

Spatial extinction date estimation: a novel method for reconstructing spatiotemporal patterns of extinction and identifying potential zones of rediscovery

Colin J. Carlson^{1,2,*}, Kevin R. Burgio^{3, 4}, Tad A. Dallas⁵ and Alexander L. Bond⁶

¹*National Socio-Environmental Synthesis Center, University of Maryland, Annapolis, Maryland 21401, USA.*

²*Department of Biology, Georgetown University, Washington, D.C. 20057, USA.*

³*Department of Ecology and Evolutionary Biology; University of Connecticut; Storrs, CT 06268, USA.*

⁴*Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, CT 06269 USA.*

⁵*Centre for Ecological Change, University of Helsinki, 00840 Helsinki, Finland.*

⁶*Bird Group, Department of Life Sciences, The Natural History Museum, Akeman Street, Tring, Hertfordshire, HP23 6AP, United Kingdom.*

*Corresponding author: ccarlson@sesync.org

Abstract

1
2 1. The estimation of extinction dates from limited and incomplete sighting records
3 is a key challenge in conservation (when experts are uncertain whether a species has
4 gone extinct) and historical ecology (when the date and mechanism of extinction is
5 controversial).

6
7 2. We introduce a spatially-explicit method of interpolating extinction date estima-
8 tors, allowing users to estimate spatiotemporal surfaces of population persistence
9 from georeferenced sighting data of variable quality.

10
11 3. We present the R package `spatExtinct`, which produces spatially-explicit
12 extinction date surfaces from geolocated sightings, including options for custom
13 randomization schemes to improve accuracy with limited datasets. We use simu-
14 lations to illustrate the sensitivity of the method to parameterization, and apply
15 the method to identify potential areas where Bachman's warbler (*Vermivora bach-*
16 *manii*) might be rediscovered.

17
18 4. Our method, and the `spatExtinct` package, has the potential to help de-
19 scribe and differentiate different drivers of extinction for historical datasets, and
20 could be used to identify possible regions of population persistence for species with
21 an uncertain extinction status, improving on non-spatial or imprecise methods that
22 are currently in use.

23
24 **Keywords:** extinction date estimation, optimal linear estimator, sighting records,
25 Bachman's warbler, *Vermivora bachmanii*

26 1 Introduction

27 Biological extinctions are one of the most fundamental processes in ecology, and despite
28 their significance and pervasiveness, they are often impossible to observe directly. Even if
29 the last known individuals of a species are kept in captivity, uncertainty can still emerge
30 depending on researchers' confidence that the species is extinct in the wild. Moreover,
31 "sightings" of a species are liable to continue long after a species is presumed extinct,
32 compounding uncertainty and potentially fueling hope of rediscovery. (Carlson *et al.*,
33 2017b) Every so often, true "Lazarus species" are found after absences of a few centuries
34 (like the Bermuda Petrel, *Pterodroma cahow*) or millions of years (like coelocanth, *La-*
35 *timeria* spp.) But the majority of species are never rediscovered, presenting conservation
36 biologists with difficult decisions: when should a species be pronounced extinct, and at
37 what point should valuable conservation resources be redirected elsewhere? (Collen *et al.*,
38 2010; David & Davis, 2017)

39 A variety of approaches have been developed to address these challenging situations.
40 Some examine the relationship between species' traits and ecology, and observed extinc-
41 tion rates to estimate the probability of rediscovery (Fisher & Blomberg, 2010; Lee *et al.*,
42 2017b). Others, which we term extinction date estimators (EDEs), make assumptions
43 about the temporal distribution of sightings leading up to extinction, to estimate the
44 most likely date of extinction and corresponding probability of persistence (Boakes *et al.*,
45 2015). Over the past two decades, a number of methodological advances have made these
46 methods more powerful and precise (Boakes *et al.*, 2015), and in conservation practice,
47 researchers are encouraged to use these different methods in combination when possible
48 (Akçakaya *et al.*, 2017).

49 However, one notable limitation of almost all of these methods is that they treat
50 extinction as a single event for an entire species. The total eradication of a species is
51 usually the product of spatially-heterogeneous population declines over time, but spatial
52 tools for reconstructing extinctions are lacking. Spatial kriging is sometimes used to
53 interpolate first and last dates of appearance (Emery-Wetherell *et al.*, 2017), but this
54 basic method has a number of limitations. In particular, modeling based on the last
55 observation is comparatively imprecise, as the extinction date estimation literature shows
56 that the last sighting is usually insufficient to make educated predictions about a species'
57 true extinction date. Moreover, the kriging approach uses a limited regional subset of

58 data, and makes no inferences based on any of the time series aspects of sighting records.
59 In the context of applying this method to recent (and unconfirmed) extinctions, the
60 kriging approach cannot be used to estimate the probability of persistence, or account for
61 uncertainty in the veracity of sightings.

62 Species distribution models (SDMs; also called ecological niche models, or ENMs) are
63 another valuable tool for reconstructing the biogeography of extinct species. SDMs con-
64 ventionally relate occurrence data (sightings) to environmental conditions via some form
65 of regression or machine learning, and make inference about the geographic distribution
66 of species based on their ecological niche. SDMs can be used to reconstruct the shifting
67 distributions over time for extinct species like the megalodon (*Carcharocles megalodon*) in
68 conjunction with extinction date estimators (Pimiento & Clements, 2014; Pimiento *et al.*,
69 2016), though this application is tenuous over shorter timescales, as most SDM methods
70 assume that distributions are at equilibrium within the scale of modeling. SDMs have also
71 been used with long-extinct species to recover biologically-meaningful information from
72 biogeographic data; for example, a recent study on the Carolina parakeet (*Conuropsis*
73 *carolinensis*) identified two distinct subspecies' ranges, and showed that only one sub-
74 species exhibited partial seasonal migration (Burgio *et al.*, 2017). In the shorter term,
75 SDMs are a critical tool for conservation planning, and can be used to help guide the
76 search for possibly-extinct species, even alongside extinction date estimators. (Makenov,
77 2018) However, for the rarest species, the necessary occurrence data may be impossible to
78 collect. Some workarounds exist, like using data from related species (Dunn *et al.*, 2015),
79 or using Bayesian belief networks to formalize *ad hoc* hypotheses about the species' niche
80 (Grainger *et al.*, 2017). Even then, environmental suitability may be a poor proxy for
81 presence especially for a species near extinction; in these cases, the total suitable area
82 is likely to be much broader than the true area of occupancy. Methods from the occu-
83 pancy modeling literature that address this pattern tend to be more data intensive, and
84 require a depth and regularity of observations and abundance data that most putatively-
85 or near-extinct species lack.

86 We therefore identify a major unaddressed need: researchers interested in reconstruct-
87 ing spatiotemporal patterns of extinction have limited options without explicit data on
88 population declines. While the theoretical underpinnings of extinction date estimators
89 could likely be extended to produce explicitly-spatial analytical approaches, these ex-

90 tensions have yet to be developed, creating an opportunity for the development of a
91 computational, approximate approach. Here, we introduce the idea of *spatial extinction*
92 *date estimators* (SEDEs) as a tool for recovering the geographic patterns of extinction and
93 identifying where possibly-extant species might be rediscovered. The method we propose
94 uses georeferenced sighting data to estimate extinction dates over a landscape, including
95 for species that have a small chance of persisting somewhere undetected. We show how
96 to optimize the method using simulations, and implement a case study with Bachman’s
97 warbler (Parulidae: *Vermivora bachmanii*), a charismatic North American bird that is
98 likely extinct.

