

# Assessing the usefulness of citizen science data for habitat suitability modelling: Opportunistic reporting versus sampling based on a systematic protocol

Laura Henckel<sup>1</sup>  | Ute Bradter<sup>1</sup> | Mari Jönsson<sup>1</sup> | Nick J. B. Isaac<sup>2</sup> | Tord Snäll<sup>1</sup> 

<sup>1</sup>Swedish Species Information Centre (ArtDatabanken), Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden

<sup>2</sup>Centre for Ecology & Hydrology (CEH), Wallingford, UK

## Correspondence

Laura Henckel, Swedish Species Information Centre (ArtDatabanken), Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden.

Email: laurahenckel@gmail.com

## Funding information

Svenska Forskningsrådet Formas; 2015-2016 BiodivERsA COFUND call for research proposals

Editor: Raimundo Real

## Abstract

**Aim:** To evaluate the potential of models based on opportunistic reporting (OR) compared to models based on data from a systematic protocol (SP) for modelling species distributions. We compared model performance for eight forest bird species with contrasting spatial distributions, habitat requirements and rarity. Differences in the reporting of species were also assessed. Finally, we tested potential improvement of models when inferring high-quality absences from OR based on questionnaires sent to observers.

**Location:** Both datasets cover the same large area (Sweden) and time period (2000–2013).

**Methods:** Species distributions were modelled using logistic regression. Predictive performance of OR models to predict SP data was assessed based on AUC. We quantified the congruence in spatial predictions using Spearman's rank correlation coefficient. We related these results to species characteristics and reporting behaviour of observers. We also assessed the gain in predictive performance of OR models by adding inferred absences. Finally, we investigated the potential impact of sampling bias in OR.

**Results:** For all species, and despite the sampling biases, results from OR overall agreed well with those of SP, for the nationwide spatial congruence of habitat suitability maps and the selection and directions of species–environment relationships. The OR models also performed well in predicting the SP data. The predictive performance of the OR models increased with species rarity and even outperformed the SP model for the rarest species. No significant impact of observer behaviour was found.

**Main conclusions:** Relatively simple analyses with inferred absences could produce reliable spatial predictions of habitat suitability. This was especially true for rare species. OR data should be seen as a complement to SP, as the weakness of one is the strength of the other, and OR may be especially useful at large spatial scales or where no systematic data collection protocols exist.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

## KEY WORDS

biodiversity monitoring, citizen science, forest birds, habitat suitability models, opportunistic reporting, presence-only, pseudo-absences, standardized protocol, volunteer recording

## 1 | INTRODUCTION

Habitat suitability models are important tools to predict species' distributions (Elith & Leathwick, 2009), and for conservation and management (Franklin, 2013; Lawler, Wiersma, & Huettmann, 2011). These models require occurrence data collected across a variety of habitat types and covering a broad spatial extent (Bonney et al., 2009). Such large biodiversity surveys are costly, and volunteer "citizen science" recording can constitute a promising alternative (Devictor, Whittaker, & Beltrame, 2010) capable of producing large datasets with value for a wide range of ecological applications (Bonney et al., 2009). Citizen science data have been successfully used to improve knowledge in many research areas such as mapping the distribution of invasive species (Delaney, Sperling, Adams, & Leung, 2008), predicting seasonal dynamics of pathogens (Altizer, Hochachka, & Dhondt, 2004) or assessing the effect of the environment on breeding success (Rosenberg, Lowe, & Dhondt, 1999). These data can also be used to model species' habitat suitability or to describe population trends and range change (Kery et al., 2010; Mair, Harrison, Rätty, et al., 2017; Snäll, Kindvall, Nilsson, & Pärt, 2011). Moreover, citizen science allows the development of a science–society interface, raising awareness of nature conservation issues and allowing the dissemination of knowledge in society (Johnson et al., 2014; Price & Lee, 2013).

Citizen science is however a broad concept that can take different forms (Brown & Williams, 2018; Pocock, Tweddle, Savage, Robinson, & Roy, 2017) from highly systematic protocols (SP) to systems only based on opportunistic reporting (OR) with no sampling design (Pocock et al., 2017; Tulloch, Possingham, Joseph, Szabo, & Martin, 2013). A trade-off often arises between data quantity and quality (Devictor et al., 2010). At one end of the gradient are data from SP based on a network of sites, which are regularly surveyed, usually during specific time periods (Pocock et al., 2017). Representative sampling is ensured through a regular, stratified or random spatial distribution of sites and a precise protocol. A complete species list is recorded for each sampled site. Such monitoring programmes can produce high-quality data with little or no bias. At the other end in OR, volunteers are free to select the sampling period and location (Pocock et al., 2017). They may record only some of the species they found. Due to its greater flexibility, this type of citizen science creates large amounts of data, but with limitations (e.g. the data are "presence-only") and biases, for example uneven sampling effort across space or time, or an oversampling of rare or emblematic species (Bird et al., 2014). Consequently, care is required to analyse OR data to avoid false inferences about species distributions.

While justifiable concerns exist around the quality of OR data, the potential biases may be counter-balanced by the much larger sample size of data collected by volunteers (Hochachka, Martin, Doyle, & Krebs, 2000; Schmeller et al., 2009). Moreover, various solutions exist to analyse this kind of data, such as presence–background

methods (e.g. Maxent) or inferring non-detections from records of other species, thus making the data suitable for occupancy–detection models (Bradter et al., 2018; Isaac, van Strien, August, de Zeeuw, & Roy, 2014; van Strien, van Swaay, & Termaat, 2013). However, Bradter et al. (2018) found that logistic regression with high-quality inferred absences performed at least as well as more specialized methods.

Despite increasing amounts and availability of data from OR, there is a lack of evaluation of their reliability for habitat suitability modelling of species with different ecology, geographical distributions and observation biases. Particularly, it is important to compare results from OR against data from SP to evaluate the reliability of OR (Mair, Harrison, Jönsson, et al., 2017). Such comparative evaluation is needed to provide guidance on the best use of existing data and to direct future data collection. Indeed, although data from OR have drawbacks they may nevertheless fill a gap for some species or habitat types poorly covered by SP. Rather than seeing these OR data as a poor, low cost alternative, we advocate to investigate their potential advantages and uniqueness, testing their performance in relation to species characteristics and observation processes. A reason is the enormous amounts available, for example currently >1.3 billion reports on [www.gbif.org](http://www.gbif.org). This further motivates validating data from OR, defining their limitations and assets, and providing recommendations on how to deal with biases and improve model quality.

The overall aim of this study is to evaluate models based on OR compared to models based on SP concerning overall predictive performance for multiple species that are reported differently by observers. We studied eight forest bird species and applied habitat suitability models to two independent citizen science datasets, one OR and one SP covering the same large area (the whole of Sweden) and time period (2000–2013). We evaluated the predictive performance of all models, and at different spatial scales. Additionally, we evaluated the congruence of the spatial predictions between OR and SP. We then investigated the relative importance of species characteristics, ecological requirements, geographical distribution, rarity, detectability, observer identifiability and interest, on the predictive performance of the habitat suitability models. Finally, we investigated whether the observed differences between models based on OR and SP can be explained by potential sampling bias between habitat types.

## 2 | METHODS

### 2.1 | Study species

We selected eight forest bird species according to the following criteria: (1) they are of particular interest to bird watchers, and hence,

**TABLE 1** Data amounts per species and dataset

	Hazel grouse	Siberian jay	Siberian tit	Long-tailed tit	Red-breasted flycatcher	Grey-headed woodpecker	Lesser spotted woodpecker	Three-toed woodpecker
Presences OR	6,105	2,828	593	15,397	1,515	5,237	16,746	4,017
Inferred absences OR	4,831	6,357	5,650	3,024	7,361	5,168	3,343	4,561
Presences SP	633	826	121	441	65	81	180	305
Absences SP	12,611	12,418	13,123	12,803	13,179	13,163	13,064	12,939

Note: The number of presences, absences and inferred absences in data from the systematic protocol (SP) and from opportunistic reporting (OR), for eight forest bird species.

citizen science data have a high potential to fill data gaps; (2) they are relatively well-studied, allowing assessment of whether the species–habitat relationships suggested by models are realistic; (3) they vary in rarity, but are not so common as not to be consistently reported by some active reporters, a pre-requisite for our method of obtaining inferred absences; and (4) they have varied climate and habitat niches (e.g., tree species dominance), and most are considered to be negatively influenced by modern forest management. Hence, our study can provide a broader perspective of how citizen science data can facilitate the conservation of boreal forest birds. Eight species were selected: the hazel grouse (HG), *Tetrastes bonasia*; the Siberian jay (SJ), *Perisoreus infaustus*; the Siberian tit (ST), *Poecile cinctus*; the long-tailed tit (LTT), *Aegithalos caudatus*; the red-breasted flycatcher (RBF), *Ficedula parva*; the grey-headed woodpecker (GHW), *Picus canus*; the lesser spotted woodpecker (LSW), *Dendrocopos minor*; and the three-toed woodpecker (TTW), *Picoides tridactylus*.

The occurrence data covered the whole country, but two species occur only in the north (SJ, ST), three only in the south (RBF, LTT, LSW) and three have a more widespread distribution (HG, TTW, GHW). Most of these species are more or less specialists of old-growth forest (e.g., cavity nesters) and are negatively affected by intensive forest management (but to a lesser extent for HG and LTT) (Jansson & Angelstam, 1999; Swenson & Angelstam, 1993). See also Appendix S1 for further details on habitat requirement for each species. Their forest preferences range from coniferous forest (SJ, TTW, ST) to mixed (HG, GHW, LTT) and deciduous forest (LSW, RBF). Apart from the long-distant migrant RBF, all species are basically resident, but eruptive movements or nomadism occurs in some species.

