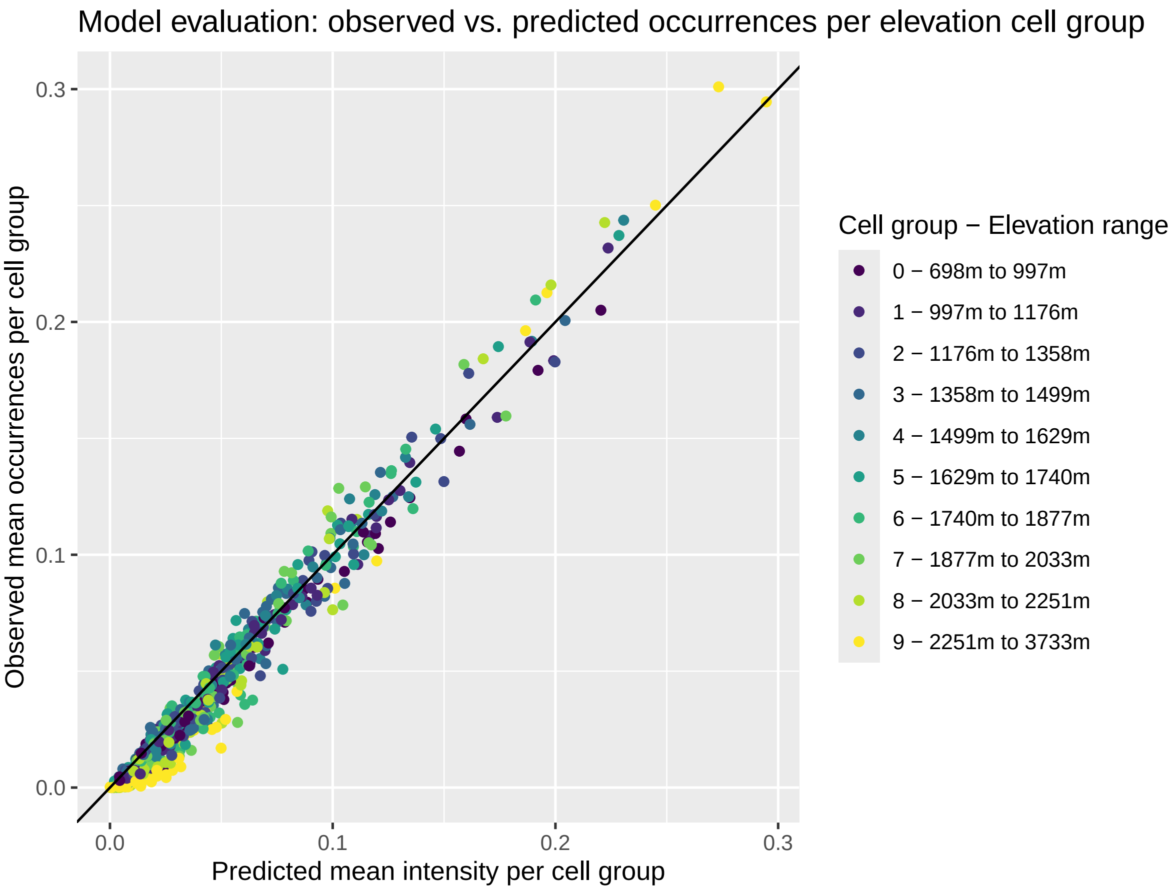

Figure 2 – Reliability plot stratified by elevation. Each point represents observed vs. predicted occurrences for a given species within an elevation class. Cells were ranked by the altitudinal gradient and then grouped into ten groups containing an equal number of cells.

associated with species from open habitats and valleys, with rocky habitat species such as *Tichodroma muraria* and *Phoenicurus ochruros*, at the boundary of the high-elevation species cluster. The cluster associated with species specialized to high-elevation habitats is smaller but well-defined, with typical species such as *Prunella collaris* and *Montifringilla nivalis*.

To compare the contribution of spatio-temporal covariates with spatial ones in the linear predictor, we computed the percentage of explained variance for each covariate with respect to 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. The results are presented in Figure 4. Our analysis revealed that, for most species, PCA Axis 1, interpreted as being related to high-elevation environments, explains the majority of the variance (Median=54%). Axis 2 and 3, interpreted as environments with medium to high elevation vegetation and grasslands, respectively, explain a substantial portion of the variance for about twenty species (Q3=27% and 15%).

Regarding spatio-temporal covariates, precipitation anomalies had the smallest percentage of explained variance, with a median below 10^{-5}%. This is further supported by only 12 out of 76 species showing a significant parameter associated with precipitation anomalies. Similarly, the explained variance for maximum temperature anomalies is low for most species (Median < 10^{-3}%); however, 45 species showed a significant parameter. It notably explained a significant amount

