Environmental and Ecological Statistics (2021) 28:609–630
https://doi.org/10.1007/s10651-021-00508-1

Modeling spatially biased citizen science effort through the eBird database

Becky Tang[1] · **James S. Clark**[2,3] · **Alan E. Gelfand**[1]

Received: 5 October 2020 / Revised: 16 February 2021 / Accepted: 22 May 2021 / Published online: 15 June 2021

Abstract

Citizen science databases are increasing in importance as sources of ecological information, but variability in effort across locations is inherent to such data. Spatially biased data—data not sampled uniformly across the study region—is expected. A further introduction of bias is variability in the level of sampling activity across locations. This motivates our work: with a spatial dataset of visited locations and sampling activity at those locations, we propose a model-based approach for assessing effort at these locations. Adjusting for potential spatial bias both in terms of sites visited and in terms of effort is crucial for developing reliable species distribution models (SDMs). Using data from eBird, a global citizen science database dedicated to avifauna, and illustrative regions in Pennsylvania and Germany, we model spatial dependence in both the observation locations and observed activity. We employ point process models to explain the observed locations in space, fit a geostatistical model to explain observation effort at locations, and explore the potential existence of preferential sampling, i.e., dependence between the two processes. Altogether, we offer a richer notion of sampling effort, combining information about location and activity. As SDMs are often used for their predictive capabilities, an important advantage of our approach is the ability to predict effort at unobserved locations and over regions. In this way, we can accommodate misalignment between point-referenced data and say, desired areal scale density. We briefly illustrate how our proposed methods can be applied to SDMs, with demonstrated improvement in prediction from models incorporating effort.

Keywords Geostatistical model · Intensity function · Log Gaussian Cox process · Nonhomogeneous Poisson process · Preferential sampling

Handling Editor: Pierre R. L. Dutilleul

✉ Becky Tang
becky.tang@duke.edu

[1] Department of Statistical Science, Duke University, Durham, NC 27708, USA

[2] Nicholas School of the Environment, Duke University, Durham, NC 27708, USA

[3] INRAE, LESSEM, University Grenoble Alpes, Saint-Martin-d'Heres, France

1 Introduction

Informal and unstructured biological surveys of wild species may lead to biased estimates of their distributions and abundances. Biases result when observers frequent locations where they expect to observe species of interest. A further source of sampling bias, which is our primary focus and novel contribution, concerns the sampling effort at a visited site with regard to, e.g., number of visits, duration of visits. This additional effort information is anticipated to enhance our ability to learn about species distributions beyond solely the point pattern of sampled locations. In fact, we demonstrate this with a simple illustrative data example. Our approach is to specify a two-stage effort model where the first stage models the point pattern of visited sites and the second stage overlays a geostatistical model for effort on these sites.

The first stage bias, that is, where sites are visited, results from differential accessibility of sampling regions; species that live within established habitats or are easily observed are often preferentially sampled. Ecologists attempt to control for biased sampling in a number of ways, imposing corrections that range from offsets in generalized linear models (GLMs) to "pseudo-absence" points intended to improve spatial coverage (see, for example, Szabo et al. 2010; Zaniewski et al. 2002). These solutions are fixed, in the sense that they are imposed by the analyst, rather than estimated as part of model fitting. Growing awareness of the modeling complications introduced by biased observations suggests a growing interest in more formal modeling approaches. These include geographically weighted kernels (Comber et al. 2013) and ensemble models (Fink et al. 2020), but the choice of model is critical (see Bird et al. 2014 for further examples). For example, Isaac et al. (2014) test eleven methods that have been proposed or used to address variation in observer activity. These methods include some notion of filtering and/or correction to adjust for the variation, such the offsets in GLMs mentioned above. The authors found that the performance of most methods deteriorated in scenarios where data was simulated with biased site selection.

To provide some context with regard to sampling effort, we first note that models for species distribution and abundance have become the primary tool for predictions of biodiversity losses (e.g., Rosenberg et al. 2019; Langham et al. 2015; Kearney et al. 2010). Global climate change and macroecology studies often take advantage of citizen science data that were collected in a structured manner, hoping that structure can avoid sampling bias. However, even studies using data obtained from structured sampling schemes rely on the assumption that the sampled locations are representative of the region. For example, the North American Breeding Bird Survey (BBS) employs stratified random sampling. BBS observations take place along roads, meaning that geographic locations without roads are left unsampled, and residential landscapes are oversampled (Dickinson et al. 2010). Additionally, bias in estimates using BBS data may arise due to spatial bias in land cover type or temporal bias in land cover change (Harris and Haskell 2007; Niemuth et al. 2007). Examining another impact of BBS roadside surveys, Griffith et al. (2010) found that as the number of vehicle counts on a route increases, observed bird counts tended to decrease.

In contrast, opportunistic citizen science data tend to have less structure (see Kelling et al. 2019 for examples of various levels of structure in citizen science surveys), while at the same time offering broader coverage. For example, BBS restricts its observa-

tions to pre-specified routes that are visited only once during the breeding season, with observations occurring in the morning. In contrast, eBird (eBird 2017; Sullivan et al. 2014), a global citizen science project for tracking avifauna with less structure than BBS, allows participants to choose the location and frequency of their observations. Increasing the geographic and temporal coverage results in a richer dataset overall. Rather than ignore the wealth of opportunistic data, we propose methods to address spatial bias so that studies using such data can address the variation in the data collection process apart from the variation in the biological process.

We develop a spatial model for sampling effort in the context of eBird data collection. Again, this data collection is unstructured; participants choose the location, duration, and frequency of their observations, and they record the observed species in the form of a checklist. eBird allows for geographic flexibility and scope at the expense of a balanced coverage of geographic space and habitat type (Isaac and Pocock 2015) identify four key areas for bias introduced by volunteer sampling, including uneven sampling in (1) time, (2) space, (3) effort per visit, and (4) uneven detectability, which can vary by observer and by species.

Possible biases in eBird data include choice of conveniently located sites, such as those close to home or roads (Tiago et al. 2017; Mair and Ruete 2016), or areas with expected or known biodiversity or rare species (Dennis and Thomas 2000; Booth et al. 2011). Additionally, users may prefer conducting observations on certain days of the week or times of the day (Courter et al. 2013). Because eBird records the numbers of visits to locations as well as the date, start time, and duration of each visit, it supplies insight into activity, hence *effort* at locations. It is anticipated that both locations and the effort at these locations will reveal spatial bias, and our objective is to formally model these biases.

Therefore, our approach quantifies the nature of spatial bias data collection, both in locations visited and in the activity level at the visited locations, thus *explicitly* addressing elements (2) and (3) (Isaac and Pocock 2015). The novelty of this approach combines location bias with location-based effort. All visited locations are not "equal" with regard to sampling effort, and spatial bias in data collected over a region will evidently affect inference regarding abundance in that region (Dennis et al. 1999). We assert that when sampling bias exists in the data, a formal method of incorporating sampling effort is needed. The analysis reported here is the first step in the context of abundance modeling to account for such spatial bias. Though our focus is on modeling effort, we provide a simple illustrative example of how modeling the bias in effort can lead to increased performance of SDMs.

We briefly review the variety of methods to account for bias in citizen science sampling effort. The simplest approach is the inclusion of spatially-referenced covariates. Other approaches choose features of the observation as a direct proxy for sampling effort, such as density of observations (Callaghan et al. 2017; Oliveira et al. 2017; Jeppsson et al. 2010) or number of species recorded (Szabo et al. 2011); use functions of such features to develop a notion of effort such as ignorance scores (Ruete 2015) or develop schemes to weight observations (Hill 2012; Johnston et al. 2020). In the case of presence-absence data, occupancy-detection models (MacKenzie et al. 2017) are common approaches for addressing biased effort. In these models, the estimated probability of detecting a species at a site, conditional on its occupancy, is used to account

for biased effort (Kery et al. 2010; van Strien et al. 2013). Other proposals include spatial filtering, where SDMs are fitted using a subset of the data in order to remove records from highly-sampled regions (Boria et al. 2014; Beck et al. 2014; Robinson et al. 2018), and mixed effects models where random effects are used to account for variability across location, observers, and/or species (Brunsdon and Comber 2012; Roy et al. 2012). With the exception of occupancy-detection models, which are not applicable to opportunistic data, these approaches do not attempt to explicitly model the sampling effort. Instead, bias is accounted for by directly plugging in features of the observations or by modifying the data.