## 2.2 | Data from the Systematic Protocol (SP)

The Swedish Bird Survey is a national monitoring programme, where the data are collected by volunteer ornithologists along  $2 \times 2$  km quadratic routes on fixed location, regularly distributed along a grid with 25-km resolution across Sweden. A total of 716 squares are distributed across the country, and a portion of them is surveyed once per year between May and July. Each square is composed of eight transects of 1 km and eight points counts, but only transects data were used in this study. We obtained a total of 13,244 transects, see

Appendix S2 for details. For each transect, bird observations were converted to presences or absences per species.

## 2.3 | Data from opportunistic reporting

Species observations were downloaded for the period 2000–2013 from the Swedish Species Observation System (Artportalen). See Appendix S3 for a detailed description of our procedure for opportunistic reporting (OR) data extraction and cleaning, which is here summarized: as some Swedish Bird Survey records are sometimes reported in the OR database, we removed these observations in order to keep data from OR independent from data from SP. We also excluded uncertain observations (e.g. uncertain location) and those with a spatial inaccuracy greater than 500 m. Following Bradter et al. (2018), we choose logistic regression for our modelling (see Section 2.8). As these models require presence–absence data, we used a specific method to infer absences based on questionnaires sent to observers. We first identified a subset of 20 observers for each species that had reported the highest number of unique locations in Sweden between 2000 and 2013, to ensure using records from many different locations, rather than many observations. To ensure good geographical coverage, we further divided Sweden into four geographical regions and identified the 20 reporters that had reported the most locations of the eight species in each region. Thereafter, we sent a questionnaire to the 94 observers identified asking about their skills and habits for identifying and reporting the eight focal species. We received 60 responses. To infer the absences for a species, we used the reports from observers that stated they always reported the focal species when found and were able to identify the species by sight and sound. We then inferred absences when and where other species but not the focal species were reported by these observers, keeping only inferred absences with a minimum search effort corresponding to more than five bird species recorded.

## 2.4 | Data amount per species and protocol

Observations from both OR and SP were aggregated into three time periods to match with the forest predictors, which were only available for 2000, 2005 and 2010 (see below): 2000–2002 (“2000”

henceforth); 2003–2007 (“2005”); 2008–2013 (“2010”). The three periods were then analysed together in a single model. The number of records obtained by species after data cleaning from the whole period 2000–2013 is detailed in Table 1.

## 2.5 | Environmental data

We chose a set of environmental predictors based on existing knowledge about the species ecology (Appendix S1) that can be grouped into three categories: climate, forest and landscape.

### 2.5.1 | Climate predictors

We used mean monthly temperature and precipitation from the EURO4M Mesan data, on a grid with a mesh size of around 5 km<sup>2</sup> (Landelius, Dahlgren, Gollvik, Jansson, & Olsson, 2016). We used mean monthly temperature and precipitation across the whole period 1989–2010 in spring (April to June) and winter (December to February) as predictors in our analyses.

### 2.5.2 | Forest predictors

Depending on the species ecology, we included the most relevant predictors among forest age, forest volume, % of spruce, % pine, % coniferous and % deciduous trees (Appendix S1). We used a spatially explicit description of the forest composition and structure based on estimates combining satellite imagery with data from the National Forest Inventory of Sweden (i.e., kNN data, Reese et al., 2003). This dataset is available for the years 2000, 2005 and 2010 with a pixel size of 25 by 25 m.

### 2.5.3 | Grain size

For each species, the environmental predictors were computed for a 1 × 1 km square centred on each transect mid-point for SP data and each record for OR. This scale has often been chosen in other studies of these species (three-toed woodpecker: Angelstam, Angelstam, Ekelund, & Schlaepfer, 2004; grey-headed woodpecker: Gjerde, Sætersdal, & Nilsen, 2005; hazel grouse: Hofstetter, Arlettaz, Bollmann, & Braunisch, 2015; long-tailed tit: Lindbladh, Felton, Trubins, & Sallnäs, 2011) and constitutes a suitable compromise between territory/home range sizes of the eight species. For the forest data, we computed the mean value of all 25 × 25 m pixels inside the 1-km<sup>2</sup> grid.

### 2.5.4 | Landscape predictors

We included the mean elevation of the site (1 × 1 km square) and the percentage of forest in 1 × 1 or 3 × 3 km square buffers, as some

species avoid open areas and prefer large or well-connected forest patches and have minimum habitat size requirements (Appendix S1). To compute this last predictor, we considered only forest 25 × 25 m pixels with a total standing tree volume larger than 25 m<sup>3</sup>/ha from the raster of forest volume (kNN data), which corresponds to the minimum volume for the classification as young forest in Brotons, Mönkkönen, Huhta, Nikula, and Rajasärkkä (2003). We also included as predictor the Euclidean distance to the nearest city to estimate site accessibility, as this may explain the sampling bias in OR (more reports from more accessible areas). Finally, we included the total number of bird observations reported in Artportalen for each sampling period and the total numbers of bird records around each observation in OR (1 km<sup>2</sup>), as a proxy of the sampling effort (see Appendix S4). As not all transects in SP are necessarily sampled every year and because the three periods vary in duration, we also tested to include the number of times a transect was surveyed as a predictor in SP models.

## 2.6 | Species characteristics

To investigate how predictive performance of the habitat suitability models varied depending on species characteristics (i.e., ecological requirements, geographical distribution, rarity or detectability), we described and classified species based on six main characteristics, see Table 2 (see also Appendix S1 for other characteristics).

## 2.7 | Observer's behaviour and data characteristics

Based on the 60 questionnaires, we assessed how observers' behaviour differed for each of the targeted species. We calculated the percentage that declared to always report observations of a focal species and that are confident in their own ability to recognize the species (by sight and sound). We further estimated this percentage among the observers that had reported the species in its main distribution area (as the recognition rate is expected to decrease when the species become rarer, but the reporting rate may increase). We also assessed the impact of prevalence (ratio presence/absence) and sample size in data from OR, as it may impact the predictive performance of the model (Barbet-Massin, Jiguet, Albert, & Thuiller, 2012; Cumming, 2000; McPherson, Jetz, & Rogers, 2004). Finally, we computed a global index for the quality of reporting, as the summed scores of recognition and reporting for each species.

## 2.8 | Statistical analysis

For each species, we proceeded as follows:

1. National-level presence–absence (SP and OR) data were modelled using logistic regression, a standard method for binary data (Faraway, 2006), using a Binomial distribution and a logit link.

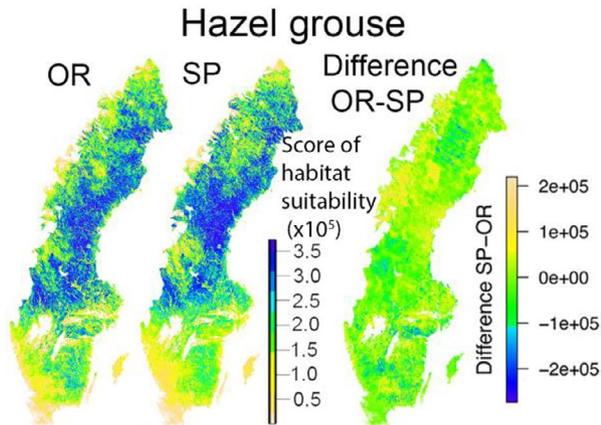
TABLE 2 Species ecological characteristics

Ecological characteristics	Definition	Expected impact on data quality	Source
Mean home range	Mean home range of the species during the whole year	Species with large home ranges are likely to have a lower probability of detection	Literature
Variation in home range through the year	Difference in home range between breeding and non-breeding season	All species except the red-breasted flycatcher are present in Sweden year-round, and the home range is larger in the non-breeding season. This change in home range between seasons may explain differences in prediction between OR and SP (the latter being collected during the breeding season only)	Literature
Body mass	Mean body mass of the species	Larger species are more likely to be detected	Literature
Longevity	Reported lifespan of the species	Affect the probability to repeatedly see the species at the same site in different time periods	Literature
Rarity	Inverse of the number of presences from the systematic protocol (SP)	Rare species may be of higher interest to observers and be more consistently reported	Estimated based on our data
Climate range	Difference in temperature (max-min) in the 90% main distribution area of the species	Species with narrow climatic distribution range might be less easily identified by observers outside the main distribution range, and SP datasets from the breeding season may not capture the full climatic range of species	Estimated based on our data

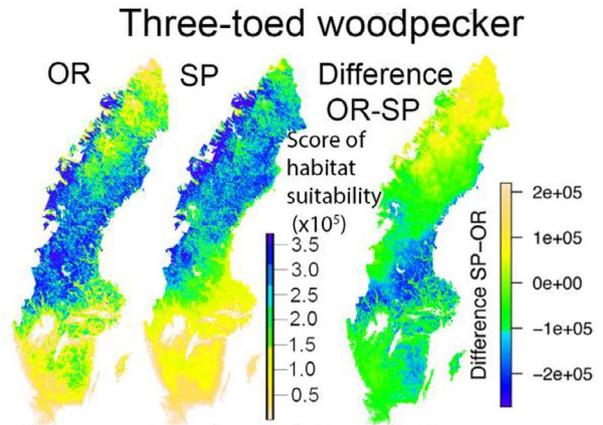
Note: The six main ecological characteristics investigated for impacts on the predictive performance of the habitat suitability models. See Appendix S1 for literature references.