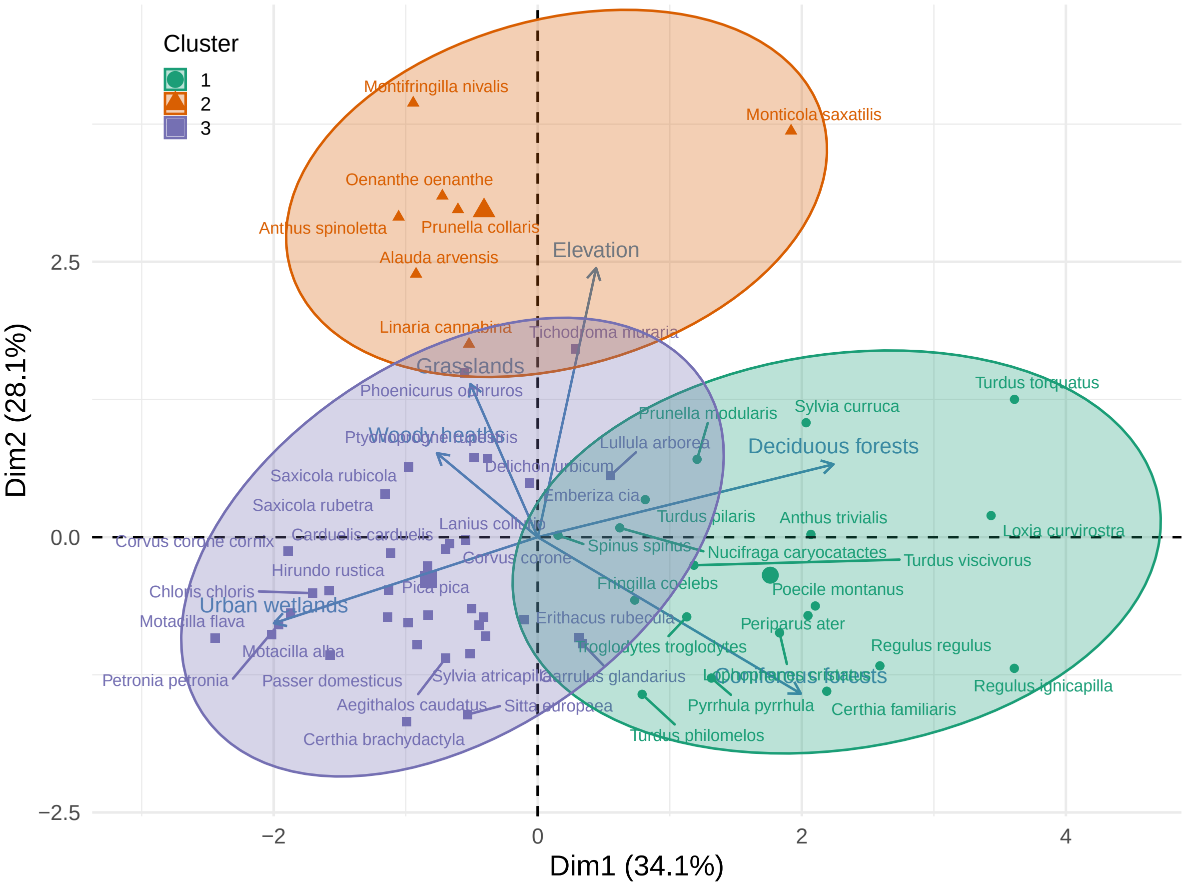

Figure 3 – Classification of 63 species in three clusters based on the estimated mean effects of spatial covariates. A Principal Component Analysis was conducted on the species with at least two credibility intervals that do not include 0. The two main axes explain 62.2% of the variance. Cluster 1: Species 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).

of information for five species: *Coccothraustes coccothraustes*, *Fringilla montifringilla*, *Lanius collurio*, *Spinus spinus*, and *Turdus pilaris*. These results align with the ecological characteristics of these species, confirming that the model effectively correlates temperature anomalies with the presence of thermophilic and cold-resistant species.

Intra-annual effect

The first two extracted harmonics capture a significant portion of the intra-annual variability, explaining 83% of the variance. The primary harmonic shows a stronger positive effect in winter months compared to summer months, while the second harmonic has a positive effect in spring and summer and a negative effect in winter (see Figure 18 in Appendix H in Lasgorceux et al., 2026c). Thus, species with positive scores on the first harmonic, referred to as the winter harmonic, are predominantly overrepresented in winter. Conversely, species with positive scores on the second harmonic, now termed the warm months harmonic, are more likely to be overrepresented in spring and summer, suggesting a potential link with migratory status.

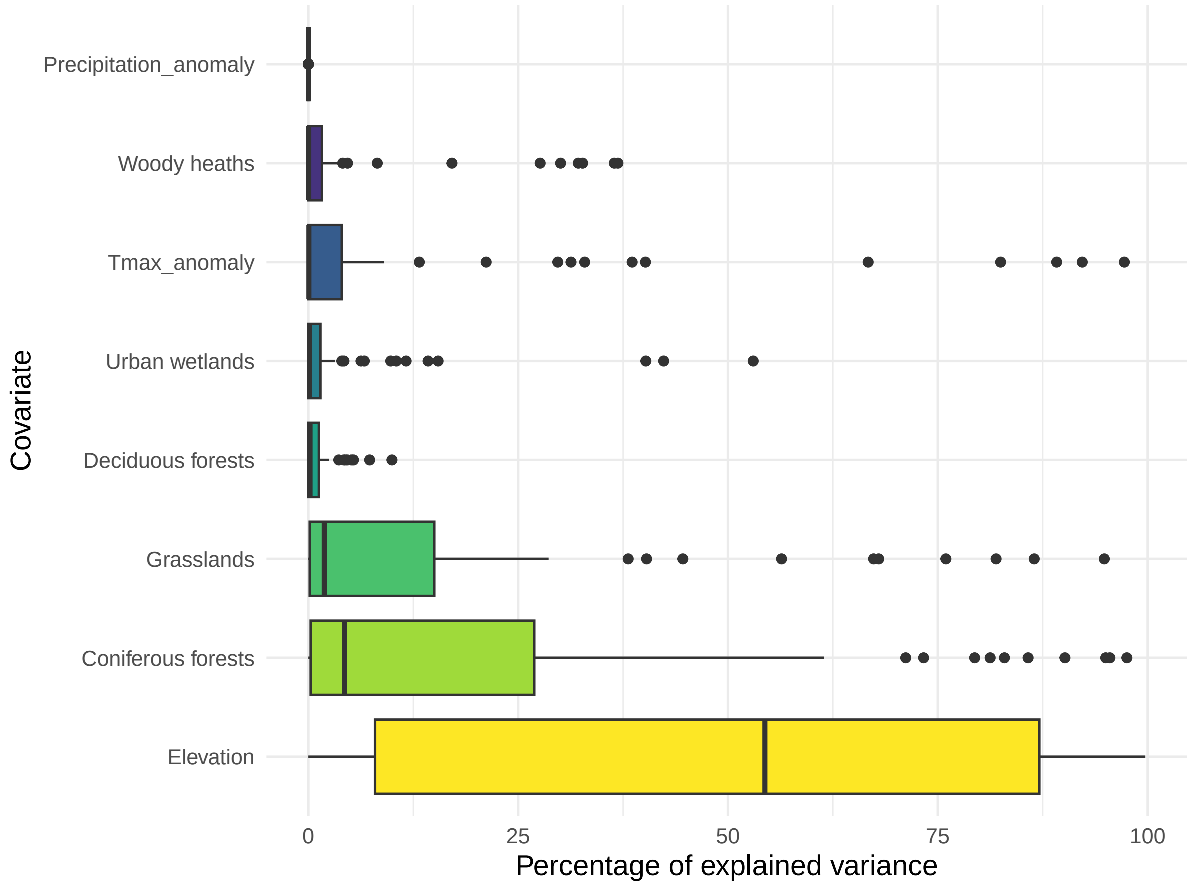

Figure 4 – Percentage of explained variance of each covariate for 74 species. The covariates include six spatial factors derived from PCA analysis and two spatio-temporal climatic factors. The interpretation of the PCA axes is detailed in the '*Materials and Methods – Data – Preprocessing of model covariates*' subsection. Covariate boxplots are ordered by their median values.

Figure 5 displays the harmonic scores for each species, categorized by migratory status within the territory of the Écrins National Park, as provided by experts from the park. This analysis confirms that species overrepresented in winter are mainly sedentary, while those overrepresented in warm months are migratory. Detailed scores for each species are provided in Table 10 in Appendix H (Lasgorceux et al., 2026c). These harmonics effectively distinguish between sedentary and migratory species in the Écrins National Park, affirming that the month-based spatio-temporal Gaussian field $W_i^{(m)}$ accurately accounts for spatial and intra-annual species variations not explained by physical covariates across the territory.

