METHODS

Spatiotemporal modelling of abundance from multiple data sources in an integrated spatial distribution model

Nicolas Strebel | **Marc Kéry** | **Jérôme Guélat** | **Thomas Sattler**

Swiss Ornithological Institute, Sempach, Switzerland

Correspondence
Nicolas Strebel, Swiss Ornithological Institute, Sempach, Switzerland.
Email: nicolas.strebel@vogelwarte.ch

Handling Editor: Lisa Manne

Abstract

Aim: In biodiversity monitoring, observational data are often collected in multiple, disparate schemes with greatly varying degrees of standardization and possibly at different spatial and temporal scales. Technical advances also change the type of data over time. The resulting heterogeneous datasets are often deemed to be incompatible. Consequently, many available datasets may be ignored in practical analyses. Here, we propose a more efficient use of disparate biodiversity data to assess species distributions and population trends.

Location: Switzerland (Europe).

Taxon: Birds.

Methods: We developed an integrated, hierarchical species distribution model with a joint likelihood for all datasets using a shared state process (e.g. latent species abundance or occurrence), but distinct observation process for each dataset. We show how the abundance submodel of a binomial N-mixture model can fuse four different data types (count, detection/non-detection, presence-only and absence-only data) and enable improved inferences about spatiotemporal patterns in abundance. As case studies, we use data from multiple avian biodiversity monitoring schemes. In the first, the goal is estimating abundance-based species distribution maps. In the second, we infer trends in population abundance across time.

Results: Accuracy and precision of abundance estimates increased when combining data from different sources compared to using a single data source alone. This is particularly valuable when data from each single data source are too sparse for reliable parameter estimation.

Main conclusions: We show that exploiting the complementary nature of 'cheap', but abundant, citizen-science data and less abundant, but more information-rich, data from structured monitoring programmes might be ideal to estimate distribution and population trends more accurately, especially for rare species. Joint likelihoods allow to include a wide variety of different datasets to (1) combine all the available information and to (2) mitigate weaknesses of one by the strength of another.

1 | INTRODUCTION

Efficient assessment of spatial and temporal patterns in species distributions, for example, to draw species distribution maps and to estimate population trends, has been a challenging but essential endeavour for ecologists and wildlife managers for a long time (Barker & Sauer, 1992; Guisan et al., 2006; Inger et al., 2015). Species distribution or abundance models that specify population trends help understanding basic ecological questions (Guisan & Thuiller, 2005; Hirzel et al., 2006; Sattler et al., 2007). Reliable distribution and trend estimates are essential for a smart resource allocation in conservation (Rodrigues et al., 2006) and species management (Sutherland et al., 2004). Traditionally, temporal or spatial patterns in species distributions have been assessed using data from a single focused monitoring scheme (e.g. Fewster et al., 2000; Kéry et al., 2005; Sauer & Link, 2011). However, the scope of interest often goes beyond the temporal or spatial limits of an established monitoring program. With the rise of mass-participation citizen-science programmes, data produced by them are increasingly being used to answer questions of scientific or management interest (Cooper et al., 2007; Dickinson et al., 2012). However, analysis of these opportunistic data, though typically very copious, often suffers from severe challenges due to variable degrees spatial sampling and detection biases (Johnston et al., 2019). Thus, combining data from different monitoring schemes seems like an obvious proposition (Freeman et al., 2007; Isaac et al., 2020; Pagel et al., 2014). Making use of multiple data sources in integrated models can improve estimator precision (Besbeas et al., 2002; Pacifici et al., 2017; Fletcher Jr. et al., 2019; Miller et al., 2019; Pacifici et al., 2019; Kéry & Royle, 2021 [Chapter 10]; Schaub & Kéry, 2022). Including datasets from designed monitoring schemes can reduce potential biases from opportunistic data when modelled together (Dorazio, 2014). In addition, data from regions or periods that are poorly covered by any one scheme may be complemented with data from another (Lebreton et al., 1995; Renner et al., 2019).

Since the 1990s, ecology has seen substantial methodological developments for combining data from different sources to improve estimation of demographic parameters (Lebreton et al., 1995) and population trends (Besbeas et al., 2002). Freeman et al. (2007) combined data from two simultaneous bird monitoring schemes within one model to estimate population changes. Pagel et al. (2014) developed a framework to estimate population trends over time by combining counts and detection/non-detection data. This idea was picked up again by Zipkin et al. (2017), who extended the framework to a dynamic model including extinction and colonization using concepts from the models of MacKenzie et al. (2002), Dail and Madsen (2011) and Rossman et al. (2016). Since then, several

integrated species distribution models that enable the fusion of data from different sources using a common state model and different observation models via a joint likelihood were described in the literature. Koshkina et al. (2017) combined detection/non-detection and presence-background data. Pacifici et al. (2017) described a spatially explicit framework to share information between higher quality and lower quality detection/non-detection data. Also, Pacifici et al. (2019) presented a method to resolve the issues caused by spatial misalignment in multiple data sources that are mutually used for species distribution modelling. Fletcher Jr. et al. (2019) proposed combining data of different quality by making use of weighted joint likelihoods. Miller et al. (2019) gave an overview on recent advances in data integration for species distribution models, putting them into the context of spatial point processes. Renner et al. (2019) proposed a combined penalized likelihood approach to combine detection/non-detection and presence-only data, accounting for spatial dependence in the data.

Here, we develop an integrated species distribution model that allows combining four different data types within a single modelling framework to estimate spatial or temporal patterns in abundance. The four data types are as follows: (1) counts resulting from (repeated) observations, as typically conducted within structured monitoring schemes; (2) detection/non-detections (also misleadingly called "presence/absence data"), for example, from checklists; (3) information on the presence of a species at a site, often originating from older data sources (e.g. prior to digital transmission of observation data); and (4) information on the ascertained absence of a species at a site. To do so, we extend the formal framework first described by Pagel et al. (2014) by an inclusion of the latter two data sources (3 and 4) that can both be considered as site-specific syntheses that lack information on the observation process. To our knowledge, this is the first attempt to combine those four data types within one single modelling framework. Our model explicitly accounts for imperfect detection by estimating detection probability in count and detection/non-detection data by combining the occupancy model by MacKenzie et al. (2002) and Tyre et al. (2003) with the binomial N-mixture model of Royle (2004).

We illustrate the novel model with two applications. First, we model spatial patterns in the abundance of two rare bird species using data from different sources collected for the Swiss Breeding Bird Atlas 2013–2016 (Knaus et al., 2018). In our second example, we assess population trends of the same species based on data collected in different monitoring schemes during the breeding seasons 1999–2018. In both cases, we compare the inferences under our new model with those from traditional methods that do not allow combining detection/non-detection and count data.