For each species, predictors were first chosen according to our knowledge of the species ecology (Appendix S1). Then, we performed a model selection retaining the model with lowest AIC (Akaike's information criterion [Akaike, 1974]), one model per species and per data type. This included testing all possible combinations of predictors and biologically reasonable quadratic effects and interactions between predictors.

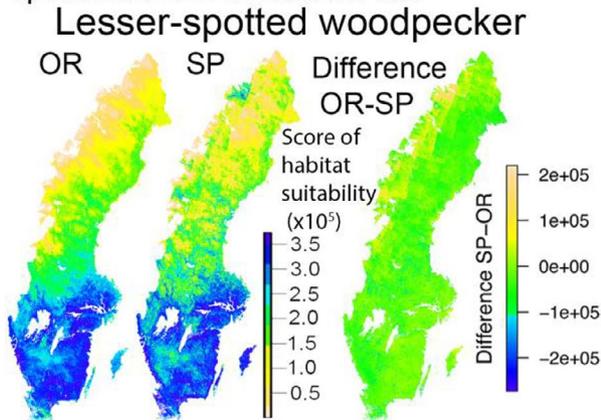
- The best model per species and dataset selected in step 1 was then evaluated by cross-validation. We evaluated the ability of the model based on OR to predict the observations from SP. We also evaluated the predictive performance of the models based on SP using a leave-one-out cross-validation, excluding successively each transect from the training data and using the excluded transect for validation. We produced a ROC curve and computed an AUC score (area under the receiver-operating curve) for each model. For further understanding of the predictive performance of the national model at finer spatial scales, we also present AUC for the main distribution area, specifically a convex polygon containing 90% of the presences in data from SP.
- As the absolute predictions from SP and OR were not directly comparable, because the non-systematic sampling changes the prevalence of presences in OR data compared to SP data, we ranked the predictions (from the lowest to the highest score) to compare the relative habitat suitability. We then quantified the congruence in spatial predictions between models from OR and SP data by calculating the Spearman rank correlation coefficient between the predictions. Predictions were computed for the whole of Sweden in a grid of 1 km<sup>2</sup> cell size, using forest predictors of 2010.
- We investigated how the model performances and differences of congruence between models from OR and SP data were related to the specific species characteristics and reporting behaviour of observers (the observation process). We tested if the differences in AUC between models based on OR and SP varied depending on these predictors (see Sections 2.6 and 2.7) using linear regression.
- We investigated the gain in predictive performance (AUC) of models based on OR by progressively increasing the number of inferred absences in the data. We randomly drew a subset of inferred absences from 0.05% to 100% of the amount of absences in the data from SP. At each step, we performed a cross-validation and used the model fitted on OR data with this subset of inferred absences to predict presences and absences in SP. We repeated this procedure 100 times to define a confidence interval, and compared the mean AUC of the models based on OR for each level of absences, with the AUC of the model based on SP. For each species, we estimated the gain in AUC by subtracting the maximum AUC (as the threshold was obtained with very few inferred absences, this maximum AUC was based on the model with 10% of absences) with the minimum AUC (AUC in the model with 0.05% of absences). This 10% threshold was chosen as it corresponds to the asymptote for all species (see Section 3) and to keep results comparable as we do not have the same total number of absences for all species.
- Finally, we assessed whether the observed differences between models based on OR and SP can be explained by potential sampling bias between habitat types. For each habitat predictor and each species, we compared the distribution of the data from OR and SP, for example distribution of presences and absences along



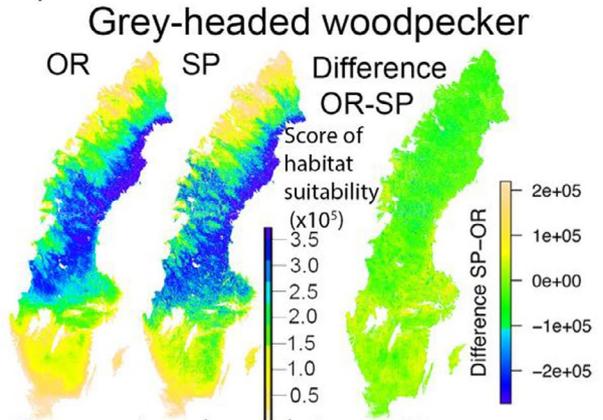
Spearman's rank correlation : 0.86



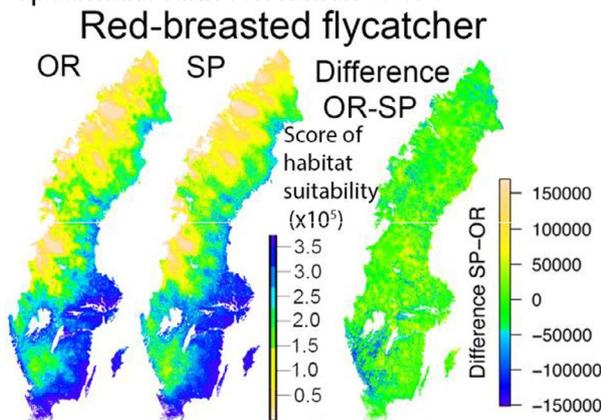
Spearman's rank correlation : 0.78



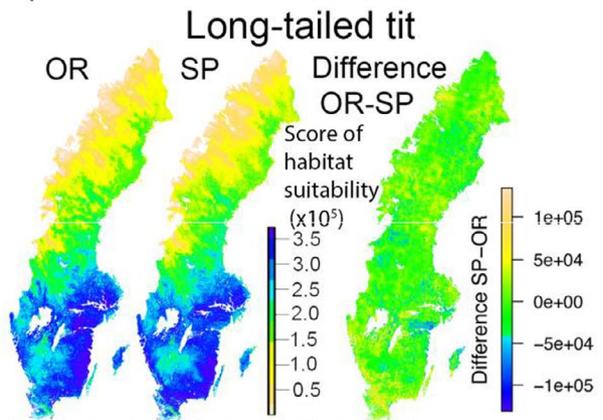
Spearman's rank correlation : 0.89



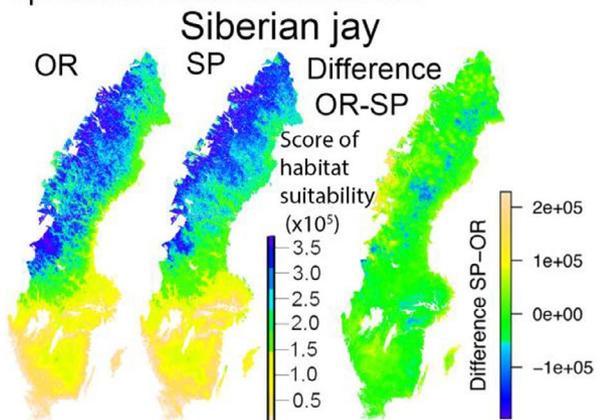
Spearman's rank correlation : 0.91



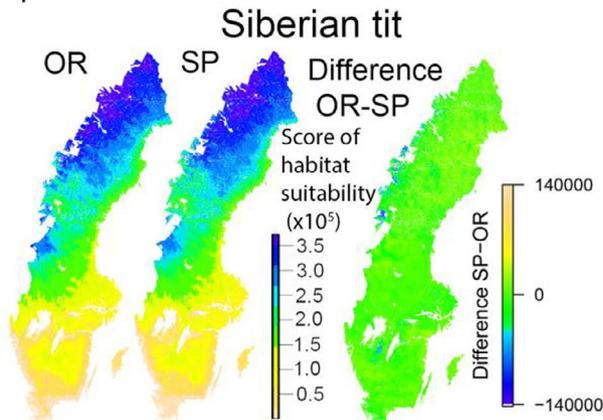
Spearman's rank correlation : 0.95



Spearman's rank correlation : 0.97



Spearman's rank correlation : 0.93



Spearman's rank correlation : 0.99

**FIGURE 1** Spatial predictions (and difference between them) using models based on data from opportunistic reporting (OR) and systematic protocol (SP) for the eight bird species. The maps on the left and middle correspond to ranked predictions for habitat suitability from OR and SP, respectively. Spearman's rank correlations between these two maps are presented for each species. The map on the right shows the difference between these two maps. Green areas correspond to low difference, while blue indicates that OR predicts a higher habitat suitability than SP, and yellow indicates the opposite

a gradient of forest age, % of forest, % of deciduous tree, etc. We also compared these distributions with the distribution of these predictors at national scale. The observed distribution for the whole of Sweden is obtained based on grid with cell size of  $1 \times 1$  km covering the whole of Sweden. Other distributions are based on an  $1 \text{ km}^2$  square around each observation (e.g. % of forest), for both presences and absences and both protocol (OR and SP). We then compared the histograms for OR, SP and the whole of Sweden per class of percentage.

All analyses were performed using R software 3.6.0 (R Core Team, 2019). Model selection was performed based on AIC using the dredge function of the package MuMIn (Barton, 2009). Cross-validation was performed using the package AUC (Ballings & Van den Poel, 2013).

### 3 | RESULTS

#### 3.1 | Congruence between models from OR and SP

##### 3.1.1 | Spatial congruence

A high congruence in habitat suitability maps from OR and SP models was observed for all the species at national scale (Figure 1). Spearman's rank correlation coefficient ranged between 0.78 (three-toed woodpecker) and 0.99 (Siberian tit). Overall, models from SP predicted a slightly higher habitat suitability in the northwest for four species, but the pattern was opposite for Siberian tit. Otherwise, differences between predictions from OR and SP models varied between species.