The above demonstrates that the notion of sampling "effort" is not always rigorously specified in the literature. In particular, there is almost always bias in the locations where sampling is performed, and this may be interpreted as bias in sampling effort. The collection of opportunistic data is especially prone to bias in sampling effort because ecologists, with limited time and resources, tend to go where they expect to find things.

Thus the effort that we focus on is effort at a location that has been visited. How often has the site been visited and how long was each visit at the site? We view this information as a mark associated with a site and hence view our data as a marked point pattern. Our approach to modeling this marked point pattern is to first model the point pattern using either a nonhomogeneous Poisson process (NHPP) or a log Gaussian Cox process (Møller et al. 1998). Then, we model the marks given the locations using a standard geostatistical model (Banerjee et al. 2014).

To offer the geostatistical model, we reduce the number and duration of visits to a scalar effort as the mark at the location. We offer several choices to do this since we cannot assert that there is a "correct" summary of effort at a site. In practice, we would insert a choice of effort into an explanatory model for species distribution or abundance and select the summary that provides the most significant explanation.

Because there will typically be misalignment between sampled sites and the species distribution or abundance/density that we seek to learn about, we can use the geostatistical model to interpolate potential effort (i.e., the amount of effort would we expect if a new location was visited) across a study region. The ability to predict potential effort at unobserved locations is important in the context of SDMs. Viewing the entire process as a marked point pattern connects us to the issue of preferential sampling (Diggle et al. 2010), as we amplify below. All models are fitted within a Bayesian framework using Markov chain Monte Carlo.

We note that the spatial scale can be important for this analysis. Many previous approaches are implemented at an areal scale, where the region is discretized into sampling units which may be modeled using spatially autocorrelated random effects (Conn et al. 2017). Because citizen science data most often arise as point-referenced locations, aggregation into areal units may lead to loss of information. Moreover, different choices of grid cell size may lead to different conclusions. Therefore, despite some recent work in modeling effort as point patterns (Geldmann et al. 2016; Sicacha-Parada et al. 2020), there remains need to develop models at point level.

Finally, in order to address model comparison, we acknowledge that while the point pattern of locations is observed, the effort variable is *constructed*. By constructed, we mean that the variable is a function or summary statistic of the observable data. Effort

may not be directly observed, but it can be defined by the modeler through a collection of observed information. So, we can only validate out-of-sample recovery of effort within a choice of effort metric. Assessment of which metric is preferred as an effort-driven covariate must be deferred to performance in an abundance model.

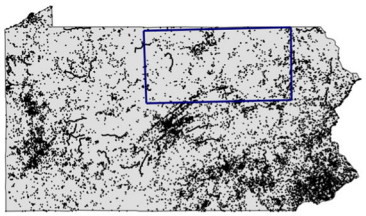
The paper has the following format. In Sect. 2 we describe the data we use with eBird as well as associated spatial regressors. In Sect. 3 we formalize our effort metrics as well as the modeling details. Section 4 takes up model fitting and checking. Section 5 provides results for the data from Sect. 2. Section 6 concludes with some discussion.

2 The data we use

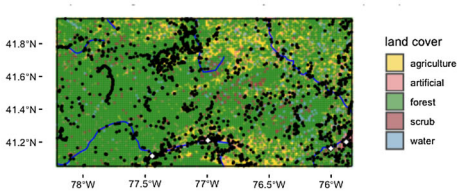
2.1 eBird data

The dataset we use in this analysis is abstracted from the eBird citizen science project where users report the type and number of bird species detected as a checklist. Users may submit "complete" checklists, where they report every species identified on the birding observation. Alternatively, users may submit "incomplete" checklists, which intentionally omit one or more species . Each checklist contains the longitude-latitude coordinate of the starting location, date, unique observer ID, and unique checklist ID. Optional information includes the duration and start time of the observation. We consider unique complete checklists only; group observations only count as one checklist. The data is filtered: checklists with NA for length of duration and/or time the observation began were removed. Additionally, checklists obtained from a long duration may be different from the typical eBirder's search. Therefore, checklists that recorded a duration length longer than 18 hours were removed. All analyses were conducted using R (Version 3.6.1) (R Core Team 2013).

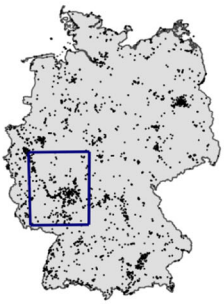
For illustration, we focus on birding activity in Pennsylvania and Germany in 2018. Pennsylvania has recorded the fifth most eBird checklists in the U.S. We choose a second region for comparison of how the drivers of observation location and activity level may differ by locality. Germany has recorded the fifth most checklists in Europe and has similar climate to Pennsylvania. Additionally, in 2018 the two regions had a similar number of observed species: 334 in Pennsylvania and 355 in Germany. Our analysis considers two subregions: 1) a 18525 km^2 region of northern Pennsylvania and 2) a 41625 km^2 region of eastern Germany. After filtering the datasets according to our criteria, which resulted in the removal of about 0.8% of observations in both regions, we have 10713 complete checklists conducted at 1910 unique locations in the Pennsylvania subregion, and 1683 complete checklists conducted at 684 unique locations in the Germany subregion. Figure 1 displays the unique birding locations in Pennsylvania and Germany in 2018, along with the observed point pattern plotted over land cover types.



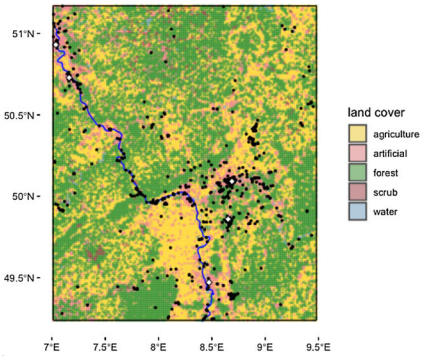
(a) Unique eBird location in Pennsylvania (2018)



(b) Unique birding location in Pennsylvania subregion (2018)



(c) Unique birding location in Germany (2018)



(d) Unique birding location in Germany subregion (2018)

Fig. 1 For Pennsylvania (top) and Germany (bottom), point patterns of unique eBird observation locations in entire region and subregions of interest overlaid on a map of land cover types with cities denoted by white diamonds. Pennsylvania land cover is obtained from the National Land Cover Database 2016, and Germany land cover is obtained from the CORINE Land Cover 2018 dataset

2.2 Spatial covariates

Associated with each unique location are the continuous covariates elevation (Shuttle Radar Topography Mission), and distances to the nearest road and nearest city (Pennsylvania: Pennsylvania Spatial Data Access, Germany: Geofabrik); binary covariates recording protected status (Pennsylvania: PAD-US, Germany: CDDA) and urban status (Pennsylvania: Protected Areas Database of the United States, Germany: European Environment Agency); and a categorical covariate for land cover type (Pennsylvania: National Land Cover Database 2016, Germany: CORINE Land Cover 2018). For each of the land cover source datasets, land cover type was aggregated into into five classes: agriculture, artificial, forest, scrub, or water (Online Resource Table S1 for aggregations). Distances to nearest road and nearest city were calculated using ArcGIS.

From Online Resource Fig. S1, we see that observations in Pennsylvania are closer to roads and cities when compared to observations conducted in Germany. The most popular sampled land cover type is artificial, comprising 0.471 and 0.401 of observations in Pennsylvania and Germany, respectively. Fewer observations come from agricultural (0.210, 0.269), forest (0.223, 0.151), and water (0.081, 0.151) areas. Areas classified as scrub land cover experienced the least observations (0.016, 0.029).

A higher proportion of German observations occurred in urban and protected areas (0.635, 0.459) as compared to observations in Pennsylvania (0.128, 0.207).

In modeling activity level, we consider what might explain the observer's behavior once they have chosen a location. For each site we know the proportion of visits that occurred on a weekend, along with the proportion of visits that began before noon. Upon plotting the proportions on a map, we do not see strong spatial patterns (Online Resource Fig. S2). In the Pennsylvania subregion, 0.476 of uniquely visited locations had all of their visits occurring in the morning and 0.320 with no morning visits, whereas these proportions are 0.323 and 0.528 for locations in the Germany subregion, respectively.