2 | MATERIALS AND METHODS

After introducing the data, we describe our general modelling framework to incorporate four different data types via a joint likelihood to model patterns in abundance over space and time. Subsequently, we illustrate our new model with two applications, focusing first on the spatial and second on the temporal dimension.

2.1 | Overview of the data

2.1.1 | Data type 1: Replicated counts

Typically, count data are used to model patterns of abundance in space or changes in abundance over time. For our subsequent applications, we considered either data collected within the Swiss Breeding Bird Atlas (Application 1; Knaus et al., 2018) or from the similarly designed Swiss Common Breeding Bird Monitoring called MHB (Application 2; Kéry & Schmid, 2006; Schmid et al., 2004). For both schemes, 1 km^2 grid cells based on the Swiss coordinate reference system (CH1903/LV03, EPSG: 21781) are surveyed two to three times annually within one breeding season (15 April–15 July) by the same observer on a cell-specific, constant transect using territory- (or spot-) mapping techniques (Bibby et al., 2000). The locations of all visually or aurally detected individuals (mostly singing males) are mapped during each survey and subsequently grouped to putative territories (Kéry & Schmid, 2006; Schmid et al., 2004). As input data for our model, we used the observed number of territories per survey, grid cell and year.

2.1.2 | Data type 2: Replicated detection/non-detection data from checklists

On the data entry platform *ornitho.ch* and the corresponding mobile-phone app *NaturaList*, users can enter their opportunistic bird observations. Users registered as volunteers of the Swiss Ornithological Institute are asked to either systematically report all observations of a defined set of rare species or to submit complete observation lists where every species detected is recorded. We used checklist data of this type for both spatial and temporal applications. This type of observational data allows deducing non-detection data (i.e. zeros) for either the set of rare species or for all species when not recorded (Kéry et al., 2010). Both complete lists and observations of the fixed set are assigned to 1 km^2 grid cells based on the Swiss coordinate reference system. To avoid inclusion in the analysis of migrating individuals, only observations made within a defined species-specific breeding period were included in the analysis (Knaus et al., 2018).

2.1.3 | Data type 3: Annual presence-only data

In a separate monitoring programme focused on wetland species, a fixed set of sites is annually surveyed for breeding birds. Effort is not standardized and annually varies between 3 and 12 surveys per site, typically conducted by highly skilled volunteers (Müller, 2015). The resulting data contain the species that were present during the breeding season at a certain site, but these data are not assigned to a specific date. This data type often originates in older data (prior to digital transmission of observation data) or from reports that do not specify individual observations (e.g. environmental impact analyses or other reports for administrative purposes, travel reports).

2.1.4 | Data type 4: Absence-only data

Switzerland has a steep altitudinal gradient (193–4634 m.a.s.l.). High-elevation areas are difficult to access and consequently less surveyed by volunteer ornithologists. Additionally, within the range of a species, density usually is lower towards its higher end of elevation. Therefore, data of type 2 and thus information on species absences from high-elevation areas are often sparse. However, most species are limited to some well-defined, species-specific altitudinal range. In this way, it is possible to define conservative species-specific altitudinal thresholds above which a species is essentially close to certain to be absent as a breeding bird, based on the sparse data that is available in high altitudes, in combination with expert knowledge. We denote all grid cells above such a threshold as absence and call them 'ascertained absence' or 'absence-only data'. Such absence data greatly ease the modelling at the edge of the elevational gradient.

2.2 | Model

Our model framework combines replicated count data, detection/non-detection data, presence-only and ascertained absences on the basis of a binomial N-mixture (Royle, 2004) and a site-occupancy model (MacKenzie et al., 2002; Tyre et al., 2003) that are specified in combination. Inference on abundance based on replicated count data is straightforward via the binomial N-mixture part of the model, while detection/non-detection data enter via the site-occupancy part. Presence-only data and absence-only data do not contain any information on the observation effort and are assumed to directly observe the true presence or absence state of a site. Therefore, we added them directly in the state part of the hierarchical model by fixing the latent occupancy state of a site at 1 (for a presence) or 0 (for an absence).

In our hierarchical model, we specified the abundance *N* and the occurrence (=presence/absence) *z* of a species at site *i* as a latent (i.e. partially or completely unobserved) variable in the state part of the model. The latent abundance *N* at site *i* is modelled as the outcome of a non-negative, discrete distribution. As usual, we will use the Poisson:

$$N_i \sim \text{Poisson}(\lambda_i). \tag{1}$$

Here, the expected abundance λ_i, indexed by site *i*, allows the inclusion of site-level covariates or random effects (e.g. to accommodate spatial dependence) via a log link:

$$\log(\lambda_i) = \alpha + \sum_{k=1}^{I} \beta_k * x_{ki}, \quad (2)$$

where α is an intercept, β stands for the site-specific covariate values and x for the parameter estimates and k indexes a number of covariates. Two different observation models link the replicated counts and detection/non-detection data to the latent abundance state at a site.

Observation model for replicated counts

The observed count $C_{i,j}$ at site i during survey j is described as the outcome of a binomial distribution with trial size N_i and success probability $\theta_{i,j}$ which is the probability that an individual present within site i is detected during survey j.

$$C_{i,j} \sim \text{Binomial}(N_i, \theta_{i,j}). \quad (3)$$

Detection probability is modelled using a logit link function, in the illustration here with an intercept α_θ and with covariates varying either by space ($x1$) or within season ($x2$).

$$\text{logit}(\theta_{i,j}) = \alpha_\theta + \beta_1 * x1_i + \beta_2 * x2_j. \quad (4)$$

Observation model for replicated detection/non-detection data

We modelled the number of detections y of a species per site i and survey j as the outcome of a binomial distribution with trial size $S_{i,j}$ and species-level detection probability $p_{i,j}$.

$$y_{i,j} \sim \text{Binomial}(S_{i,j}, p_{i,j}). \quad (5)$$

Here, $S_{i,j}$ is the number of surveys conducted at site i during survey occasion j. The success probability $p_{i,j}$ here is the detection probability of the species, that is, of all individuals combined that are present at site i during survey j, or, in other words, the probability to detect at least one individual during survey j at site i.

We linked these data to the underlying abundance at site i by expressing the per-species detection probability p as a function of the per-individual detection probability θ from above and the population size N_i of the specific site according to the model of Royle and Nichols (2003):

$$p_{i,j} = 1 - (1 - \theta_{i,j})^{N_i}. \quad (6)$$

Thus, the per-species detection probability at site i during survey j, $p_{i,j}$, is 1 minus the probability to miss all N_i individuals present, which for each individual happens independently with probability $1 - \theta_{i,j}$.