##### 3.1.2 | Model predictors

Although there were some differences, the main selected predictors were the same in models based on data from OR and SP (Appendix S5). Furthermore, predictors and their associated coefficient estimates were mainly in accordance with our expectations based on available knowledge about the species ecology (Appendix S1). However, some unexpected effects were also observed for some predictors in models based on data from OR, as for example a small negative effect of forest age on grey-headed woodpecker occurrence (Appendix S5).

##### 3.1.3 | Model evaluation

For almost all species, models built on OR data can discriminate between presences and absences in SP almost as well as models built

on SP data themselves (Figure S6.1 and Table S6.1). For one species (red-breasted flycatcher), the model built on OR data was even better at predicting SP data than the model based on SP. The largest difference between the predictions of the models from OR and SP was observed for the three-toed woodpecker (delta AUC SP-OR = 0.04, Figure S6.1). Models from OR and SP both had a satisfactory ( $>0.7$ ) or good ( $>0.8$ ) predictive performance. The AUC scores for the final models ranged from 0.75 (red-breasted flycatcher) to 0.93 (Siberian tit) for SP, and from 0.74 (three-toed woodpecker) to 0.93 (Siberian tit) for OR (Figure S6.1, Table S6.1).

When assessing the model performance for making predictions in the main distribution area, the predictive performance decreased for all species, but we did not observe strong differences between the predictions of models based on OR and SP. Models based on OR were almost as good as those based on SP in discriminating presences and absences in data from SP, with only slightly lower AUC scores. For the red-breasted flycatcher, the model based on data from OR remained better than the one based on data from SP (AUC = 0.59 for SP vs. AUC = 0.65 for OR). We noticed however that model performances for the species with the highest AUC at national scale (e.g. Siberian tit) did not perform better for the main distribution area (Figure S6.2 and Table S6.1).

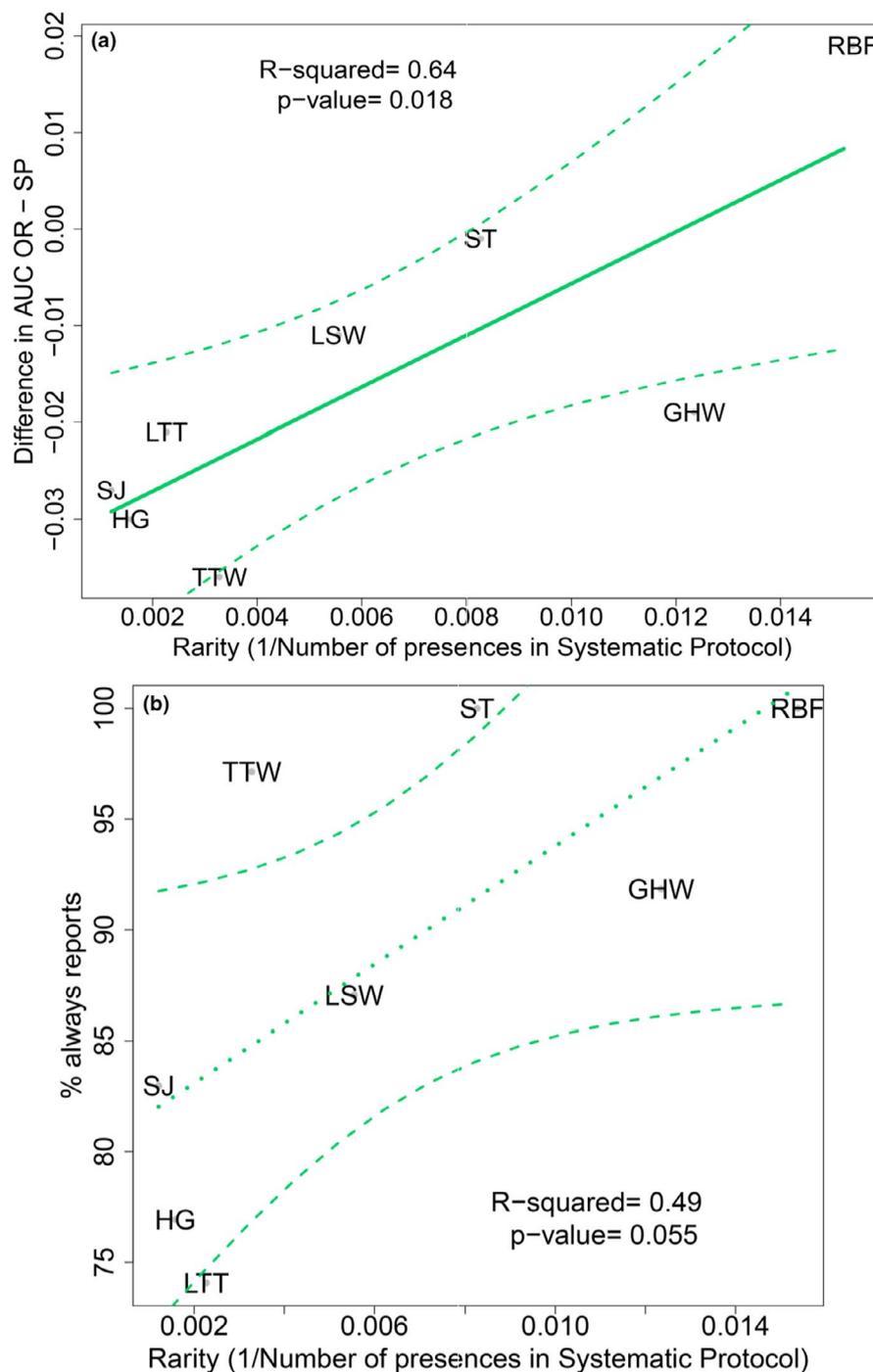
#### 3.2 | Effects of sampling biases

OR data contained much more observations in the most densely populated areas. Distance to cities was a better proxy for sampling effort than total number of observations per time period in  $1 \text{ km}^2$ , which did not improve the model (Appendix S4). Besides the geographical bias towards populated areas, OR data were also biased by habitat type, but only for open habitats (Appendix S7). The most densely forested landscapes, with higher percentages of forest in  $1 \text{ km}^2$  squares, were under-sampled in the OR data.

#### 3.3 | Differences explained by species ecological characteristics

The predictive performance of models based on data from OR became significantly closer to those based on data from SP with increasing species rarity (Figure 2a,  $R^2 = .64$ ,  $p = .018$ ). For the rarest species (red-breasted flycatcher), the model based on data from OR was even better than the one based on SP at predicting presences and absences in the SP data (Figure S6.1). Species further tend to be more consistently reported (percentage of observers that always report the species) in OR with increasing rarity (Figure 2b,  $R^2 = .49$ ,

**FIGURE 2** Predictive performance of the models in relation to the rarity of the species. (a) Difference in predictive performance between models based on data from opportunistic reporting (OR) and systematic protocol (SP) in relation to the rarity of the species. The predictive performance is estimated based on the AUC (area under the curve) of the models. The rarity of the species is defined as the inverse of the number of presences in SP. The species are designated by their acronyms: Siberian jay (SJ), Siberian tit (ST), hazel grouse (HG), three-toed woodpecker (TTW), grey-headed woodpecker (GHW), lesser spotted woodpecker (LSW), long-tailed tit (LTT), red-breasted flycatcher (RBF). (b) Reporting rate in relation to the rarity of the species. The reporting rate is defined as the percentage of observers in the OR questionnaire that always report the species at the national scale. The rarity of the species is defined as the inverse of the number of presences in SP



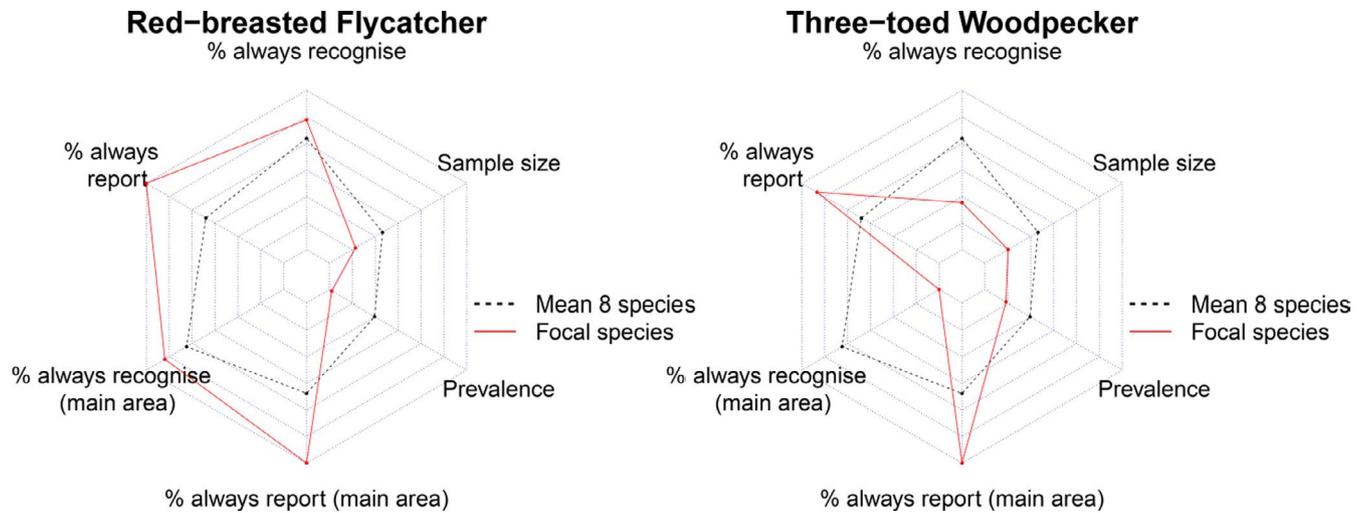
$p = .055$ ). We found no support for other associations between the predictive performance of the OR model and species characteristics (Appendix S8).