3 Modeling details

Here, we detail the modeling ideas for developing effort and potential effort-driven regressors to install in regression models for species abundance. The data consist of a point pattern of n unique locations, $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ in a region D, along with a random number of visits, $N(\mathbf{s}_i)$ at site $\mathbf{s}_i$ with associated durations $t_j(\mathbf{s}_i)$, $j = 1, 2, \ldots, N(\mathbf{s}_i)$. In addition, we have site-level regressors. As we clarify below, some are appropriate to explain the point pattern of sites and some are appropriate to explain the effort at the site.

3.1 Modeling the spatial point pattern

We consider two point process models to explain the spatial pattern of the eBird observation locations, $\mathcal{S}$ in D: a nonhomogeneous Poisson process (NHPP) and a log Gaussian Cox process (LGCP) (Møller et al. 1998). Working with Poisson processes implies that the number of locations that occur in a region D is drawn according to a Poisson distribution with intensity $\lambda(\mathbf{s})$. The points are located, conditionally independent, using the location density $f(\mathbf{s}) = \lambda(\mathbf{s})/\lambda(D)$.

In the NHPP, we assume the intensity takes the form $\log \lambda(\mathbf{s}; \theta) = \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}$. That is, the intensity on the log scale is modeled as a regression using a set of spatially-referenced covariates $\mathbf{X}(\mathbf{s})$. In the LGCP, we introduce a Gaussian process random effect $z(\mathbf{s})$ into the log intensity: $\log \lambda(\mathbf{s}; \theta) = \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta} + z(\mathbf{s})$. Normal random variables provide the customary model specification for random effects, which can yield extra flexibility for local adjustments in addition to what the covariates may be able to explain. With spatially dependent random effects, this leads to a Gaussian process specification. Under the LGCP, the log intensity is a realization of a stochastic process. Regardless, with an observed point pattern $\mathcal{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n)$, the likelihood takes the form

$$\mathcal{L}(\theta; \mathcal{S}) = \exp\{-\lambda(D; \theta)\} \prod_{i=1}^{n} \lambda(\mathbf{s}_i; \theta) = \exp\left\{-\int_D \lambda(\mathbf{s}; \theta) d\mathbf{s}\right\} \prod_{i=1}^{n} \lambda(\mathbf{s}_i; \theta). \quad (1)$$

For the NHPP, the integral in (1) is evaluated numerically. For the LGCP, the integral is stochastic; we evaluate it using regularly spaced grid points $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_J\}$ over

D, along with the observed points $\mathcal{S}$. Therefore in order to evaluate the likelihood, we require realizations of the Gaussian process at both the observed locations $\mathcal{S}$ but also at the grid points $\mathcal{U}$. In practice, these two sets of points are disjoint, so calculating the likelihood and approximating $\lambda(D)$ requires evaluating an $n + J$-dimensional normal density. To fit the model would require matrix inversion and determinant calculation of the high-dimensional covariance matrix, which becomes infeasible with large $n + J$ (as in our application). Therefore, we use a nearest neighbor Gaussian process (NNGP) as the model for $z(\mathbf{s})$ (Datta et al. 2016). Details are provided in Online Resource S1.

3.2 Effort metrics at an observed location

We consider specification of effort $w(\mathbf{s})$ as a function of the activity data that is available at an observed $\mathbf{s} \in D$. This effort variable is *constructed* or formulated and, in fact, is constructed in several ways. Importantly, since the scale of $w(\mathbf{s})$ will determine the scale of a regression coefficient associated with it, we should adopt an appropriate scale in its specification.

For the eBird data, the activity information we can employ at an observed point $\mathbf{s}_i$ includes the number of complete checklists at $\mathbf{s}_i$ and the durations of those visits. So, $w(\mathbf{s}_i)$ will be some function of this set of values, with potential choices below. We may also have some covariates $\mathbf{U}(\mathbf{s}_i)$ associated with $\mathbf{s}_i$; in our analysis these are the proportion of weekend visits at $\mathbf{s}_i$, the proportion of morning start times at $\mathbf{s}_i$, and land cover type at $\mathbf{s}_i$. Importantly, from a modeling perspective, we do not expect the covariates that help explain location to be identical to those that explain activity. Therefore, while the covariates $\mathbf{X}(\mathbf{s})$ used to drive $\lambda(\mathbf{s})$ maybe overlap with $\mathbf{U}(\mathbf{s})$–as is the case of the land cover covariate in the following analysis– we do not take $\mathbf{X}(\mathbf{s}) = \mathbf{U}(\mathbf{s})$.

Again, the modeler can specify effort in a variety of ways. In order to use effort as a predictor for abundance at a unique location, we must model the overall observer effort at that location. We consider a subset of the following choices:

- $w(\mathbf{s}) = N(\mathbf{s})$ or $w(\mathbf{s}) = \log N(\mathbf{s})$, total number of visits or log total number of visits.
- $w(\mathbf{s}) = T(\mathbf{s})$ or $w(\mathbf{s}) = \log T(\mathbf{s})$, total time or log total time.
- $w(\mathbf{s}) = T(\mathbf{s})/N(\mathbf{s})$, the activity per unit time.

The observed effort in the subregions under $w(\mathbf{s}) = \log T(\mathbf{s})$, $w(\mathbf{s}) = T(\mathbf{s})/N(\mathbf{s})$, and $w(\mathbf{s}) = \log N(\mathbf{s})$ are plotted in Figure 2. The spatial plots reveal clustering of high effort in both regions under $w(\mathbf{s}) = \log T(\mathbf{s})$ and $w(\mathbf{s}) = \log N(\mathbf{s})$, suggesting possible dependence between location and effort.

3.3 Modeling effort along with intensity and preferential sampling

Below, we will demonstrate the improved performance of the LGCP compared with the NHPP so here we will only consider the former in conjunction with three models for $w(\mathbf{s})$:

(i) $w(\mathbf{s}) = \mathbf{U}(\mathbf{s})^T\boldsymbol{\alpha} + \psi(\mathbf{s})$

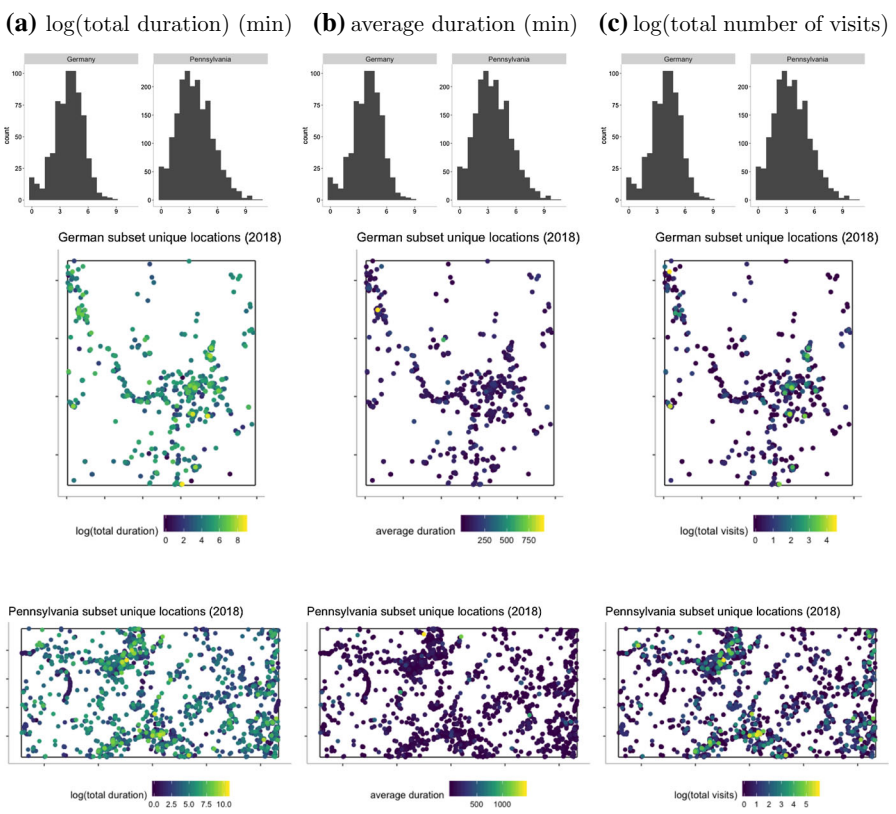


Fig. 2 Histogram of activity levels at unique locations for both subregions in Germany and Pennsylvania, followed by spatial plots