Annual presence-only data and annual absence-only data

These additional two data types were assumed to perfectly observe the true presence or absence state of a site during a given year. Note that for absences, this contrasts with the traditional assumption of site-occupancy and binomial N-mixture models that absences are never perfectly observed. These data are introduced into the model without a formal observation process by simply fixing the occupancy state for site i, z_i, at a value of 1 for sites where the species was recorded as a breeding bird based on annual presence-only data, and at a value of 0 for sites where the species was certain to be absent as a breeder (see below). To allow for the provision of fixed values for the latent presence or absence state at some sites in our implementation of the model in JAGS (Plummer, 2017), we formally specified the relationship between the latent occurrence (i.e. presence/absence) state of a site and the abundance at the site as stochastic (although it is in fact deterministic such that $z_i \leftarrow N_i > 0$):

$$z_i \sim \text{Bernoulli}(I(N_i > 0)). \quad (7)$$

The indicator function $I(.)$ evaluates to 1 when its argument is true and to 0 otherwise. This allowed us to fix z for some sites to 0 or 1.

2.3 | Model fitting

For the applications presented below, we processed data in R (R Core Team, 2018) and fitted our models in JAGS (Plummer, 2017) using the R-package 'jagsUI' (Kellner, 2019). Starting values must be selected carefully when launching the analysis in JAGS to avoid conflicts between data, model and the state of the chains at initialization. We chose starting values for the latent abundance state N in the BUGS model as follows: If replicated counts (data type 1) were available, we used the maximum number of territories found within one survey in the corresponding site. Starting values were set to 1 for sites where a species was reported on opportunistic lists only (data type 2), and to NA for sites covered by data types 3 and 4. For all sites where no data were available at all, starting values for the latent abundance state N were set to 0.

2.4 | Application 1

2.4.1 | Spatial modelling of abundance

Here, we apply our model for species distribution mapping, utilizing data collected for the Swiss breeding bird atlas (Knaus et al., 2018). We analysed data from two species which both have a restricted distribution as breeding birds in Switzerland, the

Common Nightingale *Luscinia megarhynchos* in the lowlands and the Rock Bunting *Emberiza cia* in the (Southern) Alps. During the breeding seasons of 2013–2016, replicated counts (type 1) were collected in 2262 grid cells and opportunistic lists (type 2, citizen-science data) in approximately 25,000 grid cells (Table 1). The spatial coverage of the latter data type is much higher. However, they do not contain information on abundance (just detection/non-detection) and suffer from a spatial sampling bias in favour of lowland sites. Annual presence-only data (type 3) were available from 92 wetland sites, complementing the spatial coverage from data of types 1 and 2 (especially for the Common Nightingale). For our absence-only data (data type 4), we fixed at 0 the latent occurrence state z_i of all grid cells located above the species-specific altitudinal threshold.

2.4.2 | Integrated species distribution model for abundance

To model abundance per grid cell, we used a set of environmental covariates (see Table S1) that were found to explain the species occurrence in the recent Swiss Breeding Bird Atlas (Knaus et al., 2018). We extended the log-linear model for abundance in Equation (2) by adding a two-dimensional penalized spline function for the *x*- and *y*-coordinates to account for spatial autocorrelation (Guélat & Kéry, 2018; Kammann & Wand, 2003). We modelled detection probability as.

$$\text{logit}\left(\theta_{i,j}\right) = \alpha_\theta + \beta_1 * \text{date}_{i,j} + \beta_2 * \text{date}_{i,j}^2, \tag{8}$$

with α being estimated separately for data of type 1, for data of type 2 originating from complete lists and for data of type 2 where observers only recorded the list of rarer species. We predicted the estimated abundance per grid cell for entire Switzerland ($n = 41,844$).

We compared our model with three more traditional methods (Table 2), namely a Poisson generalized linear model (henceforth: GLM), a binomial N-mixture model (henceforth: N-mix model; Royle, 2004) and a site-occupancy model (MacKenzie et al., 2002; Tyre et al., 2003) for which abundance is neglected. We label our new model the 'integrated SDM' in these comparisons. In all these model types, we retained identical the model structure in terms of covariates and spline-based modelling of spatial autocorrelation. For the GLM, the modelled response was the total number of territories per grid cell based on all surveys within one season combined (thus an aggregation of data type 1, ignoring detection probability); data types 2, 3 and 4 were ignored for the GLM. For the N-mix model, we did include detection probability (i.e. used the number of territories detected per grid cell and survey of data type 1), but again ignored data from types 2, 3 and 4. For the site-occupancy model, we used all data that were also considered for the integrated SDM, but quantized the number of territories detected per grid cell and replicate (data type 1) to become detection/non-detection data, thus incorporating all available data, but just modelling occupancy probability instead of abundance as the state variable in the hierarchical model (detection probability included). To fit the models, we ran three Markov chains for 50,000 (site-occupancy model) or 100,000 (all other models) iterations and discarded the first half as a burn-in. We monitored convergence (Rhat values < 1.1 according to Brooks & Gelman, 1998) of the parameter estimates for detectability and abundance (α in Equation 2). For the Rock Bunting, Rhat values exceed 1.1 for one parameter ('wetlands') in the site-occupancy model, for two parameters (wetlands and arable land) in the integrated SDM and for the linear and quadratic term of the covariate 'buildings' in the N-mix model and the GLM. For the Nightingale, in all four analyses, Rhat values exceeded 1.1 only for the parameter estimate for 'altitude' (a covariate strongly correlated with other covariates). Code and data to replicate the analysis can be found on https://doi.org/10.5281/zenodo.5840377.

TABLE 1 Sample sizes per species and data type, in terms of the number of grid cells with data