### 3.4 | Impact of observer's behaviour and data characteristics

Contrary to our expectations, the relative predictive performance of models based on data from OR in comparison with SP (difference in AUC) could not be explained by differences in observer's skills

or consistency in reporting, so the agreement between models was not higher for well reported or more easily recognizable species. The global index of reporting quality did not explain the differences in AUC between models based on data from OR and SP ( $R^2 = -.24$ ,  $p = .22$ , Figure S9.1). Similarly, the prevalence and sample size did not significantly impact the predictive performance of the models.

Although no significant relationship was detected, it can be noticed that our best model based on data from OR, that exceeded the predictive performance of the model based on SP, was obtained for the red-breasted flycatcher that was both among the most easily recognizable and consistently reported species (Figure 3). The model



**FIGURE 3** Reporting scores for the most comprehensively (red-breasted flycatcher) and the most restrictively (three-toed woodpecker) reported species in comparison with the mean for the eight species for different reporter's behaviour and data characteristics: the percentage of observers that always report the species (at national scale), the percentage of observers that can identify the species (at national scale), the percentage of observers that always report the species (in the main distribution area of the species, defined by a convex polygon containing 90% of the presences in systematic protocol), the percentage of observers that can identify the species (in the main distribution area of the species), the sample size in opportunistic reporting (OR) and the prevalence (ratio between the number of presence and the number or absences) in OR

with the poorest performance was obtained for three-toed woodpecker, a species considered to be among the most difficult to identify by observers (poorly recognized by call) (Figure 3).

### 3.5 | How many inferred absences do we need?

The predictive performance of models based on data from OR increased with increasing number of inferred absences (Figure 4). However, very few inferred absences were required to reach the maximum AUC. For all species, the plateau in AUC was reached around the 1% level of the total number of absences in data from SP (around 10 inferred absences). However, the relatively high variability in AUC at the 1% level for most species suggests that a higher number of inferred absences (up to 5% or 10% depending on species) will lead to more stable results (Figure 4). Note that the inferred absences are of high quality in that each inferred absence corresponds on average to 23 to 83 observations depending on species (minimum six; Figure 4a). The increase in predictive performance of the model varied among species, ranging between 1% (Siberian tit) and 20% (hazel grouse) to reaching performance asymptote at around 1% inferred absences. The AUC increased more for models with an initial low AUC at a low percentage of inferred absences (Figure 4b).

## 4 | DISCUSSION

Our study provides evidence that habitat suitability models from OR can provide similar predictions of habitat suitability as models from SP, for multiple species with varied characteristics, ecological requirements and observation biases. These findings widens the inferences

compared to studies of single bird (Bradter et al., 2018) and fungal (Mair, Harrison, Jönsson, et al., 2017) species. Indeed, models from both OR and SP data were highly congruent, and in accordance with the available knowledge on species ecology. Furthermore, models from OR predicted data from SP almost as accurately as the models from SP themselves. Some of the local mismatches (e.g. three-toed woodpecker in the north) may be explained by the lack of northern OR data, or by the poor quality of some forest predictors in this area, especially where forest age is high (Bradter et al., 2018). For one species (red-breasted flycatcher), the model based on data from OR provided even better predictions than those based on SP. This result is especially robust given the fact that our cross-validation of the OR models used an independent dataset constituting a more difficult test than the test used to assess the predictive performance of the model based on data from SP. Indeed, our thorough data cleaning procedure and our protocol to infer absences based on questionnaires should have made the quality of our OR dataset higher than many earlier evaluated datasets (Barbet-Massin et al., 2012). However, this work was straightforward and we thus show that with some relatively simple analyses and inferred absences, OR can provide reliable spatial predictions of habitat suitability at large spatial scales or in areas where no systematic protocols exist. In accordance with Bradter et al. (2018) and Huang and Frimpong (2015), we demonstrate that logistic regression using inferred absences produce reliable results that may outperform other methods based on presences paired with background or pseudo-absence data.

The high congruence between OR and SP models is striking considering the sampling biases in data from OR. OR data contain much more data from densely populated areas and more open habitats. This bias towards urban area or more accessible sites seems frequent in studies using citizen sciences data (Callaghan et al., 2020; Millar,

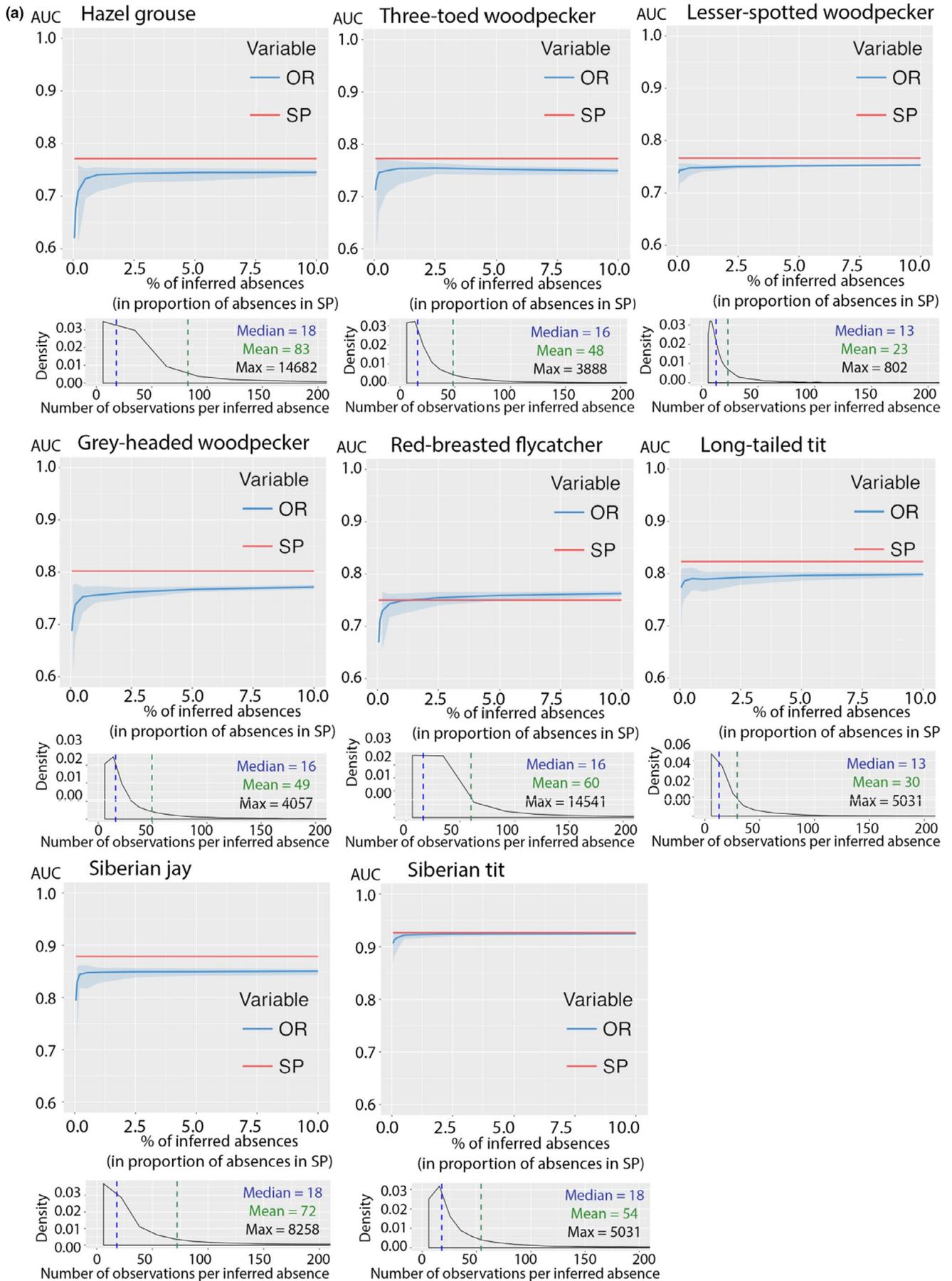
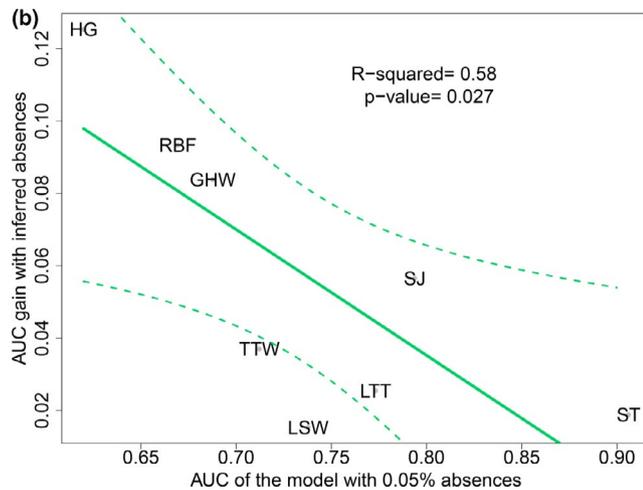