Illustrations of $W_i^{(m)}$ for each species i and each month m, through PDF and interactive Leaflet maps, are available in the Data, Scripts, and Code section (Lasgorceux et al., 2026a). An inspection of these maps reveals several interpretable patterns. For certain species, a clear seasonal signal emerges that reflects migratory behaviour (e.g. *Phylloscopus collybita*), consistent with the FPCA results, while for others, the signal appears more related to observation effort and accessibility. Areas with easier access or better observation conditions tend to stand out consistently across months, such as the north-eastern part of the Park around the Lautaret road (e.g. *Turdus merula*, *Erithacus rubecula* in October). However, these maps do not clearly capture the altitudinal movements within the Park for certain species such as *Prunella collaris*, *Monticola*

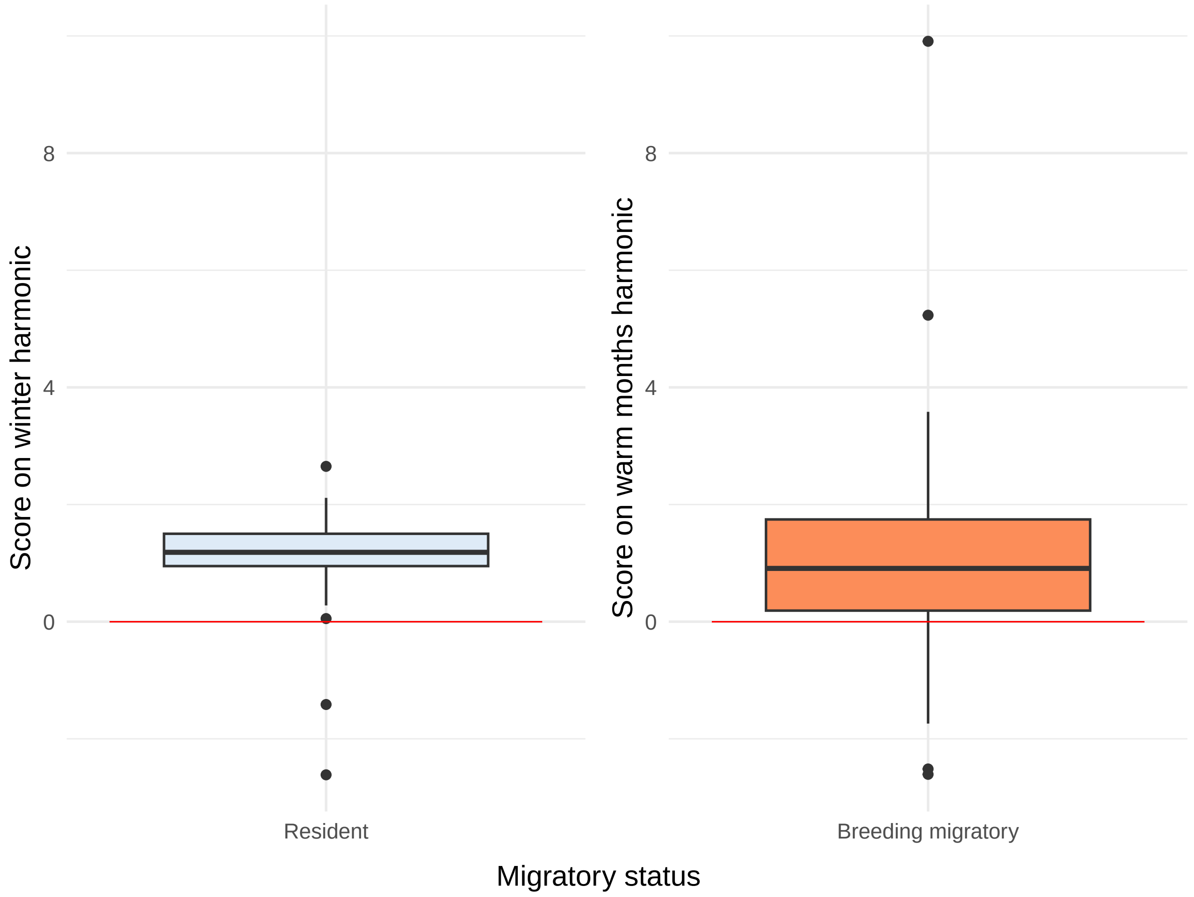

Figure 5 – Distribution of species scores on the two dominant FPCA harmonics of the month-based spatio-temporal Gaussian field $W_i^{(m)}$. The first harmonic (winter component) is shown for resident species; the second (warm months component) for breeding migratory species, based on expert classifications from Écrins National Park. The two harmonics account for 83% of the total variance. The red line marks a score of zero.

saxatilis, or *Phoenicurus ochruros*. Some artefacts are harder to explain, such as anomalies for species such as *Alauda arvensis* and *Cinclus cinclus*.

Inter-annual effect

The estimated percentage variations in relative abundance in the Écrins National Park between 1994 and 2021, based on 1000 *a posteriori* samples, are illustrated in Figure 6. Species associated with the high-elevation habitats cluster (7 species, Figure 3) exhibit an average decline of -15.1% in relative abundance during 1994–2021. This is mainly due to the strong decrease of four species: *Monticola saxatilis* (-51.2%), *Anthus spinoletta* (-22.5%), *Prunella collaris* (-16.6%) and *Montifringilla nivalis* (-15.7%). Conversely, species affiliated with the forest cluster (19 species) experienced a growth of 8.3% over the same period by removing the outlier trend of species *Regulus ignicapilla* (+220%). Figure 6 shows that 8 out of the 19 forest species exhibited a positive trend while only 2 of them showed a negative trend based on the 90% credibility interval. Species linked to the valleys cluster (37 species) experienced an average decrease of -1.2% in relative abundance.