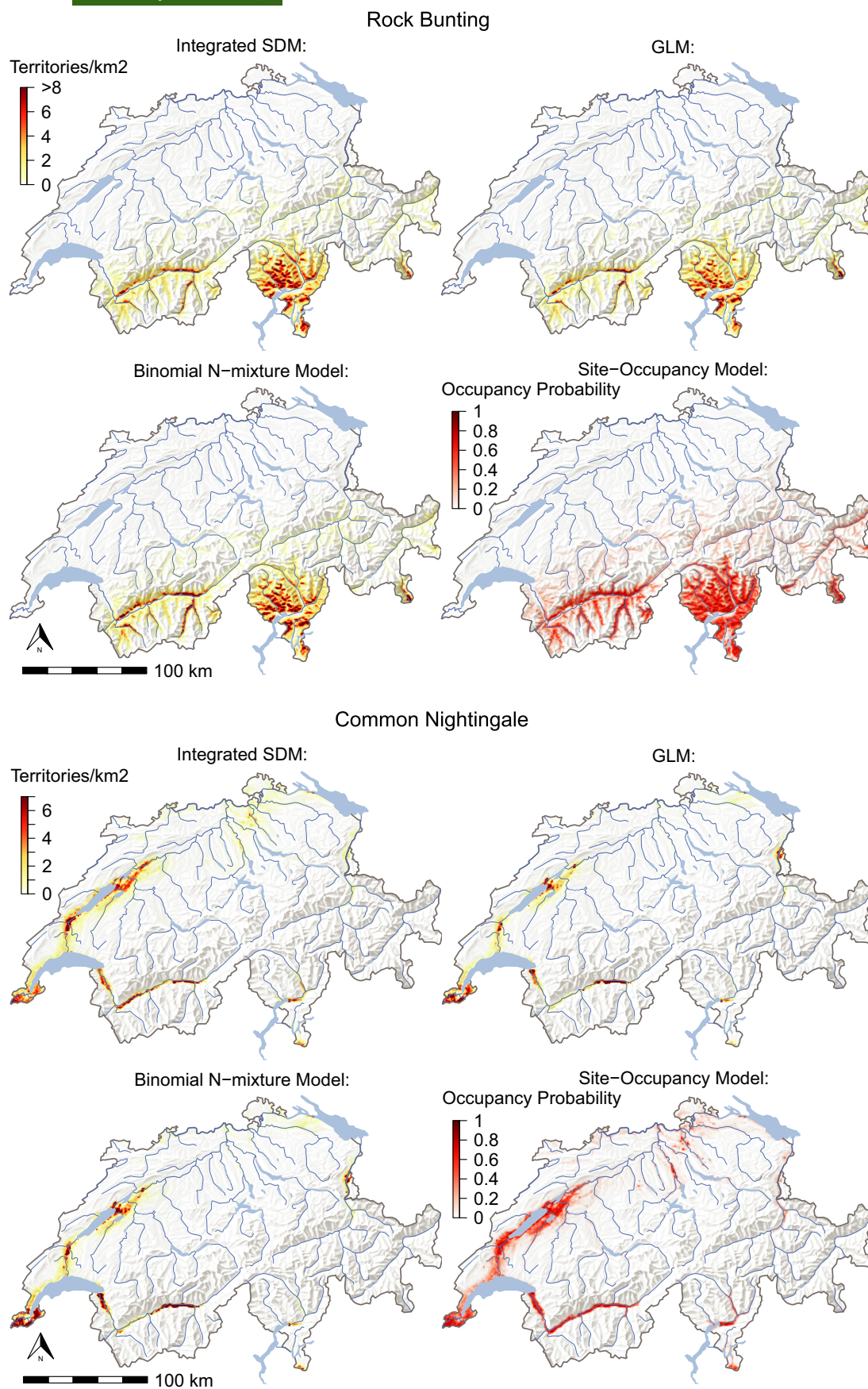|	Replicated counts (data type 1)	Detection/non-detection data (data type 2)	Presence-only data (data type 3)	Absence-only data (data type 4)
Rock Bunting	222/2262	853/24,655[a]	0	1653 (>2800 m a.s.l.)
Common Nightingale	94/2262	990/22,534	7	7787 (>1800 m a.s.l.)

Note: Data types 1 and 2: number of grid cells where the species was found/the total number of grid cells visited in 2013–2016. Presence-only data: number of sites where such data were available, but where no observation of data types 1 and 2 was available. Absence-only data: number of grid cells where the presence of the species was set to zero (corresponding altitudinal threshold in brackets).

[a]Numbers of visited cells differ, as breeding season is defined differently for the two species.

TABLE 2 Overview on the four modelling approaches in Application 1. For all models, we included the same set of environmental variables (Data S1) and a two-dimensional penalized spline function to consider spatial autocorrelation

	Integrated SDM	GLM	N-mix model	Site-occ model
Data included	Types 1–4	Aggregation of type 1	Type 1	Types 1–4
Modelled quantity	Abundance	Abundance	Abundance	Occupancy (P/A)
Detection probability	Included	Ignored	Included	Included

Rock Bunting

Integrated SDM:

Territories/km2

>8
6
4
2
0

GLM:

Binomial N−mixture Model:

Site−Occupancy Model:

Occupancy Probability

1
0.8
0.6
0.4
0.2
0

N

100 km

Common Nightingale

Integrated SDM:

Territories/km2

6
4
2
0

GLM:

Binomial N−mixture Model:

Site−Occupancy Model:

Occupancy Probability

1
0.8
0.6
0.4
0.2
0

N

100 km

FIGURE 1 Species distribution maps for Rock Bunting and Common Nightingale, obtained by the integrated species distribution model and three simpler models that do not use the full information utilized by the former. We show posterior medians of predictions of the expected number of territories (top and bottom left for the plots in both species), and occupancy probability (bottom right for both species). Examples for additional species are found in Data S1

To compare out-of-sample prediction performance of the integrated SDM with that of an N-mix model (that uses data of type 1 only), we defined a hold-out sample by randomly selecting 10% of the sites with data type 1 and turned all count data from these sites to NAs (861 breeding season counts from 226 sites for the Rock Bunting, 618 breeding season counts from 226 sites for the Nightingale). Then, we fitted the models using the reduced dataset, predicted the out-of-sample count data and calculated the Pearson correlation coefficient and the root mean squared prediction error (Hooten & Hobbs, 2015) of the mean predictions and the counts. In addition, we approximated the log predictive density for the out-of-sample counts (Gelman et al., 2014; Hooten & Hobbs, 2015) using Monte Carlo integration (Hooten & Hefley, 2019). To do so, for each posterior draw retained from the analysis in JAGS ($n = 300$), we simulated 1,000,000 data points for every out-of-sample count using the posterior sample estimates of λ and θ for the corresponding site and date from the JAGS output. For each out-of-sample count, we then calculated the proportion of simulated values that were equal to the corresponding out-of-sample count. Note that the counts are integers and so this proportion is an estimate of the probability of the observed datum given the parameter estimates under each model. Then we multiplied the resulting values across all hold-out counts. This was repeated for all 300 posterior draws we retained from the JAGS analysis. To get an estimate for the log predictive density, we then averaged the resulting 300 values and took the natural log of this average. If at least one count did not match any of the 1,000,000 simulated data points (because the observed count was outside of the range of the predicted values), then the product of the resulting values for the corresponding posterior sample became zero and the logarithm of this was negative infinity. In the binomial-N-mixture model fitted to the Nightingale data, this was the case for all the 300 posterior draws.

The modelled quantity in both the integrated SDM and the N-mix model is latent abundance. In contrast, in the GLM, the modelled quantity is observed abundance, while in the site-occupancy model it is latent occurrence. Therefore, we could not include the latter two in the comparison.