99 2 The Models

100 2.1 Extinction Date Estimators

101 How do we know if a species is extinct? Extinction date estimators (EDEs) determine the
102 status of a species based on a set of “sightings” including observations, photographs, and
103 physical evidence (like carcasses, specimens, or scat). Sightings can have different levels
104 of support and of validity, and often continue long after a species is extinct. A sighting
105 dataset can be expressed as an ordered set $\mathbf{t} = (t_1, \dots, t_n)$, and extinction date estimators
106 make an assumption about the distribution that generates those sightings before (and
107 sometimes after) an extinction event to estimate the true date of extinction T_E (Carlson
108 *et al.*, 2017a). One of the most popular methods, the optimal linear estimator (OLE) is
109 a non-parametric method first used to estimate the extinction date of the dodo (Roberts
110 & Solow, 2003). It assumes that the k last few sightings of a species follow a Weibull
111 distribution:

$$\hat{T}_E = \sum_{i=1}^k w_i t_{n-i+1} \quad (1)$$

$$w = (e' M^{-1} e)^{-1} M^{-1} e \quad (2)$$

113 where e is a vector of k 1’s, and M is a k by k matrix, for which

$$M_{ij} = \frac{\Gamma(2\hat{\nu} + i)\Gamma(\hat{\nu} + j)}{\Gamma(\hat{\nu} + i)\Gamma(j)} \quad (3)$$

114

$$\hat{v} = \frac{1}{k-1} \sum_{i=1}^{k-2} \ln \left(\frac{t_n - t_{n-k+1}}{t_n - t_{i+1}} \right) \quad (4)$$

115 An upper 95% confidence bound is given for the OLE by

$$\widehat{T}_{ci}^u = T_n + \frac{T_n - T_{n-k+1}}{c(\alpha) - 1} \quad (5)$$

$$c(\alpha) = \left[\frac{-\log(\alpha/2)}{k} \right]^{-\hat{v}} \quad (6)$$

116 The OLE is one of the most popular EDEs, in large part due to its strong performance
117 even in the face of confounding factors like temporal variation in sampling rates (Ri-
118 vadeneira *et al.*, 2009). The OLE performs well as an extinction date estimator with
119 limited data, with experimental work finding no universally-optimal sample size and rec-
120 ommending that the method can be best used with all available data (Clements *et al.*,
121 2013). One notable downside of the OLE method is that it produces extremely wide con-
122 fidence intervals, especially with larger datasets (Rivadeneira *et al.*, 2009). However, the
123 wide upper confidence bound can be a strength in cases where extreme levels of caution
124 are desired from an extinction date study.

125 One significant drawback of the OLE, and similar estimators, is that the inclusion
126 of any invalid data proportionally produces significant error in the estimates (Roberts
127 *et al.*, 2010). More recently, a new class of EDEs have been proposed that account for
128 variation in certainty and validity among different sightings (Boakes *et al.*, 2015). These
129 methods tend to be Bayesian, and assume that valid sightings can only exist prior to the
130 extinction date (Solow *et al.*, 2012; Solow & Beet, 2014; Lee *et al.*, 2014). While some of
131 these models can account for variable degrees of confidence in different data sources (Lee
132 *et al.*, 2014), recent work has indicated that expert evaluation of sightings beyond certain
133 and uncertain sightings may be unnecessary (Lee *et al.*, 2017a).

134 In this study, we adapt two models from Solow & Beet (2014), which assume that
135 data could contain a mix of valid and invalid sightings. The dataset of n sightings \mathbf{t} in an
136 interval $(0, T]$ is split in these models into certain (t_c , with length n_c) and uncertain (t_u ,
137 with n_u sightings including $n_u(\hat{T}_E)$ before \hat{T}_E) sub-datasets by researchers. The rubric for
138 that split is subjective, but typically, certain sightings involve incontrovertible evidence

139 like a physical specimen, or a clear and uncontroversial photograph. In contrast, uncertain
140 sightings have a much broader range from grainy but plausible video footage (as in the
141 case of the ivory-billed woodpecker), down to unsubstantiated reports of modern-day non-
142 avian dinosaurs. (Smith, 2015) In some cases, it may be helpful for data management to
143 add extra levels of resolution within “uncertain,” such as expert versus novice sightings,
144 even if the model makes no distinction.

145 If valid sightings occur at a true rate Λ and invalid sightings occur at a true rate Θ ,
146 the proportion of valid sightings is given as

$$\Omega = \frac{\Lambda}{\Lambda + \Theta} \quad (7)$$

147 By allowing for a mix of certainty and validity within sightings before extinction, the
148 model makes inferences about the strength of evidence after any hypothesized extinction
149 date. Solow & Beet (SB) model 1 assumes that certain and uncertain sightings follow the
150 same Poisson process. The conditional likelihood of the dataset \mathbf{t} if the species is extinct
151 is

$$p(\mathbf{t}|E) = \int_{t_L}^T p(\mathbf{t}|\tau_E)p(\tau_E)d\tau_E \quad (8)$$

152 where t_L is the date of the last certain sighting (the starting point of when extinction
153 is possible). Based on the likelihood of the underlying Poisson process for sightings, the
154 likelihood of the dataset given any extinction date is

$$p(t_u|\hat{T}_E) = \int_0^1 \omega^{-n_u}(1 - \omega)^{n-n(\hat{T}_E)}(n - 1)!(\hat{T}_E + \frac{1 - \omega}{\omega}T)^{-n}d\omega \quad (9)$$

155 where ω is a stand-in for Ω , to allow integration over all possible values of Ω , the true
156 value of which is unknown. In model 2, certain and uncertain sightings are generated by
157 two independent Poisson processes, and the conditional likelihood of the whole dataset is
158 the product of the likelihoods of the respective sub-datasets:

$$p(\mathbf{t}|\hat{T}_E) = p(t_c|\hat{T}_E)p(t_u|\hat{T}_E) \quad (10)$$

$$p(t_c|\hat{T}_E) = \frac{(n_c - 1)!}{(\hat{T}_E)^{n_c}} \quad (11)$$

$$p(t_u|\hat{T}_E) = \int_0^1 \omega^{-n_u} (1 - \omega)^{n_u - n_u(\hat{T}_E)} \left(\hat{T}_E + \frac{1 - \omega}{\omega} T\right)^{-n_u} d\omega \quad (12)$$

159 where ω is again a stand-in for Ω .

160 The probability the species went extinct in $(0, T]$, an event E , can be expressed using
161 Bayes' theorem:

$$p(E|\mathbf{t}) = \frac{p(\mathbf{t}|E)p(E)}{p(\mathbf{t})} = \frac{p(\mathbf{t}|E)p(E)}{p(\mathbf{t}|E)p(E) + p(\mathbf{t}|\bar{E})(1 - p(E))} \quad (13)$$

162 The prior probability of extinction $p(E)$ is somewhat hard to set, so for explicit calculation,
163 it is often uninformatively set to 0.5 (extinction and persistence are equally likely). If
164 $p(E) = p(\bar{E}) = 0.5$, that formula can be reduced to

$$p(E|\mathbf{t}) = \frac{p(\mathbf{t}|E)}{p(\mathbf{t}|E) + p(\mathbf{t}|\bar{E})} \quad (14)$$

165 which can be readily interpreted as the “probability of persistence” for the given year. In
166 some cases, researchers may opt instead to use the Bayes factor, which expresses relative
167 support for the alternative hypothesis and is given as

$$B = \frac{p(\mathbf{t}|E)}{p(\mathbf{t}|\bar{E})} \quad (15)$$

168 A higher Bayes factor implies stronger support for extinction, where a value of 3 or higher
169 could be taken as strong evidence the species had gone extinct. While the advantage of
170 the Bayes factor is that it avoids the problem of setting $p(E)$ altogether, both the Bayes
171 factor and $p(E|\mathbf{t})$ require setting the conditional likelihood of the data $p(\mathbf{t}|E)$, which
172 decomposes into

$$p(\mathbf{t}|E) = p(\mathbf{t}|\hat{T}_E)p(\hat{T}_E) \quad (16)$$

173 and conversely $p(\mathbf{t}|\bar{E})$ is evaluated using the same function but setting \hat{T}_E as T . The first
174 term can be evaluated as derived above; but the prior probability of a given extinction
175 date $p(\hat{T}_E)$ is again subjective and difficult to set. The selection of priors for Bayesian
176 EDEs is an important part of correctly implementing these methods (Solow, 2016); Solow
177 & Beet (2014) suggest either a uniform, linear, or exponential decline after the last certain
178 sighting, and while we have usually elected to use a uniform prior (Carlson *et al.*, 2017b),

179 some researchers may elect to make more informed or constrained choices.