(ii) $w(\mathbf{s}) = \mathbf{U}(\mathbf{s})^T\boldsymbol{\alpha} + \delta z(\mathbf{s})$
(iii) $w(\mathbf{s}) = \mathbf{U}(\mathbf{s})^T\boldsymbol{\alpha} + \delta z(\mathbf{s}) + \psi(\mathbf{s})$

where $\mathbf{U}(\mathbf{s})$ are the coefficients explaining effort, $z(\mathbf{s})$ is the Gaussian process incorporated into the LGCP, and $\psi(\mathbf{s})$ is another Gaussian process independent of $z(\mathbf{s})$. (i) is a simple geostatistical model resulting in $\mathcal{W} \equiv \{w(\mathbf{s}_j),\ j = 1, 2, \ldots, n\}$ independent of $\mathcal{S} \equiv \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$, enabling assignment of effort $w(\mathbf{s}_0)$ at unobserved locations $\mathbf{s}_0$ given $\mathcal{S}$.

Model (ii) is a so-called "shared process" model (Diggle et al. 2010; Pati et al. 2011), i.e., $z(\mathbf{s})$ is shared between both specifications. In particular, in this formulation, $z(\mathbf{s}_i)$ becomes a significant regressor for explaining $w(\mathbf{s}_i)$ if δ is significantly different from 0. This situation is referred to as *preferential sampling*; the response $w(\mathbf{s}_i)$ is explained, in part, by the point pattern of the visited sites. The third form (iii) brings in both Gaussian processes, the one reflecting preferential sampling and the one from the purely geostatistical model. Here, we again may ask whether δ is significant? Unless the sampling bias is severe, typically the flexibility of the $\psi(\mathbf{s})$ process will annihilate the significance of the inherited $z(\mathbf{s})$ process as a regressor. In any event, the sign of δ

is interpreted in terms of *local* adjustment. That is, the sign of δ along with the sign of $z(\mathbf{s}_i)$ determines whether the expected effort is pushed up or pulled down at $\mathbf{s}_i$.

A crucial remark here concerns kriging effort to a new location. For any of the covariates that are taken from $\mathbf{X}(\mathbf{s})$, e.g., land cover type, we will have regressors at $\mathbf{s}_0$ for prediction of $w(\mathbf{s}_0)$. However, we will not have certain regressors in $\mathbf{U}(\mathbf{s}_0)$, such as the percent of weekend visits. The solution we adopt here is to set those regressor in $\mathbf{U}(\mathbf{s}_0)$ to their mean levels while employing the known $\mathbf{X}(\mathbf{s}_0)$, along with the contributions from the Gaussian processes, to obtain a posterior predictive distribution for $w(\mathbf{s}_0)$.

3.4 Integration over subregions

We have noted that a common challenge which arises with species distribution modeling is a misalignment between data collection and covariates; see, for example Agarwal et al. (2002) or Lee and Sarran (2015). We may have abundances at areal scales with intensity and effort at point level. To scale from point level to areal level, we can obtain an "intensity weighted" effort of the form $\lambda(\mathbf{s})w(\mathbf{s})$ at a point $\mathbf{s}$, that is, an effort which is up- or down-weighted according to local intensity, and then integrated over subregion B to obtain

$$\frac{1}{|B|}\int_B \lambda(\mathbf{s})w(\mathbf{s})d\mathbf{s}, \tag{2}$$

the intensity-weighted effort per unit area in B.

We can consider the simpler forms, $\frac{1}{|B|}\int_B w(\mathbf{s})d\mathbf{s}$ and $\int_B \lambda(\mathbf{s})d\mathbf{s}$, the effort per unit area over subregion B and the cumulative intensity over subregion B, respectively. According to our specifications for $w(\mathbf{s})$ and $\lambda(\mathbf{s})$, all are stochastic integrals and are computed using realizations of the needed stochastic processes over a suitable fine grid of points.

4 Model fitting, checking, and inference

4.1 Model fitting for the NHPP and LGCP

The log intensity of the NHPP is specified as $\log\lambda(\mathbf{s}) = \mathbf{X}(\mathbf{s})^T\boldsymbol{\beta}$. The posterior for $\boldsymbol{\beta}$ becomes $p(\boldsymbol{\beta}|\mathcal{S}) \propto \mathcal{L}(\boldsymbol{\beta};\mathcal{S}) \times \pi(\boldsymbol{\beta})$ with the likelihood taken from (1) and $\pi(\boldsymbol{\beta})$ a diffuse normal prior. Fitting the model using Markov Chain Monte Carlo (MCMC), we take Metropolis steps to update the $\boldsymbol{\beta}$ parameters.

The LGCP model specifies $\log\lambda(\mathbf{s}) = \mathbf{X}(\mathbf{s})^T\boldsymbol{\beta} + z(\mathbf{s})$ where the stochastic integral $\lambda(D)$ in (1) is approximated via quadrature with quadrature points as the centers of a 5km lattice over the region (Online Resource S3). We model the spatial effects $z(\mathbf{s})$ with an NNGP arising from a valid parent Gaussian process with an exponential covariance function, $C(\mathbf{s}, \mathbf{s}+h) = \sigma^2_{\mathbf{z}}e^{-\phi||h||}$. Because we are modeling the intensity on the log scale, we assume the parent Gaussian process is centered at $-\sigma^2_{\mathbf{z}}/2$ so that $E[e^{z(\mathbf{s})}] = 1$. We place an Inverse Gamma(1,1) prior on $\sigma^2_{\mathbf{z}}$, whose infinite variance yields a highly uninformative prior.

Due to identifiability issues associated with $\sigma^2_{\mathbf{z}}$ and ϕ (Zhang 2004), we adopt an empirical Bayes approach and fix ϕ at 1km and 3km for Germany and Pennsylvania, respectively. Experimentation with other values revealed that inference is not sensitive to values of ϕ near these values. Another potential identfiability issue concerning $\sigma^2_{\mathbf{z}}$ may arise: as the Gaussian process is centered at $-\sigma^2_{\mathbf{z}}/2$, it may be difficult to separate $\sigma^2_{\mathbf{z}}$ from the intercept in $\boldsymbol{\beta}$. One way to address this issue is to reparameterize the Gaussian process such, introducing a $z'(\mathbf{s}) = z(\mathbf{s}) + \sigma^2_{\mathbf{z}}/2$, which will result in a Gaussian process with zero mean and the same covariance function. Brief details for the NNGP are provided in Online Resource S1. The form of the joint posterior distribution is shown in Online Resource S2.

4.2 Fitting the shared process models

For modeling effort $w(\mathbf{s})$, in case (i) we have a simple geostatistical model of the form $w(\mathbf{s}) = \mathbf{U}(\mathbf{s})^T\boldsymbol{\alpha} + \psi(\mathbf{s})$, where $\psi(\mathbf{s})$ is a Gaussian process with an exponential covariance function and spatial decay parameters fixed at the same values as above, and adopt a weak inverse Gamma prior for σ^2_{ψ}. For the regression coefficients $\boldsymbol{\alpha}$, we assign independent diffuse normal priors. With the absence of quadrature points, we fit $\psi(\mathbf{s})$ with an elliptical slice sampler (Murray et al. 2010). Due to conjugacy, posterior updates for $\boldsymbol{\alpha}$ and σ^2_{ψ} are drawn from multivariate normal and Inverse Gamma full conditionals, respectively.