2.4.3 | Results in Application 1

We estimated the number of territories for all 1 km^2 grid cells in Switzerland. As a point estimate, we use the posterior median (Figure 1). As in Knaus et al. (2018), estimates exceeding a species-specific maximum abundance were truncated for plotting. We defined this threshold to be the 99% quantile for all predictions ≥0.05 territories, rounded to the nearest integer.

For the Rock Bunting, the spatial patterns found in abundance are quite similar for the integrated SDM, the N-mix model and the GLM, whereas the estimates of the latter tended to be slightly lower as they do not correct for imperfect detection. Also, for the Nightingale, the
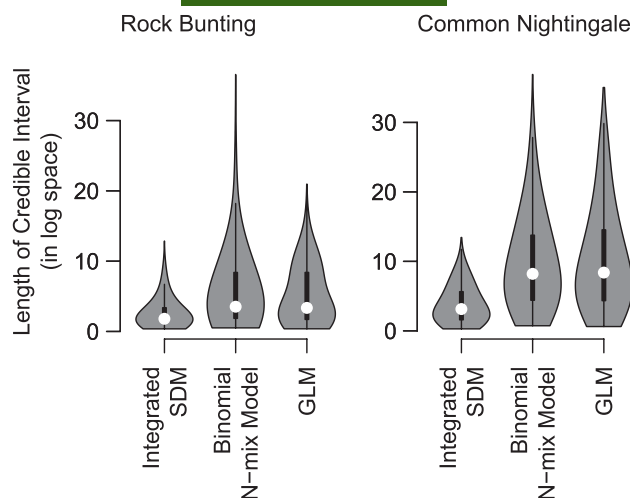
FIGURE 2 Length of the 95% credible interval of the estimated abundance per grid cell over all grid cells in Switzerland, assembled as violin plots (Adler & Kelly, 2021). Each violin plot summarizes 41,844 values. No results are shown for the site-occupancy model, since the estimates are in different mathematical spaces (here log, vs. logit in the site-occupancy model) and therefore cannot be compared reasonably

species distributions estimated by the integrated SDM, the N-mix model and the GLM look rather similar. However, the map produced by the GLM and the N-mix model miss some pockets where the species is estimated to occur according to the integrated SDM and the site-occupancy model, for example, in Northern Switzerland. This is not astonishing, as the latter two utilize data with much greater spatial coverage. The spatial pattern found by the site-occupancy model and the integrated SDM was very similar for both species. However, patches of high abundance cannot be identified, and the lack of inclusion of abundance can even mask patterns, like the strong preference of Rock Bunting for southern exposed slopes than can be seen clearly on the other maps. Thus, the site-occupancy model cannot provide nuanced predictions for species with local abundances far greater than one.

Abundance per grid cell was estimated considerably more precisely by the integrated SDM than by the N-mix model and the GLM (Figure 2).

For the considered model selection metrics, the integrated SDM outperformed the N-mix model in both species (Table 3).

2.5 | Application 2

2.5.1 | Modelling of abundance over time

For the 20 years of 1999–2018, we modelled annually varying abundance of Rock Bunting and Common Nightingale based on data from several sources. Both species may reach densities up to several territories per 1 km^2. We used territory mapping data (data type 1 from the Swiss common breeding bird monitoring (Schmid

TABLE 3 Out-of-sample model comparison results for the N-mix model and the integrated SDM using three model selection criteria

	Correlation coefficient		RMSPE		Log predictive density	
	N-mix model	Integrated SDM	N-mix model	Integrated SDM	N-mix model	Integrated SDM
Common Nightingale	0.47	**0.53**	4.36	**0.73**	−Inf[a]	**−121.82**
Rock Bunting	0.65	**0.74**	0.54	**0.48**	−221.57	**−195.93**

Note: The 'better' result per metric and species is indicated in bold. Abbreviation: SDM, species distribution model RMSPE: root mean squared prediction error.

[a]The approximation for log predictive density of the N-mix model for Nightingale resulted in −Inf, as the prediction differed substantially from the count in at least one case.

TABLE 4 Sample sizes in terms of the number of grid cells with observations of the species per year

	Replicated counts (data type 1)	Replicated detection/non-detection data (data type 2)	Presence-only data (data type 3)
Common Nightingale	Min: 1, max: 12, tot: 28	Min: 141, max: 576, tot: 1608	Min: 6, max: 15, tot: 26
Rock Bunting	Min: 7, max: 44, tot: 97	Min: 104, max: 390, tot: 639	Min: 0, max: 0, tot: 0

Note: Indicated is the minimum and maximum of grid cells with observations per year, the last figure (tot) indicates the total number of grid cells where the species was ever detected during the breeding seasons 1999–2018. For data of type 3, we only utilized sites where no detection data were available from data type 1 or 2 in the same year.

et al., 2004) and the Swiss Biodiversity Monitoring (Hintermann et al., 2000). In the former, 267 grid cells of 1 km^2 are annually surveyed since 1999. In the latter, a different set of 267 grid cells of 1 km^2 are surveyed on a rotational basis once every 5 years, starting in 2001 (NA in the 4 years between). As the Nightingale mainly occurs in riverine habitats and in rare occasions in hedgerows and forest borders (mainly W Switzerland), it is only rarely included (1–12 areas/year) in the regular grid of the Swiss common breeding bird monitoring and the Swiss Biodiversity Monitoring (Table 4). The Rock Bunting is recorded somewhat more by these regular survey schemes. In addition, we used checklist data (citizen science data) collected since 1999 (data type 2) which is far more common than data type 1. Data type 3 from the wetlands monitoring was obtained on an annual basis and was used as in Application 1. Here we did not utilize absence-only data (type 4), since we only included grid cells where at least one breeding season record was made during 1999–2018.

2.5.2 | Model extension for the temporal dimension

We extended the previous model to estimate temporal trends in abundance. The latent abundance (i.e. number of territories), $N_{i,t}$, in grid cell i and year t is treated as a Poisson random variable with an expected value $\lambda_{i,t}$ that we expressed (on the log link scale) as the sum of a year effect α and a site effect ε as follows:

$$\log(\lambda_{i,t}) = \alpha_t + \varepsilon_i, \text{ with } \varepsilon_i \sim \text{Normal}\left(0 \ \sigma_{\text{site}}^2\right). \tag{9}$$

As before, we had two different observation models that linked the observed data to the latent abundance states for each site/year combination. For these, we only needed to adapt the base model

(described above) by adding a temporal dimension. The observed number of territories $C_{i,j,t}$ at grid cell i during survey j and in year t (data type 1) was treated as a binomial random variable with index $N_{i,t}$ and success probability $\theta_{i,t}$, which again is the per-individual detection probability.

$$C_{i,j,t} \sim \text{Binomial}\left(N_{i,t}, \ \theta_{i,j,t}\right). \tag{10}$$