180 2.2 Spatial Extinction Date Estimators

181 Here we propose a new class of models we term *spatial extinction date estimators* (SEDEs):
182 spatially-explicit interpolations of extinction date estimators using georeferenced sighting
183 data. Whereas EDEs treat extinction as a one-time irreversible event for an entire species,
184 spatial extinction date estimation decomposes extinction into landscape-level extirpation,
185 a set of one-time, irreversible events at the end of local population declines. To do so, our
186 proposed method assigns extinction (extirpation) dates to every cell of a pre-determined
187 grid meant to represent the species' range or the landscape of interest. In the non-random
188 model, for every grid cell, the k nearest neighbor sightings are taken from the centroid,
189 and are run through a specified extinction date estimator. In the random model, a dataset
190 of N nearest neighbors are generated for each cell, and estimators are run with a set of
191 k records randomly selected without replacement. (Sampling with replacement produces
192 severely distorted estimates, as even one or two extra late sightings can produce centuries-
193 late extinction dates.)

194 SEDEs use the same sighting datasets as typical EDEs, with the only additional
195 requirement that every record be georeferenced (i.e., sightings are recorded as presence
196 records with date and locality). Depending on data availability, essentially any EDE
197 could be implemented in this modeling framework. Here, we illustrate how SEDEs can
198 be constructed using the optimal linear estimator (OLE), and Solow & Beet's (2014)
199 Bayesian method for incorporating sighting uncertainty (SB), which we selected based
200 on their ubiquity in the literature, and their demonstrated strong performance relative
201 to other methods. As for non-spatial implementations, the OLE method should only
202 be used for certain sightings (Roberts *et al.*, 2010), while the SB model is designed for
203 use with mixed-certainty data (a common problem in the sighting record of extinct or
204 putatively-extinct species).

205 In that they reconstruct range contraction over time, SEDEs are, at least in princi-
206 ple, temporally-dynamic species distribution models: though they use no environmental
207 covariates to make predictions (as "ecological niche models" do), they similarly make a
208 model-based inference about the geographic range of a species based on known occurrence
209 points. For species that are extinct, SEDEs can be used to describe the spatiotemporal

210 process of extinction, which may give clues as to the mechanism. For example, a panzootic
211 disease might spread in a spatial wavefront from a single location (Lips *et al.*, 2008); in
212 contrast, habitat loss or land use change might correlate in a patchwork fashion with local
213 extinctions across a landscape (Preston *et al.*, 2012). In addition to reconstructing the
214 pattern of extinction across a landscape, SEDEs can be used to identify possible zones of
215 persistence for species with an uncertain extinction status. Even if a species-level EDE
216 suggests a low probability of persistence, spatially-subsetted data may indicate potential
217 areas with low support for extinction. This can be done by identifying areas where either
218 the OLE or SB model estimates \hat{T}_E later than the present year. (The confidence intervals
219 for the OLE method could also be used for this purpose, though we discourage this as
220 anything other than an extremely conservative estimate, given how wide these confidence
221 intervals tend to be; see Clements *et al.* [2013]). Additionally, the SB model can also
222 be used to calculate the posterior probability of persistence for a given year, based on a
223 given set of prior assumptions. Zones of potential persistence can then be delineated with
224 a desired confidence level.

225 **3 The R Package**

226 We developed the R package `spatExtinct` to implement SEDEs for use in historical ecol-
227 ogy and conservation research. The package utilizes some pre-existing functions, including
228 the OLE implementation in the package `sExtinct` (Clements, 2013); and includes some
229 new functions, including an R implementation of Solow & Beet’s model, which has been
230 previously published (Carlson *et al.*, 2017b) but is streamlined in our package.

231 **3.1 Estimating Spatial Extinction Surfaces**

232 The primary function of `spatExtinct` is to spatially interpolate the models we describe
233 above, using spatially-explicit occurrence data. The OLE and SB models are respectively
234 implemented in the `spat.OLE` and `spat.SB` functions. These methods are fairly com-
235 putationally intensive and work on a cell-by-cell basis, but the package includes options
236 for parallelization and adaptive estimation (with a set convergence threshold to reduce
237 the number of runs). The only required data to run the basic functions are: (1) a data
238 frame with sightings’ decimal longitude and latitude, date (year or any other internally-

239 consistent way of denoting time), and scored sighting quality (expert-verified, plausible, or
240 uncertain); and (2) a raster onto which extinction dates can be projected. For the raster,
241 local grids can be used based on the area of interest; or, if researchers are interested in
242 patterns across a species' entire range, we suggest that species distribution models can be
243 used to define appropriate boundaries for interpolation, as they can be generated from the
244 same sighting data used by the package, and provide an intuitive constraint on outermost
245 area of occupancy. The format in which `spatExtinct` uses data is readily usable by
246 species distribution modeling packages like `dismo`. (Hijmans *et al.*, 2013)

247 3.2 Estimating Zones of Persistence

248 The primary utility of `spatExtinct` is estimating the last likely year of presence on
249 a cell-by-cell basis across landscapes. However, there is one readily-obvious extension
250 for species that may *not* be entirely extinct: `spatExtinct` can be used multiple ways
251 to identify potential zones of persistence. This can be done by using `spat.OLE` and
252 `spat.SB` to find areas where \hat{T}_E is later than the current year, but we have also included
253 an explicitly probabilistic function `spat.SB.probs` that estimates the probability of
254 persistence in a set year. That approach would work based on a hypothesis test that the
255 date of extinction T_E is not before the current time T , where assuming some significance
256 cutoff α , we delineate cells (i, j) for which

$$P(T_E \not\leq T)_{ij} \geq \alpha \quad (17)$$

257 For instance, if we wanted to identify areas where there is at least a 10% chance the
258 species is not yet extinct, we would set $\alpha = 0.10$ and map all cells meeting that criterion.
259 We suggest that this more effective, or at the least more subjective on the user end, than
260 simply cutting off by $\hat{T}_E \not\leq T$.

261 We also suggest that researchers can easily interface SEDEs and ENMs, as a simple
262 but powerful approach to optimizing rediscovery efforts with almost no *a priori* assump-
263 tions. Ecological niche models represent the probability of a species' presence relative to
264 a given set of environmental variables, and so a combined probability of rediscovery can
265 be conceptualized as

$$P(\text{rediscovery}) \approx P(T_E \not\leq T | \text{suitable}) * P(\text{suitable}) \quad (18)$$

266 This relative approach does not represent a “true probability” of rediscovery, but can be
267 used to prioritize search efforts in areas of suitable habitat with plausible undiscovered
268 populations, or to guide reserve design for cryptic or rarely-sighted species, for example.
269 The implementation of that process is particularly flexible; for instance, we can identify
270 at least three approaches:

- 271 1. Using a thresholded SDM as the base grid for `spat.OLE` or `spat.SB` and identi-
272 fying zones where $(\hat{T}_E)_{ij} \not\leq T$
- 273 2. Using a thresholded SDM as the base grid for `spat.SB.probs` and identifying
274 zones where $P(T_E \not\leq T)_{ij} \geq \alpha$
- 275 3. Using raw ENM suitability values and `spat.SB.probs` on the same landscape,
276 taking the product to approximate $P(\text{rediscovery})$, and identifying areas where the
277 combined probability is above either a pre-determined threshold or a quantile (e.g.
278 mapping the top 10% of sites as hotspots of possible rediscovery).