Turning to model (ii), the shared process model links the point process model with the model for activity level. More specifically,

$$\begin{aligned} \log\lambda(\mathbf{s}_i) &= \mathbf{X}(\mathbf{s}_i)^T\boldsymbol{\beta} + z(\mathbf{s}_i) && \text{locations} \\ \log w(\mathbf{s}_i) &= \mathbf{U}(\mathbf{s}_i)^T\boldsymbol{\alpha} + \delta z(\mathbf{s}_i) && \text{effort.} \end{aligned}$$

Again, we model the $z(\mathbf{s}_i)$ with an NNGP. Due to the dependence between the two models, the updates for $z(\mathbf{s}_i)$ are modified slightly but are still normally distributed. The updates of the spatial effects for the quadrature points remain unchanged due to the fact that we chose the reference set for the NNGP to be $\mathcal{S}$, the set of observed locations. We now assign δ a diffuse normal prior. The parameters of interest in the joint posterior distribution are $\boldsymbol{\beta}, \mathbf{z}, \sigma^2_{\mathbf{z}}, \boldsymbol{\alpha}, \delta$. As when fitting the LGCP independently, we use random walk Metropolis steps to update σ^2 and each of the $z(\mathbf{s}_i)$'s. Posterior updates for $\boldsymbol{\alpha}$ and δ are conjugate multivariate and univariate normals, respectively.

Lastly, for model (iii) we now have two Gaussian processes along with a preferential sampling term:

$$\begin{aligned} \log\lambda(\mathbf{s}_i) &= \mathbf{X}(\mathbf{s}_i)^T\boldsymbol{\beta} + z(\mathbf{s}_i) && \text{locations} \\ \log w(\mathbf{s}_i) &= \mathbf{U}(\mathbf{s}_i)^T\boldsymbol{\alpha} + \delta z(\mathbf{s}_i) + \psi(\mathbf{s}_i) && \text{effort.} \end{aligned}$$

Updates for the parameters and Gaussian processes are similar to the above, appropriately modified to include the additional terms in the model for $\log w(\mathbf{s}_i)$.

4.3 Model assessment and comparison

First we take up model assessment and comparison between the NHPP and the LGCP with regard to the observation locations. To compare the two models, we employ cross validation by separating the data into a training set and a testing set. Following (Leininger and Gelfand 2017), we use $p-$thinning (Illian et al. 2008) to create the test set $\mathcal{Y}^{\text{test}}$ for prediction by removing each point $\mathbf{s}_i \in \mathcal{S}$ with probability $(1-p)$, with the remaining points forming the training set $\mathcal{Y}^{\text{train}}$. $\mathcal{Y}^{\text{train}}$ is roughly p percent of the original dataset, with intensity $\lambda^{\text{train}}(\mathbf{s}) = p\lambda(\mathbf{s})$. $\mathcal{Y}^{\text{test}}$ has resulting intensity $\lambda^{\text{test}}(\mathbf{s}) = (1-p)\lambda(\mathbf{s})$, and the two sets are independent conditional on $\lambda(\mathbf{s})$. To obtain posterior predictions for $\mathcal{Y}^{\text{test}}$, we can convert the posterior draws of the training intensity via $\lambda^{\text{test}}(\mathbf{s}) = \left(\frac{1-p}{p}\right)\lambda^{\text{train}}(\mathbf{s})$. Using these intensities, we obtain $N^{\text{pred}}(A)$, the posterior predictive distribution for $N^{\text{test}}(A)$, the number of points within a region $A \subset D$. We do this by integrating $\lambda^{\text{test}}(\mathbf{s})$ over A.

In order to assess the performance of the models, we obtain predictive residuals within A as $R_{\text{pred}}(A) = N^{\text{test}}(A) - N^{\text{pred}}(A)$, as suggested by Leininger and Gelfand (2017). A valid model would expect residuals close to 0. We can do this using several sizes for A, partitioning D into $A_1, \ldots, A_K$ for some number K. Thus, we evaluate the residuals on the A_k and obtain the posterior predictive empirical coverage of 0. If the model is suitable, the empirical coverage and the nominal coverage are expected to be similar.

Then, to compare the NHPP and LGCP models, we use the rank probability score (RPS; Gneiting and Raftery 2007) which compares the entire predictive count distribution to the observed held-out count. We obtain the predictive distribution for $N^{\text{test}}(A)$ from both models and compare them to the observed number of points in region A of the test set, $N^{\text{test}}(A)$. Distributions more concentrated around the observation are preferred and lead to smaller RPS, providing our criterion. We then compare the posterior predicted number of points to the observed, held-out number of points by computing the RPS, averaged across the A_k if partitioning D.

Next, we turn to model comparison for effort at the visited sites. We reiterate that there is no meaningful comparison to be made among different metrics; comparison can only be made within one definition of activity level. Using training and test sets obtained as described above, we can predict activity at the held-out locations, $w(\mathbf{s}_0)$ via kriging. We compute root mean squared error (RMSE), comparing the predicted activity to the observed.

5 Results

In this section, we first implement the NHPP and LGCP-NNGP models to analyze the locations of observers. Then bringing in the notion of effort, we consider models (i) and (ii) for the following specifications for effort $w(\mathbf{s})$:

(a) $w(\mathbf{s}) = \log T(\mathbf{s})$
(b) $w(\mathbf{s}) = T(\mathbf{s}_i)/N(\mathbf{s})$

(c) $w(\mathbf{s}) = \log N(\mathbf{s})$

Due to the shared process in (ii), inference for the point process will also be affected. Therefore, we also obtain estimates for the LGCP-NNGP (from here on, simply LGCP) coefficients under (ii) for (a)-(c), followed by estimates for effort $w(\mathbf{s})$ under (i) and (ii) for (a)-(c) for both regions.

5.1 Point pattern intensities

Table 1 displays the posterior means and 95% credible intervals for the regression coefficients for both the NHPP and LGCP fit on the Pennsylvania point pattern. We see that the models mostly agree on significance of coefficients. Because the introduction of the Gaussian process provides local spatial adjustment to the intensity, i.e., less spatial smoothing, the LGCP may lead to differences in the significance and/or sign of coefficients when compared to the NHPP. The negative coefficient for elevation signifies that few people are likely to travel to areas at high elevations. This, along with the negative coefficient for distance to nearest road, indicates a notion of convenience, as people may not be as inclined to travel far or high to go birding. Under the LGCP, Pennsylvania eBirders are more likely to begin their observations in artificial or water areas than forests, but are less likely to visit areas characterized as urban. Histograms of the posterior densities of the intercepts and spatial variance terms for the LGCP reveal that posteriors are not sensitive to the choice of priors (Online Resource Fig. S3).

To assess and compare the NHPP and the LGCP fit using all the covariates, we follow the p-thinning approach with $p = 0.5$. For model assessment, we partition D into 225 fixed area subsets and obtain 90% prediction intervals for the predictive residuals. We do this for both the training data as well as the test data. For the NHPP, the empirical predictive coverage for the train and test data are 74% and 75%, respectively. For the LGCP, these empirical coverages are 89% and 86%. The LGCP achieves an empirical coverage very close to the nominal level of 90%, whereas the NHPP exhibits undercoverage.

For both models, we compare the posterior predicted number of points to the true, held-out number of points $N^{\text{test}}(D)$ in the entire Pennsylvania subregion. We then perform this comparison for subsets of D by partitioning D into 9, 25, 49, and 225 fixed area subsets and computing the RPS averaged across the nine subsets for both models. Comparing the averaged RPS across the various partitioning schemes, Table 2 (first two rows) demonstrates that the LGCP outperforms the NHPP in terms of predictive performance. Therefore, for the remainder of the analysis we choose to use the LGCP to model the locations of observations when fitting the effort models (i) and (ii).