As for Application 1, we modelled detection probability again as a linear and quadratic function of survey date, but in addition added random year effects to allow for systematic differences among the years.

$$\text{logit}\left(\theta_{i,j,t}\right) = \alpha_t + \beta_1 * \text{date}_{i,j,t} + \beta_2 * \text{date}_{i,j,t}^2, \text{with } \alpha_t \sim \text{Normal}\left(\mu_\alpha \ \sigma_{\text{year}}^2\right), \tag{11}$$

We modelled the number of records of a species per grid cell, year and survey (y, data type 2) as the outcome of a binomial distribution with trial size $S_{i,j,t}$ and detection probability $p_{i,j,t}$; which is again the probability to detect at least one of the territories, that is, has a species-level interpretation.

$$y_{i,j,t} \sim \text{Binomial}\left(S_{i,j,t}, \ p_{i,j,t}\right) \tag{12}$$

As before, we expressed the relationship between the detection probabilities per grid cell and per individual as a deterministic function of local abundance according to the Royle and Nichols (2003) model, that is, exactly as in Equation (6), but now with an additional index t for year.

$$p_{i,j,t} = 1 - \left(1 - \theta_{i,j,t}\right)^{N_{i,t}}. \tag{13}$$

Annual presence-only data (type 3) were included as in the base model (Equation 7) by fixing the occupancy state for grid cell i in year t to $z_{i,t} = 1$ if a species was listed as breeding in the corresponding grid cell and year.

By adding the annual abundance estimates over all grid cells and dividing these values by the one of 2018, we obtained an estimated index of annual population size relative to the estimate for 2018, which is this treated as a baseline. As in Application 1, we compared the results from the integrated SDM with those obtained from a N-mix model, a site-occupancy model, and a Poisson GLM (detection probability ignored) as implemented in program TRIM (Table 5). This latter is widely used to estimate national trends from annual count data in Europe (Pannekoek & van Strien, 1991). Neither of which used the full information utilized by the integrated SDM. For the N-mix model, we used the number of territories detected per grid cell and survey (data type 1). For the site-occupancy model, we again used all data that were also used in the integrated SDM but quantized the number of territories detected per grid cell and survey to become detection/non-detection data (detection probability included).

We again used Bayesian inference and fitted the models in JAGS using vague priors for all parameters, running three chains over a maximum of 50, 000 iterations and discarding the first half of the iterations as a burn-in, which was sufficient to achieve convergence (Rhat < 1.1; Brooks & Gelman, 1998) for all estimated parameters in the integrated model and the site-occupancy model. For the N-mix model, convergence was achieved for all annual estimates for Rock Bunting and for most annual estimates for the Nightingale. For the Poisson GLM fitted with TRIM in R-package 'rtrim' (Bogaart et al., 2018), we treated as the response the observed total number of territories per year and grid cell for the two to three annual surveys combined. Code and data to replicate our analysis can be found on https://doi.org/10.5281/zenodo.5840377.

2.5.3 | Results in Application 2

The different methods yielded an estimate of the relative annual population size compared to the reference year 2018 (Figure 3). For the Nightingale, the integrated model estimated a substantial increase in the breeding population particularly over the first 15 years. The increase estimated by the site-occupancy model was less in the

Nightingale. In the results given by TRIM (Poisson GLM) and the N-mix model, no trend is visible in the estimates for Nightingale. In both models, the width of the confidence intervals and the annual fluctuations imply that the available data are too sparse for accurate and precise trend estimation. For the Rock Bunting, the estimated trend in breeding population estimated by the different methods that do consider information on abundance in the data is similar, indicating an increase in breeding population size. The 95% confidence intervals of the TRIM estimates are smaller than the analogous 95% credible intervals obtained from the integrated SDM and the N-mix model. In contrast to the abundance-based models, the site-occupancy model estimates a rather stable breeding population over the study period for the Rock Bunting.

3 | DISCUSSION

In biodiversity monitoring and ecology, multiple data sources are often available when examining the distribution of a species over a large geographical area or its population trend over or a period of multiple years. We developed an integrated species distribution model (SDM) that jointly exploits the information in multiple, disparate datasets. With our new model, count data and detection/non-detection data are jointly used to estimate spatial patterns in abundance or population trends. The information content of the sparser, but more valuable, count data (directly informing abundance) is fully exploited, which is not the case in, for example, site-occupancy models that only allow analysing detection/non-detection data. The 'cheaper', but typically more abundant, detection/non-detection data can increase spatial coverage of the data considerably. In addition, detection records that are not assigned to a specific date, but have an annual temporal resolution only, can be added into the estimation. Finally, it is also possible to include information on ascertained absences of a species, for example, for sites with covariate values above or below a certain threshold that makes presence of a species essentially impossible. Such absence data add yet more information to the model and can lead to more accurate parameter estimates and predictions. This is especially valuable for sites that are rarely visited, or, more generally, if the parameter space over which we want to predict extends beyond the extent in the data, but we can be certain that a species is absent at one end of that range, as is typically the case for the broad elevation gradient of

TABLE 5 Overview on the four modelling approaches in Application 2

	Integrated SDM	TRIM	N-mix model	Site-occupancy model
Data included	Types 1–3	Aggregation of type 1	Type 1	Types 1–3
Modelled quantity	Abundance	Abundance	Abundance	Occupancy (P/A)
Detection probability	Included	Ignored	Included	Included
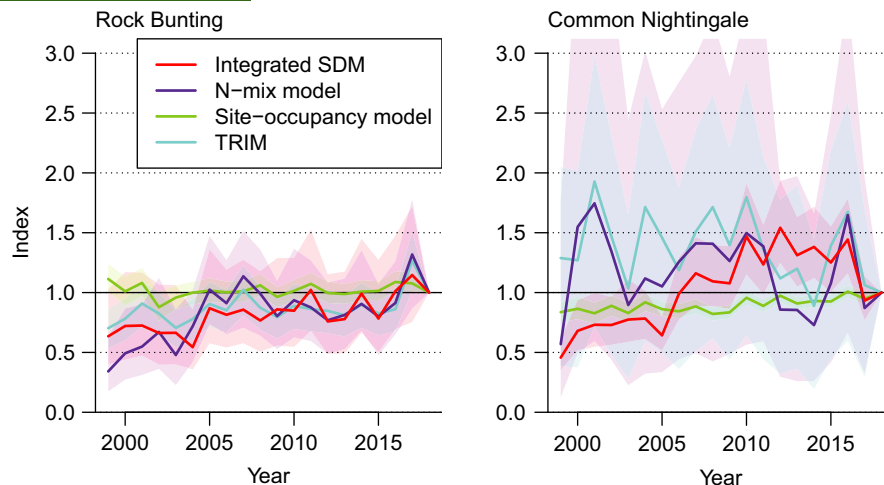
FIGURE 3 Point estimates and 95% uncertainty intervals of the annual breeding population of Common Nightingale and Rock Bunting in Switzerland, indexed such that the year 2018 is assigned a value of 1 and thus treated as a reference. We took posterior means as a point estimate in the first three models, which we fitted with Bayesian methods, and the mean estimate based on the imputed time totals for the last. Likewise, uncertainty is described by a 95% credible interval for the Bayesian analyses, and by a 95% confidence interval from TRIM. Also note the truncated *y*-axis for enhanced readability