FIGURE 4 (Continues)



**FIGURE 4** How much absences are enough? (a) Increase in predictive performance with the percentage of absences included in the models of each bird species (upper graphs). The percentage of absences on the x-axis is expressed in relation to the number of absences in systematic protocol (SP). 100% is the equivalent of including the same number of inferred absences as the number of absences in the SP, see Table 1 for the total number of inferred absences per species. As the maximum AUC (area under the curve) is rapidly reached, we scaled the x-axis from 0% to 10% to improve visibility. As each inferred absence correspond in average to many bird observations, we also present for each species the number of observations per inferred absences to have an estimation of the sampling effort that each inferred absence represents (lower graphs). (b) Gain in predictive performance with inferred absences among species. For each species, we estimated the gain in predictive performance by subtracting the maximum AUC (see threshold in the curve Figure 4a) with the minimum AUC. As the threshold for maximum AUC is obtained with very few absences, we defined the maximum AUC based on the model with 10% of absences and the minimum AUC based on the model with 0.05% of absences. The AUC gain of adding inferred absence decrease when the AUC of the model with 0.05% of absences increase. This means that adding inferred absences will have more impact on improving model quality if the model without inferred absences has a low quality (low or medium AUC)

Hazell, & Melles, 2019; Neyens et al., 2019; Sumner, Bevan, Hart, & Isaac, 2019). We included the distance to cities in our models, as a proxy for accessibility, which better accounted for this bias than the number of reports per spatial unit. As only presences were reported in our dataset, this last variable may fail to represent well the real sampling effort. In accordance with our results, other studies show that large amounts of data in OR can compensate for bias and produce reliable results when sampling thresholds are met (Callaghan et al., 2020; Horns, Adler, & Şekercioğlu, 2018). Accessibility also probably explains the bias towards open habitat in OR data, although also other factors can be involved, for example higher visibility for bird watching, or higher interest for forest edges that combine various habitat types. This known bias (e.g., Kallimanis, Panitsa, & Dimopoulos, 2017) may therefore induce better reporting of edge species than strict forest interior species.

Models based on OR produce predictions congruent to those based on SP for all eight species. More specifically, we detected an increase in relative predictive performance of the OR models (compared with SP) among rare species, in accordance to Sardà-Palomera et al. (2012), possibly because the SP sites typically rarely match the habitat requirement of these species (Ottvall et al., 2009). The low predictive performance of SP models for rare species can also be due to the low prevalence of these species in SP data. Such impact of prevalence on predictive performance has previously been shown (Cumming, 2000; Huang & Frimpong, 2015; McPherson et al., 2004). This is unlikely the case in OR but may be a problem in SP. Indeed, we did not find any effect of prevalence on the predictive performance of OR models, probably due to the high sampling effort for all focal species. On the contrary for SP, for red-breasted flycatcher our dataset contained only 65 presences for 13,179 absences. This may also explain the lower predictive power of the model based on SP for this species compared to the OR model.

We did not detect any relationship between difference in predictive performance of OR and SP models and the reporting quality between species. However, the rarest species were the most consistently reported in OR, in agreement with studies on butterfly and dragonfly species (van Strien et al., 2013). We therefore agree with previous work, which includes organism groups other than birds, that OR data may be of particular value for species where systematic reports do not exist or require high sampling effort, for example longhorn beetles (Snäll, Forslund, Jeppsson, Lindhe, & O'Hara, 2014) or lynx (Louvrier et al., 2019). The reporting rate of rare species is also likely to increase with observers' expertise (Johnston, Fink, Hochachka, and Kelling (2018). Other study observed an interaction effect between the prevalence/ubiquity of the species and the sampling effort that impact on the predictive performance of models based on OR (Steen, Elphick, & Tingley, 2019). Note though that in amphibians and reptiles, common rather than rare species seem to be better reported (Tiago, Pereira, and Capinha (2017).

Accounting for the impact of observer expertise on detection probability can improve the model fit and the predictive performance of models (Johnston et al. (2018), but including it in models does not necessarily improve estimates of population trends (Eglington, Davis, Joys, Chamberlain, & Noble, 2010). We agree with Tulloch and Szabo, (2012) that even if an observer effect is not included in the model, collecting information on observer behaviour helps identifying OR data issues and controlling them.

Also the risk of misidentification varies depending on both rarity and observer skills, with more false-positive reports of rare species by skilled observers but more false-positive reports of common species by less experienced observers (Farmer, Leonard, & Horn, 2012). The impact of misidentifications on model performance is difficult to assess and vary between studies (Cruickshank, Bühler, & Schmidt, 2019; Ruiz-Gutierrez, Hooten, & Campbell Grant, 2016). We assume that the risk of misidentification of rare species is quite low in our dataset due to high self-validation control carried out by the bird watching community.

We did not find any other relationship between the predictive performance of the OR models and species ecological characteristics. However, we focused on a selection of forest specialists, thus a small number of species resulting in low statistical power. Another study on 195 birds species showed a significant impact of body size, diet and habitat specialization on detectability (Johnston et al., 2014). Moreover, lower population density, small body size or less charismatic species may also result in lower detection (Fitzpatrick, Preisser, Ellison, & Elkinton, 2009; Steger, Butt, & Hooten, 2017). However, our worst model concerned the three-toed woodpecker, considered the most difficult species to identify by observers (especially by call only). This suggests that species characteristics such as identifiability, volume or frequency of the song/call has stronger impact on detection in ecosystems with low visibility, for example forest. Species that are difficult to identify, for example lacking distinguishing features or requiring knowledge of vocalizations, are often under-reported in citizen science (Crall et al., 2011; Ratnieks et al., 2016; Swanson, Kosmala, Lintott, & Packer, 2016).

Our results demonstrate that adding even a small number of inferred absences can increase the predictive performance of the OR models by up to 20%, depending on the initial predictive performance of the model. Absence locations in this study were actually visited by reporters who consistently report a focal species and are skilled in the identification of the species. Furthermore, our absence locations fulfilled a minimum search effort threshold. Contrary to Lobo and Tognelli (2011), we note that only a low number of inferred absences were required to improve our models which cover a large extent, but inferred absences can be considered of high quality compared to classical methods using pseudo-absences (Barbet-Massin et al., 2012; Bradter et al., 2018), where little is known about the pseudo-absence locations. Although Lobo and Tognelli (2011) suggest that the number of pseudo-absences may be more important than their location, covering a large environmental gradient is nevertheless preferable to limit biases (Bradter et al., 2018; Lobo & Tognelli, 2011; Thuiller, Brotons, Araújo, & Lavorel, 2004). Here, we selected the observers to which we sent the questionnaire not only based on the number of records they had reported, but also based on the spatial distribution of their observations, to achieve a wide spatial spread. This simple method focuses more on quality than quantity, with quite few inferred absences but covering a wide area, increasing the chances that the environmental space is well covered. This method can be easily applied to other contexts or taxa, and requires only a questionnaire answered by a small subset of the most active observers. However, it may not be applicable to very common species, as few reporters may consistently report such species. It may then be impossible to obtain enough inferred absences or to achieve a wide enough spatial distribution of absences. Alternatively, inferred absences based on historical presences of other species in atlas data may be applicable (Frimpong, Huang, & Liang, 2016; Huang & Frimpong, 2015). We however encourage observers to report non-detections or use check-lists, as

this additional information would likely increase the robustness of models based on OR data (Isaac & Pocock, 2015).

In conclusion, although OR data contain biases, models based on data from OR can provide reliable predictions of habitat suitability. This is especially true for rare species difficult to detect without a high observation effort in time and space. We think that OR data should be seen as a complement to SP, rather than an alternative, as the weakness of one is the strength of the other (Miller, Pacifici, Sanderlin, & Reich, 2019). Different methods also exist to combine these data (Fletcher et al., 2019; Isaac et al., 2020; Miller et al., 2019b). Moreover, OR often promotes the establishment of long-term networks of observers and provide an information exchange platform between scientists, society and conservationists (Sullivan et al., 2009). However, it is clear that OR programmes cannot fully substitute long-term SP programmes, the latter often designed to answer specific ecological questions dependent on replicated sampling (Bayraktarov et al., 2019).

#### ACKNOWLEDGMENTS

We thank all the volunteer observers for reporting species in the Swedish Species Observation System (Artportalen), the participants of our questionnaire and the volunteers and coordinators of the Swedish Breeding Bird survey (Lund University). We thank Johan Nilsson, Johan Södercrantz and Ragnar Hall for advising on the design of our questionnaire and Johan Nilsson for extracting observations from Artportalen. We also thank Laura Janousek for her useful comments on earlier versions of the manuscript. The opportunistically collected data were obtained from the Swedish Species Observation System funded by the Swedish Environmental Protection Agency. This research was funded through the 2015–2016 BiodivERsA COFUND call for research proposals, with the national funder Formas and SLU.