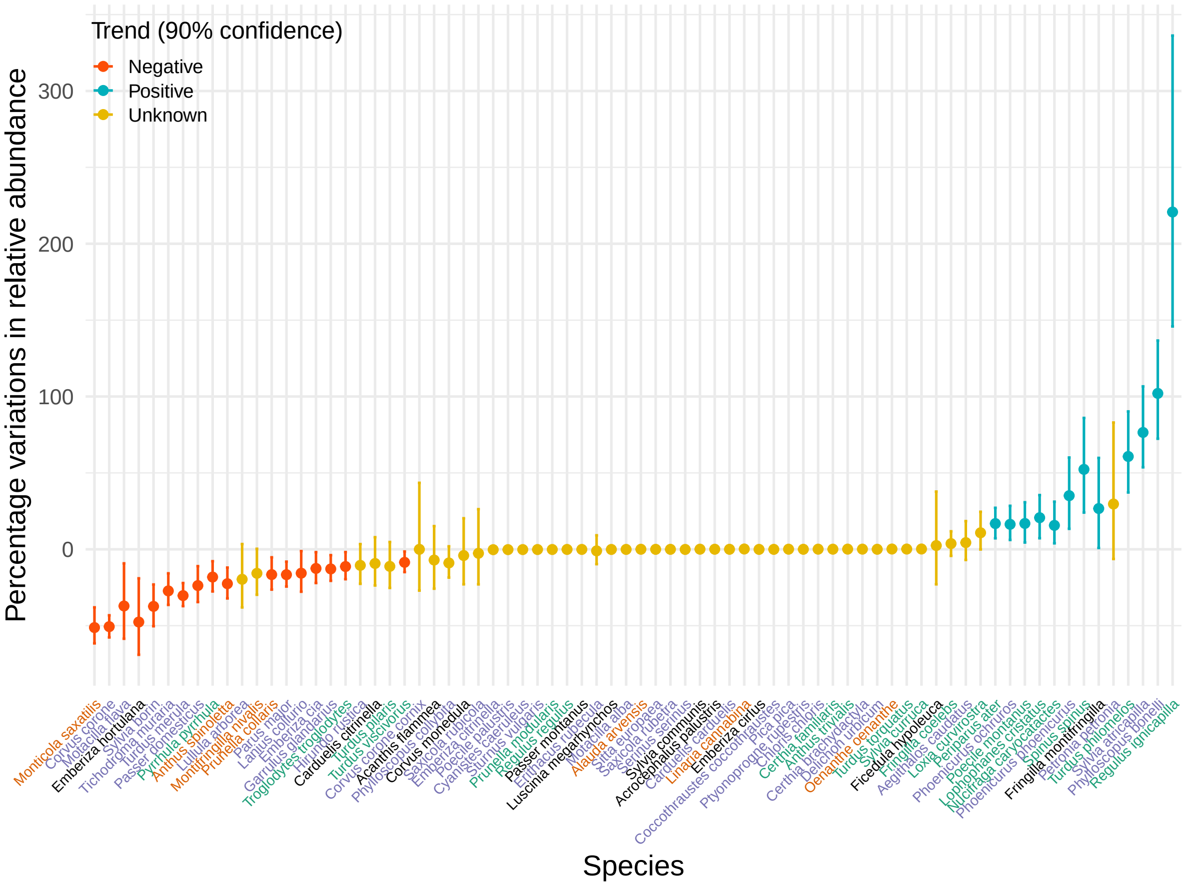

Figure 6 – Illustration of the estimated median percentage variations in relative abundance and credibility intervals for 74 retained species in the Écrins National Park spanning the period 1994–2021. Trends are derived from 1000 samples of the *a posteriori* distribution of the inter-annual effect $exp(f_i(p))$. Details can be found in the '*Materials and Methods – Model evaluation and post-processing – Inter-annual effects*' subsection. 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 3. Species with fewer than two significant covariates were not classified and are shown in black.

Trend comparison with other datasets

To assess the robustness of our modelling approach, we restricted our comparison to species with significant trends in the different datasets. For STOM, only two species show significant trends in both STOM-based models and the Écrins National Park (ENP) opportunistic data: *Phoenicurus ochruros* and *Prunella collaris*. Trends were consistent for *P. collaris* (decline in both datasets), whereas *P. ochruros* increased in ENP but declined in the survey-based datasets; this particular case is discussed further below.

Table 1 summarizes the sign of significant trends for ENP vs. STOC and ENP vs. MHB, grouped by main habitat (see also Appendix I for the details of the contingency tables). Overall, about two-thirds of the species lie on the diagonal (i.e., the same sign of trend in ENP and in the corresponding survey), indicating a substantial agreement in the direction of change when trends are significant. Below, we highlight only the main patterns by habitat rather than discussing each species individually.

All four alpine species considered in MHB show declining trends in both ENP and MHB, supporting a consistent signal of decline in high-elevation specialists. These species are not monitored by STOC.

For forest birds, agreement between ENP and national surveys is moderate. About half of the species with significant trends show the same sign in ENP and MHB, and the correlation between corrected trends is positive. Discrepancies mainly involve very common species (e.g. *Parus major*, *Turdus merula*) that increase in MHB but appear to decline in ENP. Because these species are widespread and abundant, such mismatches are likely driven by changes in reporting behaviour in the opportunistic data that are not fully captured by our sampling-effort proxy. For species with significant trends in both ENP and STOC, most show a more favourable trend in ENP than in STOC, which is consistent with a protective effect of the national park.

For farmland birds, trends in ENP and in the MHB and STOC surveys generally point in the same direction: four of the five species with significant trends share the same sign in the surveys. The only exception is *Corvus corone*, which increases in MHB but decreases in ENP; as a very common species, this mismatch is likely explained by a decline in reporting rates in the opportunistic ENP data rather than by a genuine local decline. The protected-area effect is particularly clear for *Emberiza hortulana*, which still occurs but declines in ENP, whereas it is already extinct as a breeding species in Switzerland; we therefore excluded this species from the correlation analysis with MHB.

Species associated with settlements were only represented by *Passer domesticus*, which shows a clear decline in all datasets, with a steeper decrease in ENP than in STOC or MHB. The stronger ENP decline is again likely amplified by a reduction in reporting rates over time in opportunistic data, although all sources agree on the overall negative trend.

Generalists show the highest rate of disagreement between ENP and national surveys. The most striking case is *Phoenicurus ochruros*, which increases in ENP but declines in STOC, STOM and MHB. In the ENP, the mean elevation of records for this species increases over time, especially during summer months (approximately +250 m over 25 years) compared to winter months (approximately +100 m over 25 years) (Figure 19, Appendix J), suggesting an upward shift of its elevational range. However, this alone cannot explain the negative trend observed in STOM, where sampling is concentrated around 2000 m. A plausible complementary explanation is that park rangers progressively increased reporting of high-altitude individuals after 2010, leading to an overestimation of the ENP trend. Other generalist species with mismatched trends (e.g. *Corvus corone*, *Garrulus glandarius*, *Parus major*, *Turdus merula*) are again very common birds for which small changes in reporting behaviour can strongly affect opportunistic indices.

Discussion

In our study, we demonstrated that under the assumption of sampling homogeneity across species (but where sampling effort is still allowed to vary in space and time), opportunistic presence-only data could be leveraged to estimate habitat preferences, migratory status, and relative abundance trends of passerine species. This finding is not novel *per se*, as numerous studies have utilized such data to model species distributions or abundance trends (Botella et al., 2020; Bradter et al., 2018; Kéry et al., 2010; Phillips et al., 2009; Valavi et al., 2022; Van Strien et al., 2013). However, it is one of only a few approaches achieving this within a unique spatio-temporal model. To our knowledge, the most similar work has been done by Giraud et al., 2016

Table 1 – Numbers of species with significant negative (“neg”) or positive (“pos”) trends in the Écrins National Park (ENP) and in the national surveys (STOC, MHB), by species group. Left block: ENP vs STOC comparison; right block: ENP vs MHB comparison. Correlations are between corrected trends (see ‘*Materials and Methods – Model evaluation and post-processing – Inter-annual effects*’).

	Comparison with STOC					Comparison with MHB				
Species group	ENP trend		STOC trend		corr	ENP trend		MHB trend		corr
	neg	pos	neg	pos	STOC/ENP	neg	pos	neg	pos	MHB/ENP
Alpine	0	0	0	0	–	4	0	4	0	0.42
Forest	2	3	3	2	0.85	6	6	3	9	0.40
Farmland	1	0	1	0	–	4	0	3	1	−0.88[†]
Settlements	0	1	0	1	–	1	0	1	0	–
Generalist	4	1	0	5	0.85	1	1	2	0	1.00

Counts are the numbers of species with significant negative (“neg”) or positive (“pos”) trends in each dataset within a species group. “–” indicates that a correlation was not computed (e.g. because of too few species).

[†] Correlation for farmland-MHB computed after excluding *Emberiza hortulana*, which no longer breeds in Switzerland.

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

The comparison between corrected trends from the three survey-based datasets and those produced by our model demonstrates its overall ability to capture significant trends across datasets. Most species showing discrepancies are in fact common species, suggesting that our assumption of homogeneous reporting is overly optimistic and only partially valid. The comparison of forest species, as well as the case of *Emberiza hortulana*, also indicates a generally better conservation status within the Écrins National Park, likely due to its protected-area status. Finally, the inconsistency between the three survey datasets and the ENP trend for *Phoenicurus ochruros* raised the possibility of an upward elevational shift for this species in the Écrins National Park. A remaining blind spot in our comparison concerns non-significant trends. Some species show significant trends in ENP but not elsewhere, or vice versa. We chose not to explore these cases in detail because it would be difficult to disentangle the underlying causes – such as protected-area effects, temporal variation in reporting prevalence, differences in agricultural practices, or population shifts into Switzerland. With smaller datasets, there is also less statistical power, such that certain trends can be statistically non significant even they really exist. In 5 to 10 years, a new comparison using updated STOM data should be more robust and may provide a valuable opportunity to reassess these patterns. Such comparisons may even allow for the correction of reporting-rate biases for certain species using survey-based benchmarks.