species in Switzerland. Another example might be the inferred absence in forest-only sites for an obligate open-country nester such as the Skylark (*Alauda arvensis*). Finally, the integration of presence-only data allows making use of disparate monitoring programmes where information on the observation process is lacking. Regarding trend estimation, integration of presence-only data comes with a certain risk if the observation effort increases over time. However, we assume that this is less problematic as long as there is no trend in observation effort (e.g. in a monitoring programme with a fixed set of annually surveyed sites, as in our case of the wetland monitoring scheme). For species distribution estimation, the inclusion of presence-only leads to biased results when some sites are more likely to be visited than others ('observer bias'; Warton et al., 2013). In our analyses, the proportion of presence-only data among all data is low, but as soon as presence-only data make up a more substantial part of all data, we recommend considering a weighting procedure like the one proposed by Fletcher Jr. et al. (2019), or the methods proposed by Dorazio (2014) or Fithian et al. (2015) for data integration. Compared to the scenarios described by Dorazio (2014), Fithian et al. (2015) or Koshkina et al. (2017) that focus on a sophisticated way of combining (low quality) presence-only data with (high quality) data from more structured surveys, our most important 'low-quality' data source consists of detection/non-detections. In contrast to detection-only, detection/non-detection data come with zeros, which reduces a potential influence of site selection bias on covariate estimates (Guillera-Arroita et al., 2015; Lahoz-Monfort et al., 2014). We assume that, in our case, the method described here is adequate to exploit the entire information content of this data, without inducing a site-selection-caused bias in abundance estimation. The results of the model selection conducted in the spatial application are in

agreement with Zipkin et al. (2017) who showed that model estimates improved when detection/non-detection were added and jointly analysed with count data, compared to analysing count data alone. In case some of the four data sources are lacking, the described analysis can be accommodated by omitting the corresponding data preparation steps or sub-models. Considering presence-only data or ascertained absences is still possible if either detection/non-detection or count data are lacking. In the latter case, the fitted model will either be an adaptation of the Royle-Nichols' (Royle & Nichols, 2003) model, or a site-occupancy model as in our applications. If both count and detection/non-detection data are lacking, the presented framework is not suitable.

If data are sparse, uncertainty intervals and interannual variability of the estimates from the N-mix model and from Poisson GLM were much larger than those from our integrated SDM (Figures 2 and 3). This is not surprising, as these methods can make use of a more limited proportion of the data only. Integrating 'cheaper' detection/non-detection data that is available for many more sites supports accurate and precise parameter estimation by the integrated SDM and the site-occupancy model. For trend estimation, the credible intervals of the site-occupancy estimates were smaller than those of the integrated SDM (Figure 3). However, the overall changes identified by the site-occupancy model were smaller than the one found by the other methods (Figure 3). Most likely, this is because changes in population size within occupied sites that do not lead to a change in the sites' occupancy state (such for instance a decline from 3 to 1) do not 'count' when analysing the data with a site-occupancy model.

Fitting the integrated SDM takes up more computer runtime which may be considered a disadvantage. The amount of time needed

to achieve convergence of the Markov chains in the Bayesian analyses is up to one order of magnitude greater for the integrated SDM, compared with the site-occupancy model that uses the same amount of data (though less information on abundance, as abundance inferences are much more indirect when modelling occupancy probability). However, as computer power is flexibly available nowadays and getting cheaper all the time, this should not be seen as a major disadvantage. Still, it can complicate data exploration and model development, as even a short test run might take quite some time. However, the improvement generated using all existing data can arguably compensate for the additional time investment in most cases. We hope that our R-code provided to combine the available data will lower the hurdle to tackle this endeavour.

Recently, several researchers have developed methods that allow integrating observation data of different quality (e.g. Dorazio, 2014; Pagel et al., 2014; Pacifici et al., 2017; Koshkina et al., 2017; Zipkin et al., 2017; Fletcher Jr. et al., 2019; Chapter 10 in Kéry & Royle, 2021). To our knowledge, however, ours is the first attempt where count data, detection/non-detection data, presence-only data and information on ascertained absences are all jointly analysed within a single, integrated SDM. We assume that our approach is particularly valuable when the amount of available data from one single data source only is too sparse to obtain reasonably precise estimates (rare species) or when changes in abundance may be masked when considering detection/non-detection only. In addition, drawbacks caused by a varying amount of data over time, space, or, more generally, covariate gradients, could be mitigated (Zipkin et al., 2017), as making use of a broader data base generally enhances the robustness of an analysis.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
Code and data to replicate the presented analyses in Application 1 and 2 is archived on the vogelwarte.ch Open Repository and Archive (https://doi.org/10.5281/zenodo.5840377).

ORCID
Nicolas Strebel https://orcid.org/0000-0003-2919-6732

REFERENCES
Adler, D. & Kelly, S. T. (2021). vioplot: violin plot. R package version 0.3.7. https://github.com/TomKellyGenetics/vioplot
Barker, R. J., & Sauer, J. R. (1992). Modelling population change from time series data. In *Wildlife 2001: Populations* (pp. 182–194). Springer.
Besbeas, P., Freeman, S. N., Morgan, B. J., & Catchpole, E. A. (2002). Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, *58*(3), 540–547.
Bibby, C., Jones, M., & Marsden, S. (2000). *Bird surveys* (p. 137). Birdlife International.
Bogaart, P., van der Loo, M., & Pannekoek, J. (2018). Rtrim: Trends and indices for monitoring data. R package version 2.0.6. https://CRAN.R-project.org/package=rtrim
Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.
Cooper, C. B., Dickinson, J., Phillips, T., & Bonney, R. (2007). Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society*, *12*(2), 11.
Dail, D., & Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, *67*(2), 577–587.
Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T., & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297.
Dorazio, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, *23*(12), 1472–1484.
Fewster, R. M., Buckland, S. T., Siriwardena, G. M., Baillie, S. R., & Wilson, J. D. (2000). Analysis of population trends for farmland birds using generalized additive models. *Ecology*, *81*, 1970–1984.
Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, *6*(4), 424–438.
Fletcher, R. J., Jr., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, *100*(6), e02710.
Freeman, S. N., Noble, D. G., Newson, S. E., & Baillie, S. R. (2007). Modelling population changes using data from different surveys: The common birds census and the breeding bird survey. *Bird Study*, *v. 54*, 61–72.
Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016.
Guélat, J., & Kéry, M. (2018). Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, *9*(6), 1614–1625.
Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., MA, M. C., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, *24*(3), 276–292.
Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C., Aspinall, R., & Hastie, T. (2006). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, *43*(3), 386–392.
Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009.
Hintermann, U., Weber, D., Zangger, A., & Schmill, J. (2000). Biodiversity monitoring in Switzerland. *Schriftenreihe für Landschaftspflege und Naturschutz*, *62*, 47–58.
Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*(2), 142–152.