Turning to (ii), the shared process model creates a dependence between observer location and activity level. Therefore estimates of coefficients will be affected by both the activity level as well as the metric. Online Resource Table S2 displays posterior mean and 95% credible intervals for models (ii) (a)-(c). In fact, all models for the locations agree on significance of coefficients, with the exception of agricultural land

Table 1 Posterior mean and 95% credible intervals for the NHPP and LGCP-NNGP models fit on the Pennsylvania and Germany subregion point patterns to address bias in location. Land cover types are with respect to the baseline cover type of forest

	Pennsylvania NHPP			LGCP-NNGP			Germany LGCP-NNGP		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
Intercept	−2.4315	−2.5563	−2.3335	−2.6912	−2.8138	−2.5858	−5.1101	−5.3205	−4.9179
Elevation	−0.1188	−0.1865	−0.0537	−0.0976	−0.1696	−0.0314	−0.0723	−0.1503	0.0092
Agriculture	0.2490	0.1378	0.3701	0.1431	−0.0111	0.3037	0.3373	0.0757	0.5810
Artificial	−0.3653	−0.5084	−0.2098	0.3733	0.2466	0.5035	1.8011	0.8173	1.3475
Scrub	−0.0735	−0.4428	0.2678	0.1997	−0.1629	0.5462	1.0422	0.5002	1.5290
Water	0.1841	0.0003	0.3737	0.3030	0.1102	0.4970	1.3151	1.0064	1.6095
Protected	0.0546	−0.0002	0.1056	0.0386	−0.0230	0.0950	0.3537	0.2798	0.4306
Urban	−0.2727	−0.3321	−0.2182	−0.1985	−0.2644	−0.1432	0.6596	0.5744	0.7516
Road_dist	−0.2050	−0.2848	−0.1346	−0.1054	−0.1795	−0.0323	−0.4905	−0.6445	−0.3624
City_dist	−0.0546	−0.1084	0.0027	−0.0118	−0.0749	0.0542	−0.2505	−0.3485	−0.1633

Table 2 Average RPS for subsets of a fixed area in $\mathcal{Y}^{\text{test}}$ within the Pennsylvania subregion, constructed via p-thinning with $p = 0.5$

	18525 km^2	2060 km^2	740 km^2	378 km^2	82 km^2
NHPP	10.433	25.828	12.476	9.129	2.468
LGCP-NNGP	10.278	21.298	9.257	7.327	2.078
Model (ii)(a)	**10.270**	**20.173**	**9.093**	**7.150**	**2.012**
Model (ii)(b)	12.730	22.800	9.967	7.966	2.103
Model (ii)(c)	10.953	22.092	9.484	7.400	2.103

Models with lowest RPS are preferred (in bold); comparisons can only be made across models, not area sizes

cover, which is significantly positive under (ii) (a)-(c) but not significant under the LGCP.

Additionally, we note that the shared process scheme is intended to aid in estimation of activity level, not the point pattern; a measurement of activity does not exist unless an observation took place. Nevertheless, we compute the RPS under (ii) (a)-(c) using the same train and test sets used to compare the NHPP and LGCP models as above, presented in the last three rows of Table 2. Model (ii) (a) outperforms all models, followed by the LGCP.

For the German subregion, posterior means of coefficients are very similar across all four models (Online Resource Table S2). Coefficients for all land cover types are significantly positive, suggesting that once adjusting for the spatial effects and other covariates, forested areas are the least visited by German birders. Plots of the estimated point pattern intensities for the Pennsylvania subregion under the LGCP and (ii) (a) are presented in Figure 3. The plots Figure 3 reveal that the shared process model (ii) better captures areas of particularly high intensity, that is, areas where many sampling events occur, suggesting a potential benefit of fitting the shared process model.

5.2 Effort

Posterior summaries for the estimated coefficients under the various metrics for effort under models (i) and (ii) in Pennsylvania are presented in Online Resource Table S3. Focusing first on model (i), the coefficient for the proportion of visits that occurred on weekends is significantly positive for all (a)-(c). This suggests higher effort occurs on weekends, presumably when people may have more free time. The coefficient for proportion of visits beginning before noon is negative for all (a)-(c), suggesting less effort occurs in the morning in Pennsylvania. Under (a) and (b), the coefficient for artificial land cover is also negative, interpreted as shorter birding activities occur in artificial areas such as apartments and cities when compared to forested areas.

Turning to the shared process model (ii), the addition of the Gaussian process as a regressor may invariably alter significance of coefficients. As such, we are primarily interested in δ to learn how observation location may be related to activity. From Online Resource Table S3, δ for the Pennsylvania region is estimated to be significant for all (a)-(c), suggesting spatial dependence between Pennsylvania observers' locations

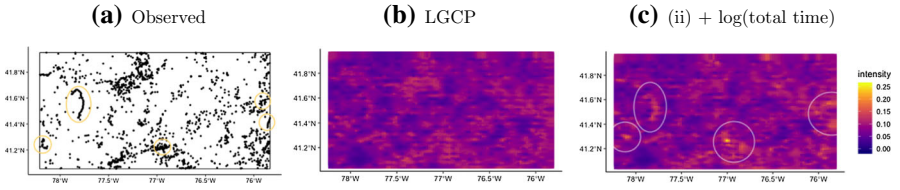


Fig. 3 **a–c** Observed point pattern along with posterior mean intensities for the region in Pennsylvania under the LGCP-only model and shared process model (ii) with effort as log(total time). Circled areas indicate where the estimated intensities under under (ii) better capture the observed areas of higher local intensity compared to the LGCP, as seen by the bright yellow. Plots of posterior mean intensities under the other formulations of effort are similar

and their effort at those locations. Online Resource Fig. S5 displays posterior mean intensities plotted against the posterior mean estimated effort under the shared process model for the two regions.

In order to compare effort under models (i) and (ii), we compute the RMSE for the predicted effort at held-out observations in Pennsylvania. Predictive RMSEs under ((i), (ii)) are (2.4354, 1.5216) under effort (a), (2.4285, 1.6064) under (b), and (5.6412, 3.3110) under (c). Thus we find that with regard to posterior prediction, the shared process model (ii) is superior to (i) for all three effort metrics. Online Resource Fig. S4 displays the observed effort surface generated from the held-out data, alongside the predicted posterior mean surfaces under (i) and (ii) for the three effort metrics. Effort surfaces under (i) have more variability and extremes compared to (ii), which is explained by the second Gaussian process providing local adjustments. In fact, surfaces under (ii) are dampened when compared to the observed effort. However, the surfaces show that predictions under (i), while perhaps closer in range to the observed values, do a poorer job of capturing the overall trends. Plots of posterior mean intensities plotted against the posterior mean estimated effort under the shared process model for the two regions are presented in Online Resource Fig. S5.

In Germany, we once again find that the shared process model removes the significance of many covariates, especially with regard to the land cover covariates (Online Resource Table S3). Interestingly, across all models and effort metrics, the coefficient for proportion of visits with morning start times is significantly positive. This is in contrast to the negative coefficient for Pennsylvania birders, and also in contrast to our exploratory analysis where we found that Germany had fewer locations with morning observations. δ is estimated to be significantly positive for (a) and (c), suggesting preferential sampling between effort and location also in Germany.

5.3 Average effort by land cover type

Here we demonstrate how to accommodate misalignment between point-referenced data collection and desired areal scale species density. Illustratively, we consider observer effort and intensity in conjunction. That is, we obtain an intensity-weighted effort per unit area in subregions that are dominated by a single land cover to learn how birding effort may differ across cover types. In both Pennsylvania and Germany

we select two subregions that are nearly uniform in artificial and agricultural land cover, respectively. Here we only consider model (ii), due to its superior prediction performance with regards to $w(\mathbf{s})$. Average effort, as proposed in 3.4, is calculated using the quadrature method, with a 1km discretization over the subregions.

In Pennsylvania under metric (a), the average intensity-weighted effort per square kilometer for artificial and agricultural land cover types are 0.3196 and 0.5129, respectively. Under (b), the values are 2.6651 and 4.4477, and under (c), 0.0710 and 0.1029. Thus under all three notions of activity, average intensity-weighted effort per square kilometer is higher in agricultural areas than artificial areas in the Pennsylvanian subregion. Looking to Germany, average intensity-weighted effort under (a) for artificial and agricultural areas are 0.5676 and 0.0982; under (b), 6.8648 and 1.8902; and under (c) 0.0817 and 0.0101. We see a relationship between these land cover types that is opposite to Pennsylvania birding effort; average birding effort in Germany is higher in artificial areas than agricultural ones.

5.4 Application: effort as a regressor

Here, we offer a simple application of effort as a regressor for modeling bird counts. We consider the observed counts for three species observed at sites in the Pennsylvania subregion from June through August 2018: dark-eyed junco (*Junco hyemalis*), mourning dove (*Zenaida macroura*), and red-winged blackbird (*Agelaius phoeniceus*). As an exploratory examination, Online Resource Fig. S6 displays the counts plotted by longitude and latitude, and colored by observed effort. For the mourning dove and red-winged blackbird, large counts tend to occur at locations where a large amount of sampling effort occurred, and these locations tend to appear in close spatial proximity. Magnitude of observed counts for the dark-eyed junco do not appear to be as strongly associated with effort.