#### DATA AVAILABILITY STATEMENT

Species observation data are available from the Swedish Lifewatch website ([www.analysisportal.se](http://www.analysisportal.se)) and in Dryad Digital Repository <https://doi.org/10.5061/dryad.8w9ghx3jj> (Henckel et al., 2020). Presence-only data and inferred absences from opportunistic reporting for the eight species are also accessible from dryad. National forest data ("kNN-Sweden") can be downloaded from <http://skogskarta.slu.se>, data on urban areas from <https://www.scb.se/hitta-statistik/regional-statistik-och-kartor/geodata/oppna-geodata/tatorter/>, data on settlements from <https://www.scb.se/hitta-statistik/regional-statistik-och-kartor/geodata/oppna-geodata/smaorter/>, and data on elevation from <http://www.lantmateriet.se/en/Maps-and-geographic-information/Maps/oppna-data/hamta-oppna-geodata/#faq:gsd-hojddata-grid-50>. The climate data ("EURO4M Mesan dataset") are available through <https://esg-dn1.nsc.liu.se/projects/esgf-liu/> (search for "mesan" after selecting a Federated ESFG-CoG Node).

#### ORCID

Laura Henckel  <https://orcid.org/0000-0002-1944-319X>

Tord Snäll  <https://orcid.org/0000-0001-5856-5539>

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Altizer, S., Hochachka, W. M., & Dhondt, A. A. (2004). Seasonal dynamics of mycoplasmal conjunctivitis in eastern North American house finches. *Journal of Animal Ecology*, 73(2), 309–322. <https://doi.org/10.1111/j.0021-8790.2004.00807.x>
- Angelstam, P., Ekelund, P., & Schlaepfer, R. (2004). Dead wood threshold values for the three-toed woodpecker presence in boreal and sub-Alpine forest. *Biological Conservation*, 119, 305–318. <https://doi.org/10.1016/j.biocon.2003.11.014>
- Ballings, M., & Van den Poel, D. (2013). *AUC: Threshold independent performance measures for probabilistic classifiers*. Retrieved from <https://cran.r-project.org/package=AUC>
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3(2), 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Barton, K. (2009). *MuMIn: Multi-model inference*. Retrieved from <https://cran.r-project.org/package=MuMIn>
- Bayraktarov, E., Ehmke, G., O'Connor, J., Burns, E. L., Nguyen, H. A., McRae, L., ... Lindenmayer, D. B. (2019). Do big unstructured biodiversity data mean more knowledge? *Frontiers in Ecology and Evolution*, 6, 239. <https://doi.org/10.3389/fevo.2018.00239>
- Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., ... Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173, 144–154. <https://doi.org/10.1016/j.biocon.2013.07.037>
- Bonney, R., Cooper, C., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K., & Shirk, J. (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59(11), 977–984. <https://doi.org/10.1525/bio.2009.59.11.9>
- 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(7), 1667–1678. <https://doi.org/10.1111/2041-210X.13012>
- Brotans, L., Mönkkönen, M., Huhta, E., Nikula, A., & Rajasärkkä, A. (2003). Effects of landscape structure and forest reserve location on old-growth forest bird species in Northern Finland. *Landscape Ecology*, 18, 377–393.
- Brown, E. D., & Williams, B. K. (2018). The potential for citizen science to produce reliable and useful information in ecology. *Conservation Biology*, 33(3), 561–569. <https://doi.org/10.1111/cobi.13223>
- Callaghan, C. T., Roberts, J. D., Poore, A. G. B., Alford, R. A., Cogger, H., & Rowley, J. J. L. (2020). Citizen science data accurately predicts expert-derived species richness at a continental scale when sampling thresholds are met. *Biodiversity and Conservation*, 29(4), 1323–1337. <https://doi.org/10.1007/s10531-020-01937-3>
- Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J., & Waller, D. M. (2011). Assessing citizen science data quality: An invasive species case study. *Conservation Letters*, 4(6), 433–442. <https://doi.org/10.1111/j.1755-263X.2011.00196.x>
- Cruikshank, S. S., Bühler, C., & Schmidt, B. R. (2019). Quantifying data quality in a citizen science monitoring program: False negatives, false positives and occupancy trends. *Conservation Science and Practice*, 1(7), e54. <https://doi.org/10.1111/csp.2.54>
- Cumming, G. S. (2000). Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography*, 27(2), 441–455. <https://doi.org/10.1046/j.1365-2699.2000.00408.x>
- Delaney, D. G., Sperling, C. D., Adams, C. S., & Leung, B. (2008). Marine invasive species: Validation of citizen science and implications for national monitoring networks. *Biological Invasions*, 10(1), 117–128. <https://doi.org/10.1007/s10530-007-9114-0>
- Devictor, V., Whittaker, R. J., & Beltrame, C. (2010). Beyond scarcity: Citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions*, 16(3), 354–362. <https://doi.org/10.1111/j.1472-4642.2009.00615.x>
- Eglinton, S. M., Davis, S. E., Joys, A. C., Chamberlain, D. E., & Noble, D. G. (2010). The effect of observer experience on English Breeding Bird Survey population trends. *Bird Study*, 57(2), 129–141. <https://doi.org/10.1080/00063650903440648>
- Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Faraway, J. J. (2006). *Extending the linear model with R (Boca Raton)*. Retrieved from [https://doi.org/10.1111/j.1467-985X.2006.00446\\_8.x](https://doi.org/10.1111/j.1467-985X.2006.00446_8.x)
- Farmer, A. R. G., Leonard, M. L., & Horn, A. G. (2012). Observer effects and avian-call-count survey quality: Rare-species biases and overconfidence. *The Auk*, 129(1), 76–86. <https://doi.org/10.1525/auk.2012.11129>
- Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M., & Elkinton, J. S. (2009). Observer bias and the detection of low-density populations. *Ecological Applications*, 19(7), 1673–1679. <https://doi.org/10.1890/09-0265.1>
- Fletcher, R. J., Hefley, T. J., Robertson, E. P., Zuckerman, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, e02710. <https://doi.org/10.1002/ecy.2710>
- Franklin, J. (2013). Species distribution models in conservation biogeography: Developments and challenges. *Diversity and Distributions*, 19(10), 1217–1223. <https://doi.org/10.1111/ddi.12125>
- Frimpong, E. A., Huang, J., & Liang, Y. (2016). IchthyMaps: A database of historical distributions of freshwater fishes of the United States. *Fisheries*, 41(10), 590–599. <https://doi.org/10.1080/03632415.2016.1219948>
- Gjerde, I., Sætersdal, M., & Nilsen, T. (2005). Abundance of two threatened woodpecker species in relation to the proportion of spruce plantations in native pine forests of western Norway. *Biodiversity and Conservation*, 14(2), 377–393. <https://doi.org/10.1007/s10531-004-6065-y>
- Hochachka, W. M., Martin, K., Doyle, F., & Krebs, C. J. (2000). Monitoring vertebrate populations using observational data. *Canadian Journal of Zoology*, 78(4), 521–529. <https://doi.org/10.1139/z99-246>
- Hofstetter, L., Arlettaz, R., Bollmann, K., & Braunisch, V. (2015). Interchangeable sets of complementary habitat variables allow for flexible, site-adapted wildlife habitat management in forest ecosystems. *Basic and Applied Ecology*, 16(5), 420–433. <https://doi.org/10.1016/j.baae.2015.02.010>
- Horns, J. J., Adler, F. R., & Şekercioğlu, Ç. H. (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., & Frimpong, E. A. (2015). Using historical atlas data to develop high-resolution distribution models of freshwater fishes. *PLoS One*, 10(6), e0129995. <https://doi.org/10.1371/journal.pone.0129995>
- Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., ... O'Hara, R. B. (2020). Data integration for large-scale models of species distributions. *Trends in Ecology & Evolution*, 35(1), 56–67. <https://doi.org/10.1016/j.tree.2019.08.006>
- Isaac, N. J. B., & Pocock, M. J. O. (2015). Bias and information in biological records. *Biological Journal of the Linnean Society*, 115(3), 522–531. <https://doi.org/10.1111/bij.12532>
- Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5(10), 1052–1060. <https://doi.org/10.1111/2041-210X.12254>