We identify four areas for future improvements in our modeling approach using solely opportunistic data:

(i) The assumption of homogeneous sampling of different species, which could be questionable due to differences in detectability, species commonness/rarity (Snäll et al., 2011), or agent identification skills and reporting prevalence (Bradter et al., 2018), albeit reasonable in practice and lacking better options. Assessing the robustness of our model to potential violations of this homogeneity assumption through dedicated simulation scenarios would therefore be a valuable direction for future research. Another perspective to account for heterogeneities in species sampling comes from Integrated Species Distribution Models (IDM) that can combine massive but biased opportunistic data with scarce, high-quality standardized data to build more robust biodiversity predictions in space and time (Dorazio, 2014; Fithian et al., 2015; Isaac et al., 2020; Miller et al., 2019). However, this approach is so far challenged by important remaining issues, such as the imbalance of opportunistic and standardized data, the heterogeneity of spatio-temporal scales across datasets, and limits in the flexibility, scaling, and interpretation of IDMs (Fletcher et al., 2019; Isaac and Pocock, 2015; Renner et al., 2019; Warton et al., 2013). Heterogeneous detectability remains an issue without data based on standardized sampling protocols and seems difficult to correct, except for trying to include prior expert knowledge into models. Different reporting behavior due to species commonness/rarity can introduce bias in estimated intercept terms, such as α_i in Equation 6, but would not have any impact on the other terms varying in space and/or time, as long as this bias in reporting behavior is homogeneous across space, time, and physical predictors. Effects related to agent identification skills and reporting prevalence could potentially be identified by including random effects for the rangers in the model. Another possibility would be to reformulate the $\mu_i(s, m, p)$ in Equation 6 to incorporate the rate of reporting as follow

$$\mu_i(s, m, p) = \frac{r_i \times \Lambda_i(s, m, p)}{\sum_i (r_i \times \Lambda_i(s, m, p))} \times \sum_i (E(s, m, p) \times r_i \times \Lambda_i(s, m, p)) . \tag{7}$$

However, the model would become less interpretable if we explicitly modeled these rates, especially when comparing our trends with those derived from standardized datasets since we would estimate the “target-group reported relative abundance of species i”. As a general recommendation to reduce biases and improve the exploitability of opportunistic presence-only data, we suggest that when observers report a species, all species of the same taxon should also be reported for better use of opportunistic presence-only data in species distribution modeling.

(ii) In our case study, very large occurrence numbers were available for the target group, such that uncertainties about the cumulative abundance of the target group can be expected to be relatively small. For modeling more precisely the cumulative abundance of the target group, for example in cases with smaller occurrence numbers or when large areas of the study domain are only very weakly sampled, we could develop an alternative approach wherein $Y_{TG}(s, m, p)$ could be initially modeled with a Poisson distribution using a mean parameter $\mu_{TG}(s, m, p)$, fitted to the data to estimate $\mu_{TG}(s, m, p)$. Subsequently, we would use the estimated Poisson mean as a proxy for $\Lambda_{TG}(s, m, p)E(s, m, p)$, rather than the raw count. This modification could potentially enhance robustness and improve the assessment of uncertainties inherent to our modeling approach.

(iii) The acquired information is relative to a target group of species since we used the target-group occurrences as a proxy for the sampling effort (Equation 4). This implies that the

potential for interpretations of results in terms of absolute abundance and its variation across covariates, space, and time remains relatively limited, as long as we do not have additional external information on the absolute abundance of the target group as a whole. For example, the mathematical developments presented in the "*Materials and Methods – Model evaluation and post-processing – Inter-annual effects*" subsection were specifically designed to compare the relative abundance trends derived from our opportunistic presence-only model with population trends obtained from standardized surveys. However, this comparison operates on the log scale, which is less directly interpretable than the percentage variation in population size typically reported by survey-based monitoring programs.
(iv) There are multiple ways to define this target group, which impacts the interpretation of abundances relative to it. The composition of this target group primarily relies on available data sources, ranging from studies covering a single species (Farr et al., 2021) to those encompassing thousands (Botella et al., 2021). Some studies focus exclusively on species within the same taxonomic group (Van Strien et al., 2013), while others consider species across different taxa (Escamilla Molgora et al., 2022). Botella et al., 2020 recommends choosing a set of species that are consistently abundant across a broad range of environmental sub-regions, which is what we did in this study.

We have recognized a pressing need for more comprehensive descriptions of what we refer to as "sampling effort" and "observer bias" when dealing with opportunistic presence-only data, as well as for methodologies to effectively incorporate it into modeling endeavors. Various strategies have been proposed to address sampling efforts and observer bias depending on the studied taxa, the available datasets, and model type. These strategies include the use of random or target-group pseudo-absences with Maxent (Phillips et al., 2006, 2009), proxy variables representing site accessibility (Henckel et al., 2020; Moreira et al., 2024; Warton et al., 2013) and refinements (Chauvier et al., 2021), integration of information from detection/nondetection data (Dorazio, 2014; Fithian et al., 2015; Giraud et al., 2016), and questionnaires (Bradter et al., 2018).

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

Acknowledgements

Preprint version 4 of this article has been peer-reviewed and recommended by Peer Community In Ecology (`https://doi.org/10.24072/pci.ecology.100750`; Yoccoz, 2026).
The authors would like to thank all the staff of Écrins National Park and its scientific team for their contributions to the dataset. The authors are also deeply grateful to Nicolas Strebel for generously sharing the Monitoring Breeding Bird data that made part of this work possible. The authors also appreciate the advice of Denis Allard on using functional principal component analysis, as well as technical support provided by Loïc Houde for the high-performance computing cluster used to implement the models in this case study.

Fundings

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

Conflict of interest disclosure

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

Data, script, code, and supplementary information availability

Data, scripts, and code can be accessed online (`https://doi.org/10.5281/zenodo.19864262`; Lasgorceux et al., 2026a).
Outputs from fitted models under various cross-validation scenarios are also available online (`https://doi.org/10.5281/zenodo.19487370`; Lasgorceux et al., 2026b).
Supplementary information is accessible online (`https://doi.org/10.5281/zenodo.19864179`; Lasgorceux et al., 2026c).

References

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

Belmont J, Martino S, Illian J, Rue H (2024). *Spatio-temporal occupancy models with INLA*. *Methods in Ecology and Evolution* **15**, 2087–2100. `https://doi.org/10.1111/2041-210X.14422`.

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

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

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

Bröcker J, Smith LA (2007). *Increasing the Reliability of Reliability Diagrams*. *Weather and Forecasting* **22**, 651–661. `https://doi.org/10.1175/waf993.1`.