279 We include no direct tool to interface ENMs and SEDEs, not due to lack of feasibility,
280 but because we stress the importance of careful user-end precision in the implementa-
281 tion of ENMs. All methods for ecological niche modeling are sensitive to sampling bias,
282 pseudoabsence generation, environmental variable set selection, and parameter tuning de-
283 cisions (Elith & Graham, 2009; Merow *et al.*, 2014); rather than include an automated
284 workflow, we encourage researchers to make careful decisions about these factors when
285 building species distribution models for rare, vulnerable species. We include an example
286 here using these methods to identify possible zones of rediscovery for Bachman’s warbler.

287 4 Applications

288 We briefly discuss two examples of how our method can be implemented, to illustrate
289 the flexibility of the package, and the relevance in both historical ecology and current
290 conservation work.

291 4.1 Simulations: Validation and Parameterization

292 We use simulated extinctions to demonstrate the relative accuracy of the methods as a
293 function of data availability and model implementation. Simulations were run on an n
294 by n square landscape where extinction dates occur on a linear gradient from 5 to $5n$
295 (e.g. a 10 by 10 landscape has cell extinction dates ranging from 5 to 50). For every
296 year in the interval of $[5, 5n)$, the landscape is constrained to areas where populations are
297 extant; p “sighting” points are generated on that surface within every year for a total of
298 $p(5n - 1)$ points (see **Figure 1** for an example). As a consequence of this, later extinctions
299 slightly autocorrelate with denser sightings. We used simulated datasets, iterated over
300 adjustable model parameters, to develop an optimization protocol for `spatExtinct`.
301 We present our analyses with two accuracy measures: correlation of estimated and real
302 extinction surfaces (is the **pattern** clear?), and mean squared error between estimated
303 and real surfaces (how **precise** are extinction date estimates?). More advanced tuning
304 issues are discussed in the Supporting Information, but here, we examine a handful of
305 basic questions:

306 4.1.1 Which model performs best?

307 Independent of randomization method, we found that Solow & Beet’s method consistently
308 performed better, with higher correlation coefficients (which were fairly insensitive to pa-
309 rameterization, unlike the OLE method) and lower error rates (**Figure 2**). Consequently,
310 for both pattern estimation and explicit extinction date estimates, the SB method is likely
311 a better one than the OLE (especially if working with mixed-certainty data; see below).
312 However, there are still cases where researchers may want to use the OLE method; most
313 notably, if attempting to delineate regions with a minor chance of persistence, the char-
314 acteristically wide upper confidence bound of the OLE method (Clements *et al.*, 2013) is
315 a strength of the approach. We suggest the best practice is simply to present both, with
316 necessary disclaimers about known levels of accuracy.

317 4.1.2 How sensitive are models to parameterization?

318 For the OLE method, we found that error rates were minimized around $k \approx 7$, with limited
319 randomization ($N = k + 1$) or no randomization ($N = k + 0$); however, an intermediate
320 level of randomization ($N = k + 5$) significantly improved correlations with real surfaces

321 **(Figure 2)**. For the SB method, smallest levels of k minimized error, with increasing
322 levels of randomization similarly leading to decreasing accuracy across k values. Again,
323 intermediate levels of randomization ($N = k + 5$) maximized correlations, especially at
324 low k values. Across methods, we suggest this indicates an unsurprising tradeoff between
325 pattern inference and the precision of estimates: randomization helps smooth estimated
326 surfaces, and makes them easier to interpret, but at the cost of local accuracy. For
327 the most precise estimates (e.g. delineating potential zones of persistence), researchers
328 should select a small neighborhood size and limited or no randomization. In contrast, mild
329 randomization may help researchers accurately interpret the spatial patterns of extinction
330 over landscapes.

331 **4.1.3 How does sample size affect accuracy?**

332 Higher sample sizes consistently improve model performance (**Figure 3**). For both the
333 OLE (A,B) and SB (C,D) method, accuracy plateaus noticeably around 40-50 points.
334 Correlations plateau more substantially, whereas increased sample size continues to reduce
335 estimates' error. However, we note that the error of SB estimates around 15-25 points
336 is still comparable to that of OLE estimates around 50 or more, highlighting its better
337 performance as an estimator.

338 **4.1.4 Can mixed certainty sighting data be effectively utilized?**

339 The most significant strength of the Solow & Beet (2014) model is the capacity to use
340 mixed-certainty sighting data, including potentially invalid sightings, to improve extinc-
341 tion date estimates. We show that this benefit still exists when the Solow & Beet models
342 are spatially interpolated (see **Figure 4**); our analyses suggest that a higher number
343 invalid points only severely reduces model performance when valid points have a high
344 uncertainty rate (see **Supporting Information**). However, a quality control option that
345 removes uncertain sightings exists in the OLE functions, for cases where researchers may
346 want to include OLE analyses despite uncertain sightings. This method, while imprecise,
347 still vastly improves the OLE's performance compared to the dramatic negative impact
348 of inaccurate sightings on unrestricted OLE analyses.

349 4.2 Case Study: Bachman’s Warbler

350 To illustrate the potential use of `spatExtinct` for estimating zones of plausible per-
351 sistence, we provide an example using sightings of Bachman’s warbler, a species that is
352 widely believed extinct despite recent, extremely controversial “sightings.” The species
353 is believed to have experienced sharp population declines between 1910 and 1930, pos-
354 sibly due to habitat loss in the southern United States or its overwintering grounds in
355 Cuba (due in part to hurricane damage); however, as for many species, the precise rea-
356 sons for decline are controversial and unresolved (Stevenson, 1972; Pimm & Askins, 1995;
357 Huntington & Barbour, 1936; Hamel, 2011).

358 Unlike higher-profile North American extinct species like the ivory-billed woodpecker
359 (*Campephilus principalis*) or passenger pigeon (*Ectopistes migratorius*), the status of
360 Bachman’s warbler has received comparatively little attention. To the extent of our
361 knowledge, the only modeling study published on the subject suggested an extinction
362 date of 1961 (upper 95% CI: 1967) based on physical evidence only, and 1964 (95% CI:
363 1967) including expert-opinion sightings (Elphick *et al.*, 2010); however, controversial and
364 unverifiable sightings have been reported as recently as 2001 (Chamberlain, 2003). It is
365 not implausible that Bachman’s warbler could go undetected for several years, given the
366 species’ biology. One naturalist’s report from the early 20th Century notes: “These birds
367 are very hard to detect...in fact I cannot recall a bird that moves as rapidly as Bachman’s
368 Warbler does in the breeding season” (Chamberlain, 2003). On the other hand, recent
369 sightings are particularly dubious given the risk of misidentification, as the species has a
370 strong resemblance to the extant hooded warbler (*Wilsonia citrina*).

371 To evaluate the status of the species, we assembled a georeferenced dataset contain-
372 ing all readily-accessible known records of Bachman’s warbler in the continental United
373 States. Sightings were sorted into an established set of three categories (Solow & Beet,
374 2014; Carlson *et al.*, 2017b): confirmed sightings with physical evidence, expert-supported
375 sightings, and unconfirmed but plausible sightings. We developed a species distribution
376 model following a similar protocol to Burgio *et al.* (2017), using the package `ENMeval`
377 to tune maximum entropy (MaxEnt) species distribution models for the data. We used
378 the data to develop a spatial surface for the probability of persistence in 2017 using the
379 `spat.SB.probs` function in `spatExtinct`, and we use the to estimate the probabilit-
380 y of persistence in 2017 based on certain and uncertain sightings. (See **Supporting**

381 **Information** for more details on data collection and model implementation.)