Hooten, M. B., & Hefley, T. J. (2019). *Bringing Bayesian models to life*. CRC Press.

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, 85(1), 3–28.

Inger, R., Gregory, R., Duffy, J. P., Stott, I., Voříšek, P., & Gaston, K. J. (2015). Common European birds are declining rapidly while less abundant species' numbers are rising. *Ecology Letters*, 18(1), 28–36.

Isaac, N. J., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., & Jarvis, S. (2020). Data integration for large-scale models of species distributions. *Trends in Ecology & Evolution*, 35(1), 56–67.

Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Gutierrez, V. R., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S. T., & Fink, D. (2019). Best practices for making reliable inferences from citizen science data: Case study using eBird to estimate species distributions. *BioRxiv*, 574392.

Kammann, E. E., & Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52, 1–18.

Kellner, K. (2019). jagsUI: A wrapper around 'rjags' to streamline 'JAGS' analyses. R package version 1.5.1. https://CRAN.R-project.org/package=jagsUI

Kéry, M., & Royle, J. A. (2021). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press.

Kéry, M., Royle, J. A., & Schmid, H. (2005). Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications*, 15, 1450–1461.

Kéry, M., Royle, J. A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., & Zbinden, N. (2010). Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conservation Biology*, 24, 1388–1397.

Kéry, M., & Schmid, H. (2006). Estimating species richness: Calibrating a large avian monitoring programme. *Journal of Applied Ecology*, 43(1), 101–110.

Knaus, P., Antoniazza, S., Wechsler, S., Guélat, J., Kéry, M., Strebel, N., & Sattler, T. (2018). *Swiss breeding bird atlas 2013–2016. Distribution and population trends of birds in Switzerland and Liechtenstein*. Swiss Ornithological Institute.

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M., & Stone, L. (2017). Integrated species distribution models: Combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution*, 8(4), 420–430.

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, 23, 504–515.

Lebreton, J. D., Morgan, B. J., Pradel, R., & Freeman, S. N. (1995). A simultaneous survival rate analysis of dead recovery and live recapture data. *Biometrics*, 51(4), 1418–1428.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), 2248–2255.

Miller, D. A., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution*, 10(1), 22–37.

Müller, C. (2015). *Monitoring Feuchtgebiete. Mit Stiefeln und Ausdauer auf den Spuren seltener Arten*. Presentation, accessed on https://www.vogelwarte.ch/downloads/files/projekte/ueberwachung/CMu_Mitarbeitertagung_Sempach_2015.pdf. Sempach.

Pacifici, K., Reich, B. J., Miller, D. A., Gardner, B., Stauffer, G., Singh, S., McKerrow, A., & Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion. *Ecology*, 98(3), 840–850.

Pacifici, K., Reich, B. J., Miller, D. A., & Pease, B. S. (2019). Resolving misaligned spatial data with integrated species distribution models. *Ecology*, 100(6), e02709.

Pagel, J., Anderson, B. J., O'Hara, R. B., Cramer, W., Fox, R., Jeltsch, F., Roy, D. B., Thomas, C. D., & Schurr, F. M. (2014). Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods in Ecology and Evolution*, 5(8), 751–760.

Pannekoek, J., & van Strien, A. (1991). *TRIM: Trends & indices for monitoring data*. Statistics Netherlands.

Plummer, M. (2017). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Release 4.3.0.

R Core Team (2018). *R: A language and environment for statistical computing. R foundation for statistical computing*. www.R-project.org

Renner, I. W., 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(12), 2118–2128.

Rodrigues, A. S., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M., & Brooks, T. M. (2006). The value of the IUCN red list for conservation. *Trends in Ecology & Evolution*, 21(2), 71–76.

Rossman, S., Yackulic, C. B., Saunders, S. P., Reid, J., Davis, R., & Zipkin, E. F. (2016). Dynamic N-occupancy models: Estimating demographic rates and local abundance from detection-nondetection data. *Ecology*, 97(12), 3300–3307.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60, 108–115.

Royle, J. A., & Nichols, J. D. (2003). Estimating abundance from repeated presence–absence data or point counts. *Ecology*, 84(3), 777–790.

Sattler, T., Bontadina, F., Hirzel, A. H., & Arlettaz, R. (2007). Ecological niche modelling of two cryptic bat species calls for a reassessment of their conservation status. *Journal of Applied Ecology*, 44, 1188–1199.

Sauer, J. R., & Link, W. A. (2011). Analysis of the north American breeding bird survey using hierarchical models. *The Auk*, 128(1), 87–98.

Schaub, M., & Kéry, M. (2022). *Integrated population models – A Bayesian hierarchical perspective using R and BUGS*. Elsevier/Academic Press.

Schmid, H., Zbinden, N., & Keller, V. (2004). *Überwachung der Bestandsentwicklung häufiger Brutvögel in der Schweiz*. Swiss Ornithological Institute.

Sutherland, W. J., Pullin, A. S., Dolman, P. M., & Knight, T. M. (2004). The need for evidence-based conservation. *Trends in Ecology & Evolution*, 19, 305–308.

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, 13(6), 1790–1801.

Warton, D. I., Renner, I. W., & Ramp, D. (2013). Model-based control of observer bias for the analysis of presence-only data in ecology. *PLoS ONE*, 8(11), 1–9.

Zipkin, E. F., Rossman, S., Yackulic, C. B., Wiens, J. D., Thorson, J. T., Davis, R. J., & Grant, E. H. C. (2017). Integrating count and detection–nondetection data to model population dynamics. *Ecology*, 98(6), 1640–1650.

BIOSKETCH

Our interest is to develop and apply methods for analysing animal monitoring data, ranging from unstructured citizen science data over data originating from structured monitoring programmes up to data including demographic information on the individual level. We mainly focus on temporal trends and species distribution modelling.

Authors' contributions: NS and TS developed the idea and concept, NS and MK the methodology. NS conducted the analysis and led the writing of the manuscript. JG, MK and TS contributed to the manuscript.