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

Chauvier Y, Zimmermann NE, Poggiato G, Bystrova D, Brun P, Thuiller W (2021). *Novel methods to correct for observer and sampling bias in presence-only species distribution models*. *Global Ecology and Biogeography* **30**, 2312–2325. `https://doi.org/10.1111/geb.13383`.

Chiffard J (2024). *Suivi Temporel des Oiseaux communs en Montagne (STOM) : Le premier bilan ! Période 2014 - 2023*. Office Français de la Biodiversité ; Parc National des Pyrénées ; Parc National des Ecrins ; Parc National du Mercantour ; Parc National de la Vanoise ; Centre d'écologie fonctionnelle et évolutive. URL: `https://ofb.hal.science/hal-05074160`.

Coron C, Calenge C, Giraud C, Julliard R (2018). *Bayesian estimation of species relative abundances and habitat preferences using opportunistic data*. *Environmental and Ecological Statistics* **25**, 71–93. `https://doi.org/10.1007/s10651-018-0398-2`.

Donald PF, Green RE, Heath MF (2001). *Agricultural intensification and the collapse of Europe's farmland bird populations*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**, 25–29. `https://doi.org/10.1098/rspb.2000.1325`.

Dorazio RM (2014). *Accounting for imperfect detection and survey bias in statistical analysis of presence-only data*. *Global Ecology and Biogeography* **23**, 1472–1484. `https://doi.org/10.1111/geb.12216`.

Dray S, Dufour AB (2007). *The ade4 Package: Implementing the Duality Diagram for Ecologists*. *Journal of Statistical Software* **22**, 1–20. `https://doi.org/10.18637/jss.v022.i04`.

Duckworth GD, Altwegg R (2018). *Effectiveness of protected areas for bird conservation depends on guild*. *Diversity and Distributions* **24**, 1083–1091. `https://doi.org/10.1111/ddi.12756`.

Eftekhari A, Pasadakis D, Bollhöfer M, Scheidegger S, Schenk O (2021). *Block-enhanced precision matrix estimation for large-scale datasets*. *Journal of Computational Science* **53**, 101389. `https://doi.org/10.1016/j.jocs.2021.101389`.

Escamilla Molgora JM, Sedda L, Diggle P, Atkinson PM (2022). *A joint distribution framework to improve presence-only species distribution models by exploiting opportunistic surveys*. *Journal of Biogeography* **49**, 1176–1192. `https://doi.org/10.1111/jbi.14365`.

Farr MT, Green DS, Holekamp KE, Zipkin EF (2021). *Integrating distance sampling and presence-only data to estimate species abundance*. *Ecology* **102**, e03204. `https://doi.org/10.1002/ecy.3204`.

Fawcett T (2006). *An introduction to ROC analysis*. *Pattern Recognition Letters* **27**, 861–874. `https://doi.org/10.1016/j.patrec.2005.10.010`.

Fick SE, Hijmans RJ (2017). *WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas*. *International Journal of Climatology* **37**, 4302–4315. `https://doi.org/10.1002/joc.5086`.

Fithian W, Elith J, Hastie T, Keith DA (2015). *Bias correction in species distribution models: pooling survey and collection data for multiple species*. *Methods in Ecology and Evolution* **6**, 424–438. `https://doi.org/10.1111/2041-210X.12242`.

Fithian W, Hastie T (2013). *Finite-sample equivalence in statistical models for presence-only data*. *The Annals of Applied Statistics* **7**, 1917–1939. `https://doi.org/10.1214/13-AOAS667`.

Fletcher RJ, Hefley TJ, Robertson EP, Zuckerberg B, Mccleery RA, Dorazio RM (2019). *A practical guide for combining data to model species distributions*. *Ecology* **100**, e02710. `https://doi.org/10.1002/ecy.2710`.

Flousek J, Telenský T, Hanzelka J, Reif J (2015). *Population Trends of Central European Montane Birds Provide Evidence for Adverse Impacts of Climate Change on High-Altitude Species*. *PLOS ONE* **10**, e0139465. `https://doi.org/10.1371/journal.pone.0139465`.

Fontaine B, Moussy C, Chiffard Carricaburu J, Dupuis J, Schmaltz L, Lorrillière R, Loïs G, Gaudard C, Couzi L (2020). *Suivi des oiseaux communs en France 1989-2019 : 30 ans de suivis participatifs*. MNHN- Centre d'Ecologie et des Sciences de la Conservation, LPO BirdLife France - Service Connaissance, Ministère de la Transition écologique et solidaire, p. 46. URL: `https://www.vigienature.fr/sites/vigienature/files/atoms/files/syntheseoiseauxcommuns2020_final.pdf`.

Gaedke-Merzhäuser L, Niekerk J, Schenk O, Rue H (2023). *Parallelized integrated nested Laplace approximations for fast Bayesian inference*. *Statistics and Computing* **33**, 25. `https://doi.org/10.1007/s11222-022-10192-1`.

Gilks WR, Richardson S, Spiegelhalter D (1995). *Markov Chain Monte Carlo in Practice*. New York, NY: Chapman and Hall/CRC. 512 pp. `https://doi.org/10.1201/b14835`.

Giraud C, Calenge C, Coron C, Julliard R (2016). *Capitalizing on opportunistic data for monitoring relative abundances of species*. *Biometrics* **72**, 649–658. `https://doi.org/10.1111/biom.12431`.

Gregory RD, Skorpilova J, Vorisek P, Butler S (2019). *An analysis of trends, uncertainty and species selection shows contrasting trends of widespread forest and farmland birds in Europe*. *Ecological Indicators* **103**, 676–687. `https://doi.org/10.1016/j.ecolind.2019.04.064`.

Gregory RD, Vorisek P, Van Strien A, Gmelig Meyling AW, Jiguet F, Fornasari L, Reif J, Chylarecki P, Burfield IJ (2007). *Population trends of widespread woodland birds in Europe*. *Ibis* **149**, 78–97. `https://doi.org/10.1111/j.1474-919X.2007.00698.x`.

Harris I, Osborn TJ, Jones P, Lister D (2020). *Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset*. *Scientific Data* **7**, 109. `https://doi.org/10.1038/s41597-020-0453-3`.

Haupert S, Sèbe F, Sueur J (2023). *Physics-based model to predict the acoustic detection distance of terrestrial autonomous recording units over the diel cycle and across seasons: Insights from an Alpine and a Neotropical forest*. *Methods in Ecology and Evolution* **14**, 614–630. `https://doi.org/10.1111/2041-210X.14020`.

Heldbjerg H, Sunde P, Fox AD (2018). *Continuous population declines for specialist farmland birds 1987-2014 in Denmark indicates no halt in biodiversity loss in agricultural habitats*. *Bird Conservation International* **28**, 278–292. `https://doi.org/10.1017/S0959270916000654`.

Henckel L, Bradter U, Jönsson M, Isaac NJB, Snäll T (2020). *Assessing the usefulness of citizen science data for habitat suitability modelling: Opportunistic reporting versus sampling based on a systematic protocol*. *Diversity and Distributions* **26**, 1276–1290. `https://doi.org/10.1111/ddi.13128`.

Hijmans RJ (2024). *terra: Spatial Data Analysis*. R package version 1.7-71. URL: `https://CRAN.R-project.org/package=terra`.