To examine the possible association of effort and counts, we split the data into train and test sets and fit five models on the train sets for each species: a Poisson regression using the environmental covariates outlined above as regressors; three Poisson regressions using the environmental covariates and effort $w(\mathbf{s})$ for the three effort metrics under the shared process model (ii); and a Poisson regression using the environmental covariates plus the covariates of proportion morning visits and proportion weekend visits as "naive" proxies for effort. We acknowledge that while a more complex model for the observed counts is warranted, here our aim is simply to present an illustration of how our work may be applied to SDMs.

For models fit with effort under (ii) as a regressor, the estimated coefficient for effort is significantly positive for all species and all specifications of $w(\mathbf{s})$. Thus, increasing effort leads to increased expected counts. As mentioned in the Introduction, SDMs are valued for their predictive abilities. We therefore explore if there are performance gains when using predicted effort, $w(\mathbf{s}_0)$, at held out locations, $\mathbf{s}_0$, to predict the species counts at $\mathbf{s}_0$. A distinct advantage of our modeling of effort emerges; it allows us to interpolate and predict effort at new locations. We compute the RPS to compare the predictive performances of the five count models. All three models which include the explicitly modeled effort as a regressor are largely superior for the three species

Table 3 Average RPS for predicting counts for three species using a model with only environmental covariates (EC), EC with the "naive effort" covariates of proportion weekend and morning visits, and EC with the explicitly modeled effort under (ii)(a)–(c)

Model	*Junco hyemalis*	*Zenaida macroura*	*Agelaius phoeniceus*
EC	0.706	6.972	16.982
EC + naive effort	0.691	6.021	21.129
EC + (ii)(a)	**0.632**	**4.756**	**5.363**
EC + (ii)(b)	0.699	6.638	12.770
EC + (ii)(c)	1.166	5.858	7.124

Models with lowest RPS are preferred (in bold)

and forms of effort, with (ii)(a) outperforming for all three species (Table 3). We find that the model without effort tends to overpredict counts with much higher frequency compared to the model containing effort.

6 Discussion

Confronting spatial bias in citizen science data is not a new problem; it is recognized that bias exists in space and effort, and both should be accounted for in adopting species distribution models. If the data representing the biological community arise from a biased sampling scheme, then depending on the goal of the analysis, inference for species distributions may not be reliable if the model does not incorporate appropriate adjustments. In particular, accounting for the bias is necessary when the outcome of interest varies due to the bias in the data. We argued that the visited eBird locations occur as a point pattern over space, fit a nonhomogeneous Poisson process and a logGaussian Cox process model within a Bayesian framework, and found that the LGCP is superior with regard to prediction of where birders go. We found that areas of lower accessibility are often less likely to be visited, and that agricultural, artificial, and water locations are often more likely to be visited than forests.

We then developed a geostatistical model for the observed effort at each location, under various definitions of effort, and learned that both time of week and day can be significant predictors of sampling effort (Table S3). Additionally, we found evidence of preferential sampling between effort and location, implying dependence in the two sources of bias. Crucially, by explicitly modeling the spatial dependence, we were able to obtain predictions for effort at other locations within the region. We then offered a simple illustration which used the predicted effort as a regressor to predict the counts of three species of birds, and demonstrated that adjusting for sampling bias can improve predictive performance.

Some of our findings align with results obtained from previous studies. In modeling the intensity of observations of Bird atlas III in Denmark, Geldmann et al. (2016) found intensities increased in areas classified as water land cover. In our own analyses, compared to a different baseline class of forest, we also found evidence of increased intensities in water areas in both Germany and Pennsylvania. Additionally, Geldmann

et al. (2016) found birding intensity increased with increasing human population density. Considering urban status as a similar covariate, we obtained similar conclusions for Germany. In their modeling of sampling effort of BirdTrack data in Great Britain, Johnston et al. (2020) found protected status was not a significant predictor, in contrast to other studies which have found a higher density of sampling in protected areas. They suggest that the discrepancy could possibly be attributed to how the characteristics of protected areas vary across countries. We find supporting evidence in our own analysis, where protected status is a positive, significant predictor for Germany birding locations but is not significant for Pennsylvania.

In fact, we found several differences in sign and significance of coefficients between the two regions. For example, urban areas are associated with higher intensity in Germany but lower intensity in Pennsylvania. Additionally, we considered how birding effort changes by land cover type, and found higher average intensity-weighted effort in agricultural areas than artificial ones in Pennsylvania, while the relationship is reversed in Germany. These findings suggest differences in the behavior of eBird observers in various geographical regions, and underscore the importance of modeling effort and including covariates in order to account for these differences. While we focused on two regions with more temperate weather and similar climates, our methods are readily extendable to analyze sampling effort in other regions.

We did not consider the temporal aspect of the data collection. In this regard, we could extend the analysis to a spatiotemporal point pattern. One would expect that time of year has an impact on the frequency of effort, due to weather or migration season. Also, we do not expect species distributions to be independent of observation effort; bias towards rare or attractive species has been found in previous studies, such as Johnston et al. (2020) and Boakes et al. (2010). Therefore, site selection and effort may change over time as species distributions change or become more well known (Lobo et al. 2007), further underscoring the potential of temporal bias in birding effort. This could be implemented in the fashion of an autoregressive model, developing effort across a sequence of time windows to see if previous effort helps to explain current effort. Doing so would address bias (1) identified by Illian et al. (2008).

Lastly, the larger goal is to develop effective SDMs. By creating a model for effort, we can introduce it as a regressor into a model for abundance in a hierarchical specification of the form [abundance|effort][effort], adding an additional shared process layer. We provided a simple demonstration of how modeled effort can be incorporated into SDMs. If we wished to model the observed counts at each site $\mathbf{s}$, then we can immediately introduce the associated effort at that site. If the objective is understanding abundance, taken say as density at areal unit scale, we can obtain integrated effort for areal units of interest as developed above in the case of subregions characterized by a land cover. The flexibility of the methods presented here allows for incorporating the bias in effort at different spatial scales.

In conclusion, citizen science data are crucial for the field of ecology, allowing for the monitoring of species diversity at large geographic scales. The nature of citizen science data, especially data collected in an opportunistic manner, can pose challenges due to the bias in sampling location and effort. As citizen science data are being applied to increasingly more complex ecological questions, developing methodology to account for such biases is essential. We hope this work encourages further model-

based approaches to capturing the biases in effort, as doing so will inevitably lead to more trustworthy models for species distributions and abundances.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10651-021-00508-1.

Acknowledgements For comments on the manuscript we thank Valentin Journe, Ruben Palacio, Renata Poulton Kamakura, Tong Qiu, Chantal Reid, C. Lane Scher, Shubhi Sharma, and Maggie Swift.

Author Contributions J.S.C. conceived of the study, and A.E.G. and J.S.C. contributed to formulation of models. B.T. obtained the data and implemented the analyses. A.E.G. and B.T. drafted the paper with contributions from J.S.C. All authors read and approved the final manuscript.

Funding The project was funded by the National Science Foundation (NSF-DEB-1754443, NSF ICER/Belmont Forum Biodiversa), NASA (AIST18-0063), and the Programme d'Investissement d'Avenir under project FORBIC (18-MPGA-0004).

Declarations

Conflicts of interest All authors declare that they have no conflict of interest.

Data availability All data used in this study are available at https://github.com/beckytang/ebird_data.

Code availability All data used in this study are available at https://github.com/beckytang/ebird_data.