- Jansson, G., & Angelstam, P. (1999). Threshold levels of habitat composition for the presence of the long-tailed tit (*Aegithalos caudatus*) in a boreal landscape. *Landscape Ecology*, 14(3), 283–290. <https://doi.org/10.1023/A:1008085902053>
- Johnson, M., Hannah, C., Acton, L., Popovici, R., Karanth, K., & Weinthal, E. (2014). Network environmentalism: Citizen scientists as agents for environmental advocacy. *Global Environmental Change-Human and Policy Dimensions*, 29, 235–245. <https://doi.org/10.1016/j.gloenvcha.2014.10.006>
- Johnston, A., Fink, D., Hochachka, W. M., & Kelling, S. (2018). Estimates of observer expertise improve species distributions from citizen science data. *Methods in Ecology and Evolution*, 9(1), 88–97. <https://doi.org/10.1111/2041-210X.12838>
- Johnston, A., Newton, S. E., Risely, K., Musgrove, A. J., Massimino, D., Baillie, S. R., & Pearce-Higgins, J. W. (2014). Species traits explain variation in detectability of UK birds. *Bird Study*, 61(3), 340–350. <https://doi.org/10.1080/00063657.2014.941787>
- Kallimanis, A. S., Panitsa, M., & Dimopoulos, P. (2017). Quality of non-expert citizen science data collected for habitat type conservation status assessment in Natura 2000 protected areas. *Scientific Reports*, 7(1), 8873. <https://doi.org/10.1038/s41598-017-09316-9>
- Kery, M., Andrew, R. J., 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(5), 1388–1397. <https://doi.org/10.1111/j.1523-1739.2010.01479.x>
- Landelius, T., Dahlgren, P., Gollvik, S., Jansson, A., & Olsson, E. (2016). A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind. *Quarterly Journal of the Royal Meteorological Society*, 142(698), 2132–2142. <https://doi.org/10.1002/qj.2813>
- Lawler, J. J., Wiersma, Y. F., & Huettmann, F. (2011). Using species distribution models for conservation planning and ecological forecasting. In: C. Drew A. Wiersma & Y. Huettmann, (Eds.). *Predictive species and habitat modeling in landscape ecology* (pp. 271–290). New York, NY: Springer. [https://doi.org/10.1007/978-1-4419-7390-0\\_14](https://doi.org/10.1007/978-1-4419-7390-0_14)
- Lindbladh, M., Felton, A., Trubins, R., & Sallnäs, O. (2011). A landscape and policy perspective on forest conversion: Long-tailed tit (*Aegithalos caudatus*) and the allocation of deciduous forests in southern Sweden. *European Journal of Forest Research*, 130(5), 861–869. <https://doi.org/10.1007/s10342-010-0478-9>
- Lobo, J. M., & Tognelli, M. F. (2011). Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation*, 19(1), 1–7. <https://doi.org/10.1016/j.jnc.2010.03.002>
- Louvrier, J., Molinari-Jobin, A., Kéry, M., Chambert, T., Miller, D., Zimmermann, F., ... Gimenez, O. (2019). Use of ambiguous detections to improve estimates from species distribution models. *Conservation Biology*, 33(1), 185–195. <https://doi.org/10.1111/cobi.13191>
- Mair, L., Harrison, P. J., Jönsson, M., Löbel, S., Nordén, J., Siitonen, J., ... Snäll, T. (2017). Evaluating citizen science data for forecasting species responses to national forest management. *Ecology and Evolution*, 7(1), 368–378. <https://doi.org/10.1002/ece3.2601>
- Mair, L., Harrison, P. J., Rätty, M., Barring, L., Strandberg, G., & Snäll, T. (2017). Forest management could counteract distribution retractions forced by climate change. *Ecological Applications*, 27(5), 1485–1497. <https://doi.org/10.1002/eap.1541>
- McPherson, J. M., Jetz, W., & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: Ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, 41(5), 811–823. <https://doi.org/10.1111/j.0021-8901.2004.00943.x>
- Millar, E. E., Hazell, E. C., & Melles, S. J. (2019). The 'cottage effect' in citizen science? Spatial bias in aquatic monitoring programs. *International Journal of Geographical Information Science*, 33(8), 1612–1632. <https://doi.org/10.1080/13658816.2018.1423686>
- Miller, D. A. W., 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. <https://doi.org/10.1111/2041-210X.13110>
- Neyens, T., Diggle, P. J., Faes, C., Beenaerts, N., Artois, T., & Giorgi, E. (2019). Mapping species richness using opportunistic samples: A case study on ground-floor bryophyte species richness in the Belgian province of Limburg. *Scientific Reports*, 9(1), 19122. <https://doi.org/10.1038/s41598-019-55593-x>
- Ottvall, R., Edenius, L., Elmberg, J., Engström, H., Holmqvist, N., Lindström, Å. K. E., ... Tjernberg, M. (2009). Population trends for Swedish breeding birds. *Ornis Svecica*, 19, 117–192.
- Pocock, M. J. O., Tweddle, J. C., Savage, J., Robinson, L. D., & Roy, E. (2017). The diversity and evolution of ecological and environmental citizen science. *PLoS One*, 12(4), 1–17. <https://doi.org/10.1371/journal.pone.0172579>
- Price, C. A., & Lee, H.-S. (2013). Changes in participants' scientific attitudes and epistemological beliefs during an astronomical citizen science project. *Journal of Research in Science Teaching*, 50(7), 773–801. <https://doi.org/10.1002/tea.21090>
- R Core Team (2019). *R: A language and environment for statistical computing*. Retrieved from <http://www.r-project.org/index.html>
- Ratnieks, F. L. W., Schrell, F., Sheppard, R. C., Brown, E., Bristow, O. E., & Garbuzov, M. (2016). Data reliability in citizen science: Learning curve and the effects of training method, volunteer background and experience on identification accuracy of insects visiting ivy flowers. *Methods in Ecology and Evolution*, 7(10), 1226–1235. <https://doi.org/10.1111/2041-210X.12581>
- Reese, H., Nilsson, M., Pahlén, T. G., Hagner, O., Joyce, S., Tingelöf, U., ... Olsson, H. (2003). Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. *AMBIO: A Journal of the Human Environment*, 32(8), 542. [https://doi.org/10.1639/0044-7447\(2003\)032\[0542:CEOFVU\]2.0.CO;2](https://doi.org/10.1639/0044-7447(2003)032[0542:CEOFVU]2.0.CO;2)
- Rosenberg, K. V., Lowe, J. D., & Dhondt, A. A. (1999). Effects of forest fragmentation on breeding tanagers: A continental perspective. *Conservation Biology*, 13(3), 568–583. <https://doi.org/10.1046/j.1523-1739.1999.98020.x>
- Ruiz-Gutierrez, V., Hooten, M. B., & Campbell Grant, E. H. (2016). Uncertainty in biological monitoring: A framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, 7(8), 900–909. <https://doi.org/10.1111/2041-210X.12542>
- Sardà-Palamera, F., Brotons, L., Villero, D., Sierdsema, H., Newson, S. E., & Jiguet, F. (2012). Mapping from heterogeneous biodiversity monitoring data sources. *Biodiversity and Conservation*, 21(11), 2927–2948. <https://doi.org/10.1007/s10531-012-0347-6>
- Schmeller, D. S., Henry, P.-Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., ... Henle, K. (2009). Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology*, 23(2), 307–316. <https://doi.org/10.1111/j.1523-1739.2008.01125.x>
- Snäll, T., Forslund, P., Jeppsson, T., Lindhe, A., & O'Hara, R. B. (2014). Evaluating temporal variation in Citizen Science Data against temporal variation in the environment. *Ecography*, 37(3), 293–300. <https://doi.org/10.1111/j.1600-0587.2011.00544.x>
- Snäll, T., Kindvall, O., Nilsson, J., & Pärt, T. (2011). Evaluating citizen-based presence data for bird monitoring. *Biological Conservation*, 144(2), 801–810. <https://doi.org/10.1016/j.biocon.2010.11.010>
- Steen, V. A., Elphick, C. S., & Tingley, M. W. (2019). An evaluation of stringent filtering to improve species distribution models from citizen science data. *Diversity and Distributions*, 25(12), 1857–1869. <https://doi.org/10.1111/ddi.12985>
- Steger, C., Butt, B., & Hooten, M. B. (2017). Safari Science: Assessing the reliability of citizen science data for wildlife

- surveys. *Journal of Applied Ecology*, 54(6), 2053–2062. <https://doi.org/10.1111/1365-2664.12921>
- Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282–2292. <https://doi.org/10.1016/j.biocon.2009.05.006>
- Sumner, S., Bevan, P., Hart, A. G., & Isaac, N. J. B. (2019). Mapping species distributions in 2 weeks using citizen science. *Insect Conservation and Diversity*, 12, 382–388. <https://doi.org/10.1111/icad.12345>
- Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, 30(3), 520–531. <https://doi.org/10.1111/cobi.12695>
- Swenson, J. E., & Angelstam, P. (1993). Habitat separation by sympatric forest grouse in Fennoscandia in relation to boreal forest succession. *Canadian Journal of Zoology*, 71(7), 1303–1310. <https://doi.org/10.1139/z93-180>
- Thuiller, W., Brotons, L., Araújo, M. B., & Lavorel, S. (2004). Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, 27(2), 165–172. <https://doi.org/10.1111/j.0906-7590.2004.03673.x>
- Tiago, P., Pereira, H. M., & Capinha, C. (2017). Using citizen science data to estimate climatic niches and species distributions. *Basic and Applied Ecology*, 20, 75–85. <https://doi.org/10.1016/j.baae.2017.04.001>
- Tulloch, A. I. T., Possingham, H. P., Joseph, L. N., Szabo, J., & Martin, T. G. (2013). Realising the full potential of citizen science monitoring programs. *Biological Conservation*, 165, 128–138. <https://doi.org/10.1016/j.biocon.2013.05.025>
- Tulloch, A. I. T., & Szabo, J. K. (2012). A behavioural ecology approach to understand volunteer surveying for citizen science datasets. *Emu - Austral Ornithology*, 112(4), 313–325. <https://doi.org/10.1071/MU12009>
- van Strien, A. J., van Swaay, C. A. M., & 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(6), 1450–1458. <https://doi.org/10.1111/1365-2664.12158>

## BIOSKETCH

The authors have expertise in modelling and evaluating citizen science data for a variety of purposes. Both institutes (ArtDatabanken and CEH) have their own bioinformatics and IT infrastructure for biodiversity citizen science data collection and analysis, involving multiple taxa at national scale (Sweden and UK). All authors are experienced and interested in testing and developing new methods, models and protocols for citizen science to produce reliable and useful information in environmental conservation, management and policy.

Author contributions: L.H. and T.S. co-conceived the idea; L.H., T.S., U.B., M.J. and N.I. defined the questions and the methodological approach; L.H. and U.B. analysed the data; L.H. wrote the manuscript; L.H., T.S., U.B., M.J. and N.I. edited the manuscript; and T.S. obtained the funding.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Henckel L, Bradter U, Jönsson M, Isaac NJB, Snäll T. Assessing the usefulness of citizen science data for habitat suitability modelling: Opportunistic reporting versus sampling based on a systematic protocol. *Divers Distrib*. 2020;26:1276–1290. <https://doi.org/10.1111/ddi.13128>