382 In total we found 118 usable historical records of Bachman’s warbler (64 certain and
383 verifiable sightings, 46 strongly plausible expert sightings, and 8 implausible or novice
384 sightings). Sightings of Bachman’s warbler dated back as far as the species’ taxonomic
385 description in 1833, and as late as the last certain sighting in 1959, with both sightings
386 recorded near Charleston, South Carolina. The last unconfirmed “sighting” in 2001 was
387 recorded in Congaree Swamp Park in Richland, South Carolina (Chamberlain, 2003).
388 After georeferencing, there were 86 spatiotemporally unique sightings recorded, at 47
389 unique localities. While this is sufficient for species distribution modeling (Proosdij *et al.*,
390 2016), it is still astonishing how limited data are for a charismatic North American species
391 that only went (probably) extinct in the last century. (Georeferenced data were limited
392 enough that we suggest interpreting the below analyses as more of a vignette for how a
393 study could be designed using `spatExtinct`, than a definitive appraisal of the status
394 of the species.)

395 The hypothesis that Bachman’s warbler has been extinct for several years was sup-
396 ported by both the optimal linear estimator (all data: $\hat{T}_E = 2004$, 95% CI = (2001,2012);
397 certain sightings only: $\hat{T}_E = 1965$, 95% CI = (1959,1987)) and the Solow & Beet method
398 (model 1: $\hat{T}_E = 1978$, $p(T_E \geq 2017) = 0.011$; model 2: $\hat{T}_E = 1963$, $p(T_E \geq 2017) =$
399 1.5×10^{-5}). Without any spatial information, these results suggest the species can be
400 safely presumed extinct in 2018; in fact, these estimates suggest the species was probably
401 already extinct by 1967, when it was included on the first federal listing of endangered
402 species in the United States (the “Class of ’67”). The species is still listed despite its
403 apparent extinction; in 2015, the U.S. Fish and Wildlife Service most recently reassessed
404 the species, maintaining its Endangered status and explaining, “We considered recom-
405 mending delisting Bachmans warbler; however, because Bachmans warbler is difficult to
406 detect and identify (Chamberlain 2003) and the lack of formal extensive search efforts
407 over the last 27 years, considerable uncertainty remains as to its status.” (Sisson, 2015)

408 Our species distribution model agreed with prior expert knowledge, suggesting a broad
409 geographic division between a coastal Atlantic range, and an inland range following the
410 Mississippi and Ohio River basins, tracking the distribution of baldcypress (*Taxodium*
411 *distichum*) in the South. Previous work suggested these ranges were continuous and
412 connected above the Gulf of Mexico, though our model suggests the ranges may have been

413 more separate (Hamel, 2011). The spatial Solow & Beet model suggested that, despite
414 one late uncertain sighting in northern Louisiana, the western range of the species was
415 likely gone by the early 20th century (**Figure 5**). The model suggested persistence in the
416 1960s or 1970s along the coast of Georgia and South Carolina, as well as (surprisingly)
417 around the Chesapeake Bay. These estimates still suggest that the species was likely
418 extinct by the mid-1970s.

419 As a final test, we used the probability of persistence ($p(T_E \leq 2017)$, generated with
420 the `spat.SB.probs` function) to test the hypothesis of local extinction and verify the
421 minimal chance of rediscovery. In some of the most uncertain and probably error-prone
422 areas (West Virginia and North Carolina) the probability was surprisingly high. However,
423 the model still found broad regions with a small but nontrivial (5-20%) posterior proba-
424 bility of presence. Using the species distribution model as an additional probability filter,
425 we found three major areas of potential rediscovery: coastal Georgia and South Carolina,
426 the Mississippi delta in Louisiana, and a broad patch from North Carolina up through
427 Maryland (**Figure 6**).

428 The idea that there might be undiscovered, isolated populations of Bachman’s warbler
429 in the southeastern United States—especially in South Carolina—is far from new. Official
430 U.S. Fish & Wildlife Service documentation still lists the species as having a range along
431 coastal South Carolina and the southernmost end of Florida (the latter being part of its
432 migratory range; USFWS [2018]). In the mid-1970s, a series of unconfirmed sightings were
433 reported in the I’on Swamp in coastal South Carolina. (Shuler, 1976) After unconfirmed
434 but plausible sightings in 2000 and 2001 in Congaree National Park, the park was surveyed
435 in 2002 without any success (Chamberlain, 2003), suggesting that even if there were late
436 persistent populations in the area, they are likely gone by today.

437 The prediction of a possible zone of persistence further north is surprising, given the
438 absence of late plausible sightings, but could possibly merit further investigation. Most
439 promising, in our opinion, is the possible zone of persistence in the Mississippi basin.
440 While the swamps of Congaree have been searched, remaining baldcypress swamps along
441 the Mississippi could just as plausibly harbor an undiscovered remnant population; the
442 2015 FWS assessment noted optimistically that efforts to rediscover the ivory-billed wood-
443 pecker in similar habitats might incidentally lead to the warbler’s rediscovery. (Sisson,
444 2015) Given that Bachman’s warbler was inconspicuous at best, it may also be possible

445 that the bird, if it still persists, will be found in unexpected areas. SDMs have been proven
446 useful in finding new populations of other elusive species (Menon *et al.*, 2010; Williams
447 *et al.*, 2009; Fois *et al.*, 2015), so it may well be worth looking in the areas suggested
448 by our models. But all evidence suggests the species is likely extinct, and the decision
449 to keep searching for Bachman’s warbler may also detract from other, more promising
450 conservation efforts—and the decision to continue the search depends at least as much on
451 values as it does on model results. (Akçakaya *et al.*, 2017; Carlson *et al.*, 2017b; David &
452 Davis, 2017; Jackson, 2006)

453 5 Discussion

454 Here, we have introduced the idea of a spatially-explicit approach to extinction date
455 estimation, and an associated R package. Spatial extinction date estimation makes an
456 important conceptual link between extinction date estimation and species distribution
457 modeling, and when parameterized correctly, we suggest that our method can be readily
458 used to reconstruct the spatial pattern of range loss during a species’ extinction. Provided
459 that sufficient data is available, we suggest that spatial extinction date estimation has
460 tremendous promise as a new tool for assessing the status of putatively-extinct species
461 like Bachman’s warbler, even when some sightings are suspected to be invalid.

462 Spatial extinction date estimation is a new method, and our simulations suggest it
463 works fairly well. However, the computational approach to approximation we use in the
464 `spatExtinct` package would likely be outperformed by an analytic approach that adapts
465 a model like Solow & Beet’s to explicitly consider spatial autocorrelation in extinction
466 dates and spatial heterogeneity in sighting rates. (In fact, we hypothesize that much
467 of the temporal heterogeneity in sighting rates that these models account find could be
468 attributed to the combination of spatial heterogeneity and species’ contracting ranges.)
469 We encourage the development of superior modeling approaches, and similarly encourage
470 caution interpreting the results of any model. Users of the method, and the R package,
471 should carefully consider the bias different models contribute: for example, Solow & Beet’s
472 model 1 is strongly influenced by the last uncertain sighting and predicts persistence more
473 commonly than model 2, which is more strongly influenced by the last certain sighting
474 (Kodikara *et al.*, 2018). The relative accuracy of the two models is influenced by the

475 proportion of uncertain sightings that are invalid, something users of our method should
476 strongly consider.

477 After the extinction of a species, the amount of biological information that can be
478 recovered rapidly declines, especially for poorly-documented recent extinctions without
479 a detailed specimen or fossil record. As the sixth mass extinction accelerates due to
480 forces like climate change, an increasing amount of information on global biodiversity is
481 irretrievably lost (Barnosky *et al.*, 2011; Ceballos *et al.*, 2015; Urban, 2015). In particularly
482 data-deficient situations, the mechanism of an extinction may be uncertain. In some cases,
483 a species' status as extinct may itself be uncertain. Spatially-interpolated extinction date
484 estimators, and their implementation in the `spatExtinct` package, are designed to
485 address both of these data-limited situations. We suggest, though, that the strongest use
486 of SEDEs is alongside other tools. Statistical methods like ENMs hold clear promise as
487 a companion to SEDEs, but sighting data can only accomplish so much. Using these
488 tools alongside specimen work, however, is likely to be especially powerful as a path of
489 inquiry. Identifying cause of death from specimens may help explain mortality patterns
490 and develop informative population models (Cunningham & Daszak, 1998), and stable
491 isotope work can help recover key information about changing patterns of diet or migration
492 (Hilderbrand *et al.*, 1996). All of these methods in conjunction can help develop a more
493 robust perspective on extinction as a spatiotemporal process, rather than a single event
494 in time, benefiting work in historical ecology and conservation biology alike.