References

Agarwal DK, Gelfand AE, Silander JA (2002) Investigating tropical deforestation using two-stage spatially misaligned regression models. J Agric Biol Environ Stat 7(3):420

Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data. CRC Press, Boca Raton

Beck J, Böller M, Erhardt A, Schwanghart W (2014) Spatial bias in the gbif database and its effect on modeling species' geographic distributions. Ecol Inform 19:10–15

Bird TJ, Bates AE, Lefcheck JS, Hill NA, Thomson RJ, Edgar GJ, Stuart-Smith RD, Wotherspoon S, Krkosek M, Stuart-Smith JF et al (2014) Statistical solutions for error and bias in global citizen science datasets. Biol Conserv 173:144–154

Boakes EH, McGowan PJ, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol 8(6)

Booth JE, Gaston KJ, Evans KL, Armsworth PR (2011) The value of species rarity in biodiversity recreation: a birdwatching example. Biol Conserv 144(11):2728–2732

Boria RA, Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecol Model 275:73–77

Brunsdon C, Comber L (2012) Assessing the changing flowering date of the common lilac in North America: a random coefficient model approach. Geoinformatica 16(4):675–690

Callaghan C, Lyons M, Martin J, Major R, Kingsford R (2017) Assessing the reliability of avian biodiversity measures of urban greenspaces using eBird citizen science data. Avian Conserv Ecol 12(2)

Comber A, See L, Fritz S, Van der Velde M, Perger C, Foody G (2013) Using control data to determine the reliability of volunteered geographic information about land cover. Int J Appl Earth Obs Geoinf 23:37–48

Conn PB, Thorson JT, Johnson DS (2017) Confronting preferential sampling when analysing population distributions: diagnosis and model-based triage. Methods Ecol Evol 8(11):1535–1546

Courter JR, Johnson RJ, Stuyck CM, Lang BA, Kaiser EW (2013) Weekend bias in citizen science data reporting: implications for phenology studies. Int J Biometeorol 57(5):715–720

Datta A, Banerjee S, Finley AO, Gelfand AE (2016) Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. J Am Stat Assoc 111(514):800–812
Dennis RLH, Thomas C (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. J Insect Conserv 4(2):73–77
Dennis RL, Sparks TH, Hardy PB (1999) Bias in butterfly distribution maps: the effects of sampling effort. J Insect Conserv 3(1):33–42
Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol Syst 41:147–172
Diggle PJ, Menezes R, Su T-L (2010) Geostatistical inference under preferential sampling. J R Stat Soc: Ser C (Appl Stat) 59(2):191–232
eBird (2017) eBird: an online database of bird distribution and abundance. Cornell Lab of Ornithology, Ithaca, New York
Fink D, Auer T, Johnston A, Ruiz-Gutierrez V, Hochachka WM, Kelling S (2020) Modeling avian full annual cycle distribution and population trends with citizen science data. Ecol Appl 30(3):e02056
Geldmann J, Heilmann-Clausen J, Holm TE, Levinsky I, Markussen B, Olsen K, Rahbek C, Tøttrup AP (2016) What determines spatial bias in citizen science? exploring four recording schemes with different proficiency requirements. Divers Distrib 22(11):1139–1149
Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378
Griffith EH, Sauer JR, Royle JA (2010) Traffic effects on bird counts on North American breeding bird survey routes. The Auk 127(2):387–393
Harris J, Haskell D (2007) Land cover sampling biases associated with roadside bird surveys. Avian Conserv Ecolo 2(2)
Hill MO (2012) Local frequency as a key to interpreting species occurrence data when recording effort is not known. Methods Ecol Evol 3(1):195–205
Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns, vol 70. Wiley, New York
Isaac NJB, Pocock MJO (2015) Bias and information in biological records. Biol J Linn Soc 115(3):522–531
Isaac NJ, van Strien AJ, August TA, de Zeeuw MP, Roy DB (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. Methods Ecol Evol 5(10):1052–1060
Jeppsson T, Lindhe A, Gärdenfors U, Forslund P (2010) The use of historical collections to estimate population trends: a case study using Swedish longhorn beetles (Coleoptera: Cerambycidae). Biol Conserv 143(9):1940–1950
Johnston A, Moran N, Musgrove A, Fink D, Baillie SR (2020) Estimating species distributions from spatially biased citizen science data. Ecol Model 422:108927
Kearney MR, Wintle BA, Porter WP (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conserv Lett 3(3):201–213
Kelling S, Johnston A, Bonn A, Fink D, Ruiz-Gutierrez V, Bonney R, Fernandez M, Hochachka WM, Julliard R, Kraemer R et al (2019) Using semistructured surveys to improve citizen science data for monitoring biodiversity. BioScience 69(3):170–179
Kery M, Royle JA, Schmid H, Schaub M, Volet B, Haefliger G, Zbinden N (2010) Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. Conserv Biol 24(5):1388–1397
Langham GM, Schuetz JG, Distler T, Soykan CU, Wilsey C (2015) Conservation status of North American birds in the face of future climate change. PLoS ONE 10(9):e0135350
Lee D, Sarran C (2015) Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. Environmetrics 26(7):477–487
Leininger TJ, Gelfand AE et al (2017) Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. Bayesian Anal 12(1):1–30
Lobo JM, Baselga A, Hortal J, Jiménez-Valverde A, Gómez JF (2007) How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? Divers Distrib 13(6):772–780
MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey L, Hines JE (2017) Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier, Amsterdam
Mair L, Ruete A (2016) Explaining spatial variation in the recording effort of citizen science data across multiple taxa. PLoS ONE 11(1):e0147796
Møller J, Syversveen AR, Waagepetersen RP (1998) Log Gaussian cox processes. Scand J Stat 25(3):451–482

Murray I, Prescott Adams R, MacKay DJ (2010) Elliptical slice sampling
Niemuth ND, Dahl AL, Estey ME, Loesch CR (2007) Representation of landcover along breeding bird survey routes in the northern plains. J Wildl Manag 71(7):2258–2265
Oliveira U, Brescovit AD, Santos AJ (2017) Sampling effort and species richness assessment: a case study on Brazilian spiders. Biodivers Conserv 26(6):1481–1493
Pati D, Reich BJ, Dunson DB (2011) Bayesian geostatistical modelling with informative sampling locations. Biometrika 98(1):35–48
R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
Robinson OJ, Ruiz-Gutierrez V, Fink D (2018) Correcting for bias in distribution modelling for rare species using citizen science data. Divers Distrib 24(4):460–472
Rosenberg KV, Dokter AM, Blancher PJ, Sauer JR, Smith AC, Smith PA, Stanton JC, Panjabi A, Helft L, Parr M et al (2019) Decline of the North American avifauna. Science 366(6461):120–124
Roy HE, Adriaens T, Isaac NJ, Kenis M, Onkelinx T, Martin GS, Brown PM, Hautier L, Poland R, Roy DB et al (2012) Invasive alien predator causes rapid declines of native European ladybirds. Divers Distrib 18(7):717–725
Ruete A (2015) Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance map. Biodivers Data J 3
Sicacha-Parada J, Steinsland I, Cretois B, Borgelt J (2020) Accounting for spatial varying sampling effort due to accessibility in citizen science data: a case study of moose in Norway. Spat Stat 42
Sullivan BL, Aycrigg JL, Barry JH, Bonney RE, Bruns N, Cooper CB, Damoulas T, Dhondt AA, Dietterich T, Farnsworth A et al (2014) The ebird enterprise: an integrated approach to development and application of citizen science. Biol Conserv 169:31–40
Szabo JK, Vesk PA, Baxter PW, Possingham HP (2010) Regional avian species declines estimated from volunteer-collected long-term data using list length analysis. Ecol Appl 20(8):2157–2169
Szabo JK, Vesk PA, Baxter PW, Possingham HP (2011) Paying the extinction debt: woodland birds in the mount lofty ranges, South Australia. Emu-Austral Ornithol 111(1):59–70
Tiago P, Ceia-Hasse A, Marques TA, Capinha C, Pereira HM (2017) Spatial distribution of citizen science casuistic observations for different taxonomic groups. Sci Rep 7(1):1–9
van Strien AJ, van Swaay CA, Termaat T (2013) Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. J Appl Ecol 50(6):1450–1458
Zaniewski AE, Lehmann A, Overton JM (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol Model 157(2–3):261–280
Zhang H (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J Am Stat Assoc 99(465):250–261

Becky Tang is a PhD candidate at Duke University's Department of Statistical Science under the supervision of Professors Alan Gelfand and James Clark. Her research lies in the fields of spatial and ecological statistics. She is a recent recipient of the NSF GRFP fellowship.

James S. Clark is a professor at the Nicholas School of the Environment and the Department of Statistical Science at Duke University and at INRAE at the University of Grenoble. His research focuses on the effects of global change on biodiversity.

Alan E. Gelfand is James B. Duke Professor Emeritus in the Department of Statistical Science and also in the Nicholas School of the Environment. His research focuses on spatial and spatio-temporal modeling of complex environmental and ecological processes.