Hill MO (2012). *Local frequency as a key to interpreting species occurrence data when recording effort is not known: A key to interpreting species occurrence data*. *Methods in Ecology and Evolution* **3**, 195–205. `https://doi.org/10.1111/j.2041-210X.2011.00146.x`.

Horns JJ, Adler FR, Şekercioğlu ÇH (2018). *Using opportunistic citizen science data to estimate avian population trends*. *Biological Conservation* **221**, 151–159. `https://doi.org/10.1016/j.biocon.2018.02.027`.

Huang J, Ling C (2005). *Using AUC and accuracy in evaluating learning algorithms*. *IEEE Transactions on Knowledge and Data Engineering* **17**, 299–310. `https://doi.org/10.1109/TKDE.2005.50`.

IGN (2018). *RGE ALTI® [Data set]*. URL: `https://geoservices.ign.fr/rgealti`.

Illian JB, Martino S, Sørbye SH, Gallego-Fernández JB, Zunzunegui M, Esquivias MP, Travis JMJ (2013). *Fitting complex ecological point process models with integrated nested Laplace approximation*. *Methods in Ecology and Evolution* **4**, 305–315. `https://doi.org/10.1111/2041-210x.12017`.

Inger R, Gregory R, Duffy JP, Stott I, Voříšek P, Gaston KJ (2015). *Common European birds are declining rapidly while less abundant species' numbers are rising*. *Ecology Letters* **18**, 28–36. `https://doi.org/10.1111/ele.12387`.

Inglada J, Vincent A, Thierion V (2018). *Theia OSO Land Cover Map 2018 [Data set]*. Zenodo. `https://doi.org/10.5281/zenodo.3613415`.

IPBES (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. Bonn, Germany. 56 pages. `https://doi.org/10.5281/zenodo.3553579`.

Isaac NJB, Jarzyna MA, Keil P, Dambly LI, Boersch-Supan PH, Browning E, Freeman SN, Golding N, Guillera-Arroita G, Henrys PA, Jarvis S, Lahoz-Monfort J, Pagel J, Pescott OL, Schmucki R, Simmonds EG, O'Hara RB (2020). *Data Integration for Large-Scale Models of Species Distributions*. *Trends in Ecology & Evolution* **35**, 56–67. `https://doi.org/10.1016/j.tree.2019.08.006`.

Isaac NJB, Pocock MJO (2015). *Bias and information in biological records*. *Biological Journal of the Linnean Society* **115**, 522–531. `https://doi.org/10.1111/bij.12532`.

Isaac NJB, Strien AJ, August TA, Zeeuw MP, Roy DB (2014). *Statistics for citizen science: extracting signals of change from noisy ecological data*. *Methods in Ecology and Evolution* **5**, 1052–1060. `https://doi.org/10.1111/2041-210X.12254`.

Janžekovič F, Novak T (2012). *PCA – A Powerful Method for Analyze Ecological Niches*. In: *Principal Component Analysis*. 8. Rijeka: IntechOpen. `https://doi.org/10.5772/38538`.

Jiguet F, Devictor V, Julliard R, Couvet D (2012). *French citizens monitoring ordinary birds provide tools for conservation and ecological sciences*. *Acta Oecologica* **44**, 58–66. `https://doi.org/10.1016/j.actao.2011.05.003`.

Jung Y (2018). *Multiple predicting K -fold cross-validation for model selection*. *Journal of Nonparametric Statistics* **30**, 197–215. `https://doi.org/10.1080/10485252.2017.1404598`.

Kamp J, Frank C, Trautmann S, Busch M, Dröschmeister R, Flade M, Gerlach B, Karthäuser J, Kunz F, Mitschke A, Schwarz J, Sudfeldt C (2021). *Population trends of common breeding birds in Germany 1990–2018*. *Journal of Ornithology* **162**, 1–15. `https://doi.org/10.1007/s10336-020-01830-4`.

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

Knape J (2016). *Decomposing trends in Swedish bird populations using generalized additive mixed models*. *Journal of Applied Ecology* **53**, 1852–1861. `https://doi.org/10.1111/1365-2664.12720`.

Lasgorceux F, Papaïx J, Bunz Y, Combrisson D, Opitz T (2026a). *Data, scripts and code for 'Space-time species distribution modeling with opportunistic presence-only data: a case study of passerines in a protected area' [Data set]*. Zenodo. `https://doi.org/10.5281/zenodo.19864262`.

Lasgorceux F, Papaïx J, Bunz Y, Combrisson D, Opitz T (2026b). *Outputs from fitted models across the cross-validation scenarios for 'Space-time species distribution modeling with opportunistic presence-only data: a case study of passerines in a protected area' [Data set]*. Zenodo. `https://doi.org/10.5281/zenodo.19487370`.

Lasgorceux F, Papaïx J, Bunz Y, Combrisson D, Opitz T (2026c). *Supplementary information for 'Space-time species distribution modeling with opportunistic presence- only data: a case study of passerines in a protected area' [Data set]*. Zenodo. `https://doi.org/10.5281/zenodo.19864179`.

Lehikoinen A, Brotons L, Calladine J, Campedelli T, Escandell V, Flousek J, Grueneberg C, Haas F, Harris S, Herrando S, Husby M, Jiguet F, Kålås JA, Lindström Å, Lorrillière R, Molina B, Pladevall C, Calvi G, Sattler T, Schmid H, et al. (2019). *Declining population trends of European mountain birds*. *Global Change Biology* **25**, 577–588. `https://doi.org/10.1111/gcb.14522`.

Lindgren F, Rue H, Lindström J (2011). *An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach*. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **73**, 423–498. `https://doi.org/10.1111/j.1467-9868.2011.00777.x`.

Link WA, Cam E, Nichols JD, Cooch EG (2002). *Of Bugs and Birds: Markov Chain Monte Carlo for Hierarchical Modeling in Wildlife Research*. *The Journal of Wildlife Management* **66**, 277–291. `https://doi.org/10.2307/3803160`.

McVicar TR, Körner C (2013). *On the use of elevation, altitude, and height in the ecological and climatological literature*. *Oecologia* **171**, 335–337. `https://doi.org/10.1007/s00442-012-2416-7`.

Miller DAW, Pacifici K, Sanderlin JS, Reich BJ (2019). *The recent past and promising future for data integration methods to estimate species' distributions*. *Methods in Ecology and Evolution* **10**, 22–37. `https://doi.org/10.1111/2041-210X.13110`.

Moreira GA, Menezes R, Wise L (2024). *Presence-Only for Marked Point Process Under Preferential Sampling*. *Journal of Agricultural, Biological and Environmental Statistics* **29**, 92–109. `https://doi.org/10.1007/s13253-023-00558-x`.

Murtagh F, Legendre P (2014). *Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?* *Journal of Classification* **31**, 274–295. `https://doi.org/10.1007/s00357-014-9161-z`.

Newton I (2004). *The recent declines of farmland bird populations in Britain: an appraisal of causal factors and conservation actions*. *Ibis* **146**, 579–600. `https://doi.org/10.1111/j.1474-919X.2004.00375.x`.

Noël F, Combrisson D, Geoffroy JJ, Nicolas MG (2023). *Synthèse des connaissances sur les Isopodes et Diplopodes terrestres du Parc national des Écrins (SE France)*. *Naturae 2023*, 151–170. `https://doi.org/10.5852/naturae2023a9`.