495 **Acknowledgements**

496 We thank Andrew Beet for the original Matlab code used in Solow & Beet (2014); Adam
497 Butler (Biomathematics and Statistics Scotland) for translating the Matlab code into R;
498 Christopher Clements for helpful feedback and brainstorming; Nathaniel J. Ruess, for
499 moral support the last six years; and Robert Rankin for statistics advice.

500 **Author Contributions**

501 CJC, KRB, and ALB conceived of the idea for the study. CJC and KRB developed
502 the modeling framework, and CJC developed the R package. TAD and CJC developed
503 the simulations. KRB collected the Bachman's warbler dataset, and CJC and KRB ran

504 subsequent analyses. All authors contributed to the writing and editing of the manuscript.

505 **Data Accessibility**

506 `spatExtinct` can be downloaded from <https://github.com/cjcarlson/spatExtinct>.

507 Code to reproduce simulation analyses, and the Bachman's warbler dataset, are available

508 on Figshare at (**link forthcoming**).

509 References

- 510 Akçakaya, H., Keith, D.A., Burgman, M., Butchart, S.H., Hoffmann, M., Regan, H.M.,
511 Harrison, I. & Boakes, E. (2017) Inferring extinctions III: A cost-benefit framework for
512 listing extinct species. *Biological Conservation*, p. in press.
- 513 Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O., Swartz, B., Quental, T.B., Mar-
514 shall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C. *et al.* (2011) Has the Earth's
515 sixth mass extinction already arrived? *Nature*, **471**, 51–57.
- 516 Boakes, E.H., Rout, T.M. & Collen, B. (2015) Inferring species extinction: the use of
517 sighting records. *Methods in Ecology and Evolution*, **6**, 678–687.
- 518 Burgio, K.R., Carlson, C.J. & Tingley, M.W. (2017) Lazarus ecology: Recovering the
519 distribution and migratory patterns of the extinct Carolina parakeet. *Ecology and*
520 *Evolution*, **7**, 5467–5475.
- 521 Carlson, C., Burgio, K., Dallas, T. & Getz, W. (2017a) The mathematics of extinction
522 across scales: from populations to the biosphere. *PeerJ PrePrints*.
- 523 Carlson, C.J., Bond, A.L. & Burgio, K.R. (2017b) Estimating the extinction date of the
524 thylacine with mixed certainty data. *Conservation Biology*, p. in press.
- 525 Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M. & Palmer, T.M.
526 (2015) Accelerated modern human-induced species losses: Entering the sixth mass
527 extinction. *Science Advances*, **1**, e1400253.
- 528 Chamberlain, D. (2003) Additional notes on Bachman's Warbler. *Chat*, **67**, 5–10.
- 529 Clements, C. (2013) `sExtinct`: Calculates the historic date of extinction given a series
530 of sighting events. R package version 1.1.
- 531 Clements, C.F., Worsfold, N.T., Warren, P.H., Collen, B., Clark, N., Blackburn, T.M. &
532 Petchey, O.L. (2013) Experimentally testing the accuracy of an extinction estimator:
533 Solow's optimal linear estimation model. *Journal of Animal Ecology*, **82**, 345–354.
- 534 Collen, B., Purvis, A. & Mace, G.M. (2010) Biodiversity research: When is a species really
535 extinct? Testing extinction inference from a sighting record to inform conservation
536 assessment. *Diversity and Distributions*, **16**, 755–764.

- 537 Cunningham, A.A. & Daszak, P. (1998) Extinction of a species of land snail due to
538 infection with a microsporidian parasite. *Conservation Biology*, **12**, 1139–1141.
- 539 David, W.M. & Davis, R.A. (2017) Hopeful monsters—in defence of quests to rediscover
540 long-lost species. *Conservation Letters*, **10**, 382–383.
- 541 Dunn, J.C., Buchanan, G.M., Cuthbert, R.J., Whittingham, M.J. & McGowan, P.J.
542 (2015) Mapping the potential distribution of the critically endangered himalayan quail
543 *Ophrysia superciliosa* using proxy species and species distribution modelling. *Bird Con-*
544 *servation International*, **25**, 466–478.
- 545 Elith, J. & Graham, C.H. (2009) Do they? How do they? WHY do they differ? On
546 finding reasons for differing performances of species distribution models. *Ecography*,
547 **32**, 66–77.
- 548 Elphick, C.S., Roberts, D.L. & Reed, J.M. (2010) Estimated dates of recent extinctions
549 for north american and hawaiian birds. *Biological Conservation*, **143**, 617–624.
- 550 Emery-Wetherell, M.M., McHorse, B.K. & Davis, E.B. (2017) Spatially explicit analysis
551 sheds new light on the pleistocene megafaunal extinction in north america. *Paleobiology*,
552 pp. 1–14.
- 553 Fisher, D.O. & Blomberg, S.P. (2010) Correlates of rediscovery and the detectability
554 of extinction in mammals. *Proceedings of the Royal Society of London B: Biological*
555 *Sciences*, p. rspb20101579.
- 556 Fois, M., Fenu, G., Lombrana, A.C., Cogoni, D. & Bacchetta, G. (2015) A practical
557 method to speed up the discovery of unknown populations using species distribution
558 models. *Journal for Nature Conservation*, **24**, 42–48.
- 559 Grainger, M.J., Ngoprasert, D., Mc Gowan, P.J. & Savini, T. (2017) Informing decisions
560 on an extremely data poor species facing imminent extinction. *Oryx*, pp. 1–7.
- 561 Hamel, P.B. (2011) *Bachman's Warbler (Vermivora bachmanii)*, version 2.0. Cornell Lab
562 of Ornithology, Ithaca, NY.
- 563 Hijmans, R., Phillips, S., Leathwick, J. & Elith, J. (2013) *dismo*. Species distribution
564 modeling. R package version 0. 7-23. 2012.

- 565 Hilderbrand, G.V., Farley, S.D., Robbins, C.T., Hanley, T.A., Titus, K. & Servheen, C.
566 (1996) Use of stable isotopes to determine diets of living and extinct bears. *Canadian*
567 *Journal of Zoology*, **74**, 2080–2088.
- 568 Huntington, J. & Barbour, T. (1936) The birds at Soledad, Cuba, after a hurricane. *Auk*,
569 **53**, 436–437.
- 570 Jackson, J.A. (2006) Ivory-billed Woodpecker (*Campephilus principalis*): Hope, and the
571 interfaces of science, conservation, and politics. *Auk*, **123**, 1–15.
- 572 Kodikara, S., Demirhan, H. & Stone, L. (2018) Inferring about the extinction of a species
573 using certain and uncertain sightings. *Journal of theoretical biology*, **442**, 98–109.
- 574 Lee, T.E., Bowman, C. & Roberts, D.L. (2017a) Are extinction opinions extinct? *PeerJ*,
575 **5**, e3663.
- 576 Lee, T.E., Fisher, D.O., Blomberg, S.P. & Wintle, B.A. (2017b) Extinct or still out there?
577 disentangling influences on extinction and rediscovery helps to clarify the fate of species
578 on the edge. *Global change biology*, **23**, 621–634.
- 579 Lee, T.E., McCarthy, M.A., Wintle, B.A., Bode, M., Roberts, D.L. & Burgman, M.A.
580 (2014) Inferring extinctions from sighting records of variable reliability. *Journal of*
581 *Applied Ecology*, **51**, 251–258.
- 582 Lips, K.R., Diffendorfer, J., Mendelson III, J.R. & Sears, M.W. (2008) Riding the wave:
583 reconciling the roles of disease and climate change in amphibian declines. *PLoS biology*,
584 **6**, e72.
- 585 Makenov, M. (2018) Extinct or extant? A review of dhole (*Cuon alpinus* Pallas, 1811)
586 distribution in the former USSR and modern Russia. *Mammal Research*, pp. 1–9.
- 587 Menon, S., Choudhury, B.I., Khan, M.L. & Peterson, A.T. (2010) Ecological niche model-
588 ing and local knowledge predict new populations of *Gymnocladus assamicus* a critically
589 endangered tree species. *Endangered Species Research*, **11**, 175–181.
- 590 Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S.,
591 Thuiller, W., Wüest, R.O., Zimmermann, N.E. & Elith, J. (2014) What do we gain

- 592 from simplicity versus complexity in species distribution models? *Ecography*, **37**, 1267–
593 1281.
- 594 Pimiento, C. & Clements, C.F. (2014) When did *Carcharocles megalodon* become extinct?
595 a new analysis of the fossil record. *PLoS One*, **9**, e111086.
- 596 Pimiento, C., MacFadden, B.J., Clements, C.F., Varela, S., Jaramillo, C., Velez-Juarbe, J.
597 & Silliman, B.R. (2016) Geographical distribution patterns of *Carcharocles megalodon*
598 over time reveal clues about extinction mechanisms. *Journal of Biogeography*, **43**, 1645–
599 1655.
- 600 Pimm, S.L. & Askins, R.A. (1995) Forest losses predict bird extinctions in eastern north
601 america. *Proceedings of the National Academy of Sciences*, **92**, 9343–9347.
- 602 Preston, K.L., Redak, R.A., Allen, M.F. & Rotenberry, J.T. (2012) Changing distribution
603 patterns of an endangered butterfly: Linking local extinction patterns and variable
604 habitat relationships. *Biological Conservation*, **152**, 280–290.
- 605 Proosdij, A.S., Sosef, M.S., Wieringa, J.J. & Raes, N. (2016) Minimum required number
606 of specimen records to develop accurate species distribution models. *Ecography*, **39**,
607 542–552.
- 608 Rivadeneira, M.M., Hunt, G. & Roy, K. (2009) The use of sighting records to infer species
609 extinctions: an evaluation of different methods. *Ecology*, **90**, 1291–1300.
- 610 Roberts, D.L., Elphick, C.S. & Reed, J.M. (2010) Identifying anomalous reports of puta-
611 tively extinct species and why it matters. *Conservation Biology*, **24**, 189–196.
- 612 Roberts, D.L. & Solow, A.R. (2003) Flightless birds: when did the dodo become extinct?
613 *Nature*, **426**, 245–245.
- 614 Shuler, J. (1976) Three recent sight records of Bachman’s warbler. *Chat*, **41**, 11–12.
- 615 Sisson, P. (2015) Bachman’s Warbler (*Vermivora bachmanii*) 5-Year Review: Summary
616 and Evaluation. Technical report, U.S. Fish and Wildlife Service.
- 617 Smith, J.F. (2015) The Inclusion of False, Falsified, and Falsifiable Data that Favor an
618 Evolutionary Worldview in the High School Science Curriculum of Public and Private
619 Schools in the Philippines. *Christian Perspectives in Education*, **8**, 2.

- 620 Solow, A., Smith, W., Burgman, M., Rout, T., Wintle, B. & Roberts, D. (2012) Uncertain
621 sightings and the extinction of the ivory-billed woodpecker. *Conservation Biology*, **26**,
622 180–184.
- 623 Solow, A.R. (2016) On the prior distribution of extinction time. *Biology Letters*, **12**,
624 20160089.
- 625 Solow, A.R. & Beet, A.R. (2014) On uncertain sightings and inference about extinction.
626 *Conservation Biology*, **28**, 1119–1123.
- 627 Stevenson, H.M. (1972) The recent history of Bachman’s Warbler. *The Wilson Bulletin*,
628 **84**, 344–347.
- 629 Urban, M.C. (2015) Accelerating extinction risk from climate change. *Science*, **348**, 571–
630 573.
- 631 USFWS (2018) Species Profile for Bachman’s warbler (*Vermivora bachmanii*).
- 632 Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., OBrien, J.M. & Schwartz,
633 M.W. (2009) Using species distribution models to predict new occurrences for rare
634 plants. *Diversity and Distributions*, **15**, 565–576.

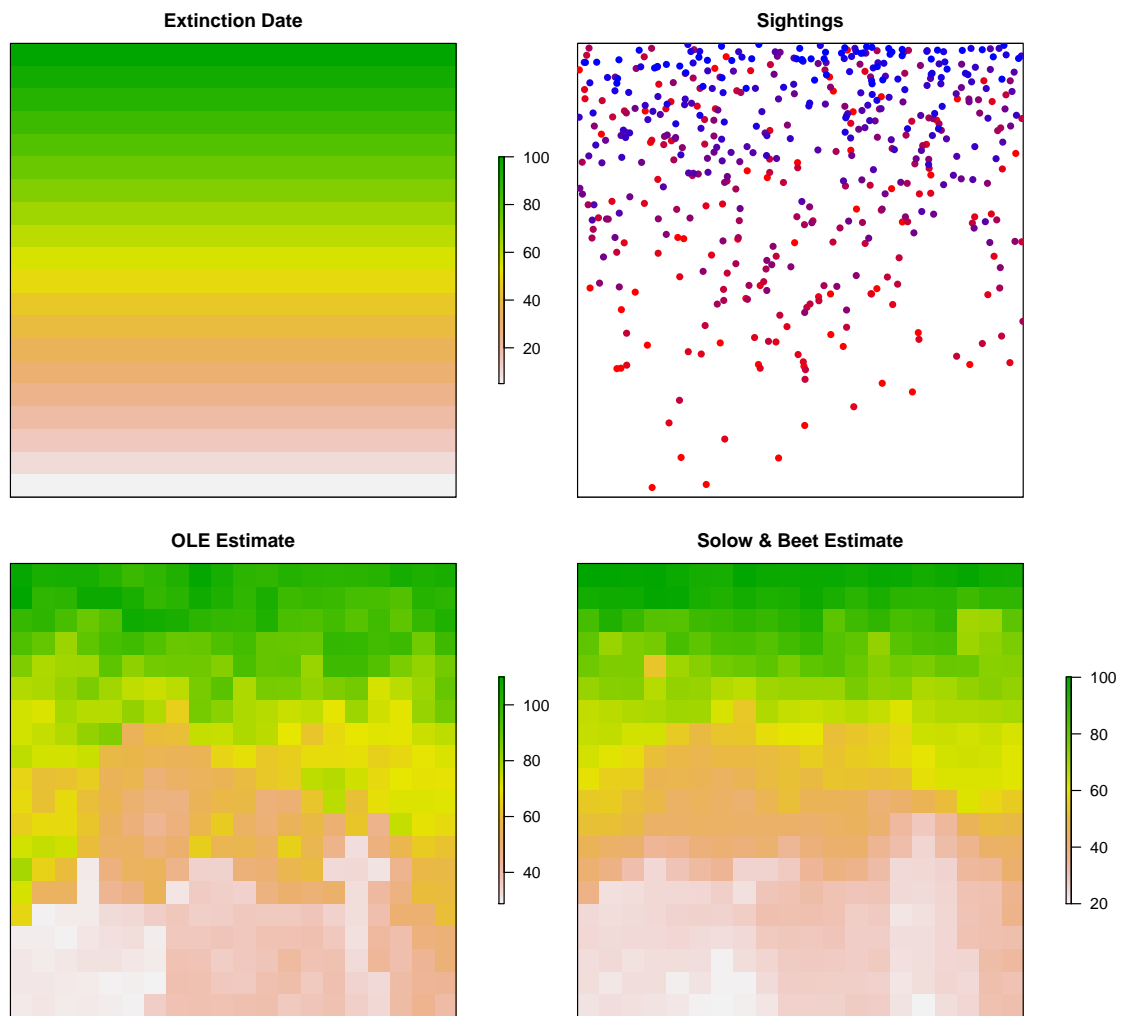


Figure 1: **An example simulation and implementation of spatial extinction date estimation.** Simulations run with 5 sightings per year over a 20 by 20 landscape. OLE and S&B models parameterized using best practices tuning described in the main text, randomized over 100 iterations.