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`.

Pedersen EJ, Miller DL, Simpson GL, Ross N (2019). *Hierarchical generalized additive models in ecology: an introduction with mgcv*. *PeerJ* **7**, e6876. `https://doi.org/10.7717/peerj.6876`.

Phillips SJ, Anderson RP, Schapire RE (2006). *Maximum entropy modeling of species geographic distributions*. *Ecological Modelling* **190**, 231–259. `https://doi.org/10.1016/j.ecolmodel.2005.03.026`.

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009). *Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data*. *Ecological Applications* **19**, 181–197. `https://doi.org/10.1890/07-2153.1`.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.r-project.org`.

Ramsay JO, Silverman BW (2005). *Functional Data Analysis*. Springer Series in Statistics. New York, NY: Springer. `https://doi.org/10.1007/b98888`.

Ramsay J, Hooker G, Graves S (2009). *Functional Data Analysis with R and MATLAB*. New York, NY: Springer. `https://doi.org/10.1007/978-0-387-98185-7`.

Reif J, Storch D, Voříšek P, Šťastný K, Bejček V (2008a). *Bird-habitat associations predict population trends in central European forest and farmland birds*. *Biodiversity and Conservation* **17**, 3307–3319. `https://doi.org/10.1007/s10531-008-9430-4`.

Reif J, Voříšek P, Š?astný K, Bejček V, Petr J (2008b). *Agricultural intensification and farmland birds: new insights from a central European country*. *Ibis* **150**, 596–605. `https://doi.org/10.1111/j.1474-919X.2008.00829.x`.

Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI (2015). *Point process models for presence-only analysis*. *Methods in Ecology and Evolution* **6**, 366–379. `https://doi.org/10.1111/2041-210X.12352`.

Renner IW, Louvrier J, Gimenez O (2019). *Combining multiple data sources in species distribution models while accounting for spatial dependence and overfitting with combined penalized likelihood maximization*. *Methods in Ecology and Evolution* **10**, 2118–2128. `https://doi.org/10.1111/2041-210X.13297`.

Renner IW, Warton DI (2013). *Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology*. *Biometrics* **69**, 274–281. `https://doi.org/10.1111/j.1541-0420.2012.01824.x`.

Rigal S, Dakos V, Alonso H, Auniņš A, Benkő Z, Brotons L, Chodkiewicz T, Chylarecki P, Carli E, Moral JC, Domşa C, Escandell V, Fontaine B, Foppen R, Gregory R, Harris S, Herrando S, Husby M, Ieronymidou C, Jiguet F, et al. (2023). *Farmland practices are driving bird population decline across Europe*. *Proceedings of the National Academy of Sciences* **120**, e2216573120. `https://doi.org/10.1073/pnas.2216573120`.

Rue H, Martino S, Chopin N (2009). *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations*. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **71**, 319–392. `https://doi.org/10.1111/j.1467-9868.2008.00700.x`.

Sanderson FJ, Kucharz M, Jobda M, Donald PF (2013). *Impacts of agricultural intensification and abandonment on farmland birds in Poland following EU accession*. *Agriculture, Ecosystems & Environment* **168**, 16–24. `https://doi.org/10.1016/j.agee.2013.01.015`.

Schulze ED, Craven D, Durso AM, Reif J, Guderle M, Kroiher F, Hennig P, Weiserbs A, Schall P, Ammer C, Eisenhauer N (2019). *Positive association between forest management, environmental change, and forest bird abundance*. *Forest Ecosystems* **6**, 3. `https://doi.org/10.1186/s40663-019-0160-8`.

Silva JP, Correia R, Alonso H, Martins RC, D’Amico M, Delgado A, Sampaio H, Godinho C, Moreira F (2018). *EU protected area network did not prevent a country wide population decline in a threatened grassland bird*. *PeerJ* **6**, e4284. `https://doi.org/10.7717/peerj.4284`.

Snäll T, Kindvall O, Nilsson J, Pärt T (2011). *Evaluating citizen-based presence data for bird monitoring*. *Biological Conservation* **144**, 804–810. `https://doi.org/10.1016/j.biocon.2010.11.010`.

Soriano-Redondo A, Jones-Todd CM, Bearhop S, Hilton GM, Lock L, Stanbury A, Votier SC, Illian JB (2019). *Understanding species distribution in dynamic populations: a new approach using spatio-temporal point process models*. *Ecography* **42**, 1092–1102. https://doi.org/10.1111/ecog.03771.

Strebel N, Antoniazza S, Auchli N, Birrer S, Bühler R, Sattler T, Volet B, Wechsler S, Moosmann M (2024). *The State of Birds in Switzerland. Report 2024*. Sempach: Swiss Ornithological Institute. https://doi.org/10.13140/RG.2.2.36730.09927.

Telfer MG, Preston C, Rothery P (2002). *A general method for measuring relative change in range size from biological atlas data*. *Biological Conservation* **107**, 99–109. https://doi.org/10.1016/s0006-3207(02)00050-2.

Timmers R, Kuijk M, Verweij PA, Ghazoul J, Hautier Y, Laurance WF, Arriaga-Weiss SL, Askins RA, Battisti C, Berg Å, Daily GC, Estades CF, Frank B, Kurosawa R, Pojar RA, Woinarski JC, Soons MB (2022). *Conservation of birds in fragmented landscapes requires protected areas*. *Frontiers in Ecology and the Environment* **20**, 361–369. https://doi.org/10.1002/fee.2485.

Traba J, Morales MB (2019). *The decline of farmland birds in Spain is strongly associated to the loss of fallowland*. *Scientific Reports* **9**, 9473. https://doi.org/10.1038/s41598-019-45854-0.

Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J (2022). *Predictive performance of presence-only species distribution models: a benchmark study with reproducible code*. *Ecological Monographs* **92**, e01486. https://doi.org/10.1002/ecm.1486.

Van Strien AJ, Van Swaay CA, Termaat T (2013). *Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models*. *Journal of Applied Ecology* **50**, 1450–1458. https://doi.org/10.1111/1365-2664.12158.

Warton DI, Renner IW, Ramp D (2013). *Model-Based Control of Observer Bias for the Analysis of Presence-Only Data in Ecology*. *PLoS ONE* **8**, e79168. https://doi.org/10.1371/journal.pone.0079168.

Warton DI, Shepherd LC (2010). *Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology*. *The Annals of Applied Statistics* **4**, 1383–1402. https://doi.org/10.1214/10-AOAS331.

Wikle CK (2003). *Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes*. *Ecology* **84**, 1382–1394. https://doi.org/10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2.

Wretenberg J, Lindström Å, Svensson S, Thierfelder T, Pärt T (2006). *Population trends of farmland birds in Sweden and England: similar trends but different patterns of agricultural intensification*. *Journal of Applied Ecology* **43**, 1110–1120. https://doi.org/10.1111/j.1365-2664.2006.01216.x.

Yoccoz N (2026). *Analysing spatial and temporal changes of species distributions in mountains: a modelling approach for data collected by park rangers*. *Peer Community in Ecology*, 100750. https://doi.org/10.24072/pci.ecology.100750.

Zamora R, Barea-Azcón JM (2015). *Long-Term Changes in Mountain Passerine Bird Communities 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.