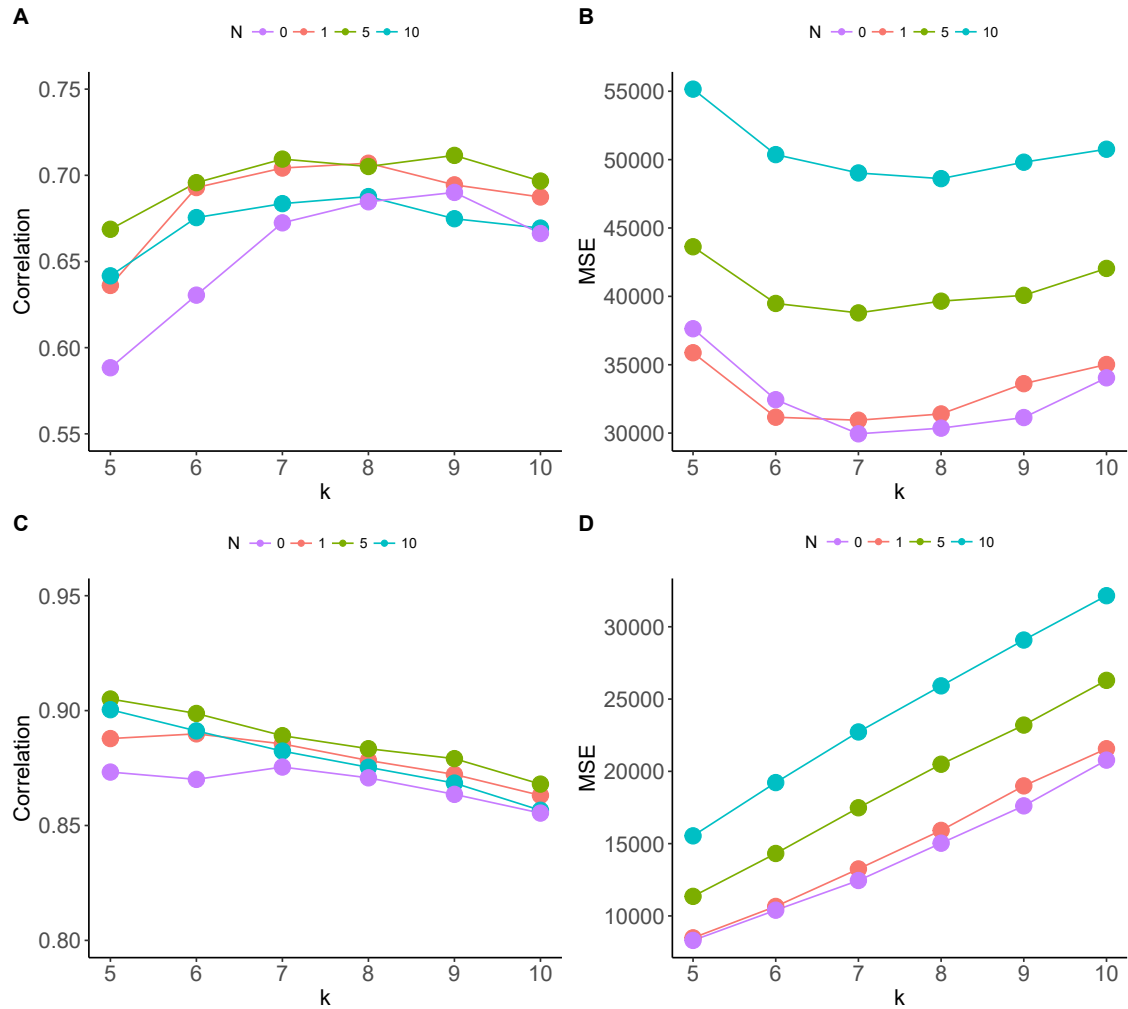


Figure 2: **Optimizing over neighborhood size.** Accuracy metrics are given for the spatial OLE (A,B) and spatial SB (C,D) models, averaged across 20 simulations for each k, N pair. We calculated the average correlation between estimated extinction surfaces and true extinction dates (A,C), and the mean squared error of those estimates (B,D), for different neighborhood sizes (k) and levels of randomization (N , where “1” indicates $N = k + 1$).

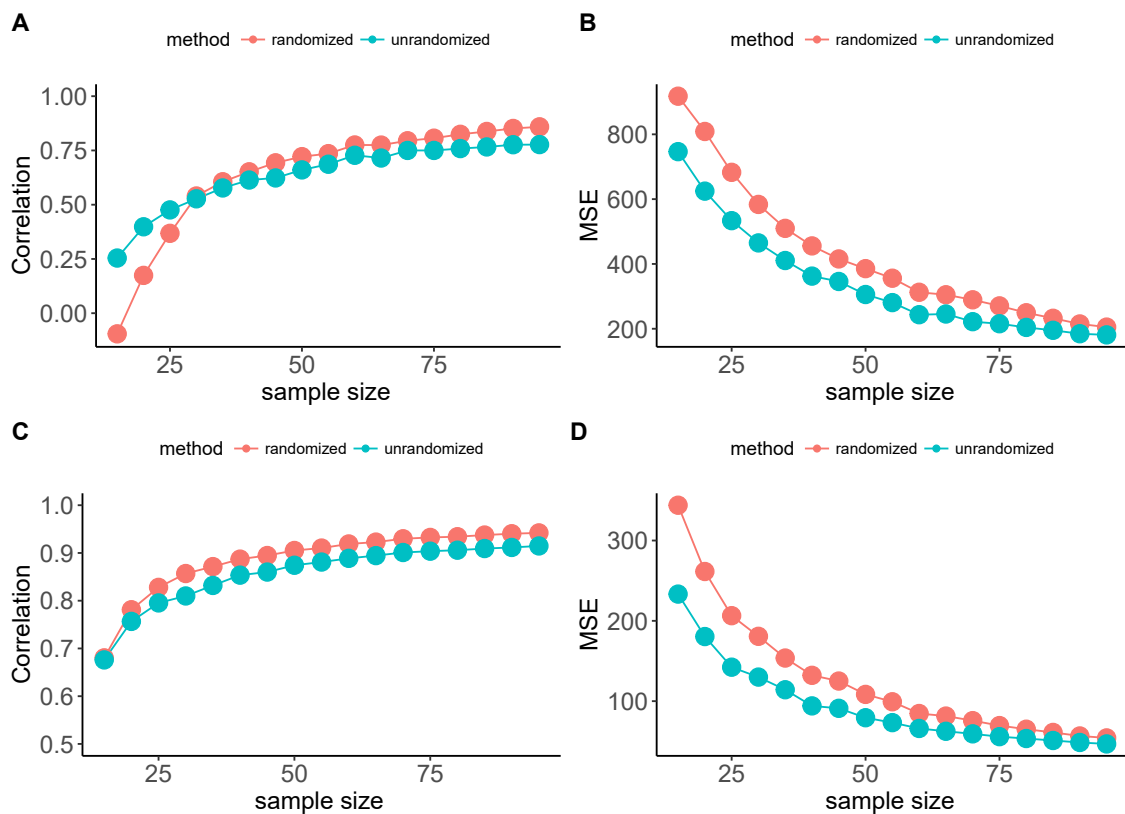


Figure 3: **Sample size improves estimates.** Results are averaged across 20 simulations for each sample size level (each with 10 resampling instances for samples within a given simulation), with neighborhood size for the OLE ($k = 7$, $N = 12$; A,B) and SB ($k = 5$, $N = 10$; C,D) models taken from the best practices determined in **Figure 2**.

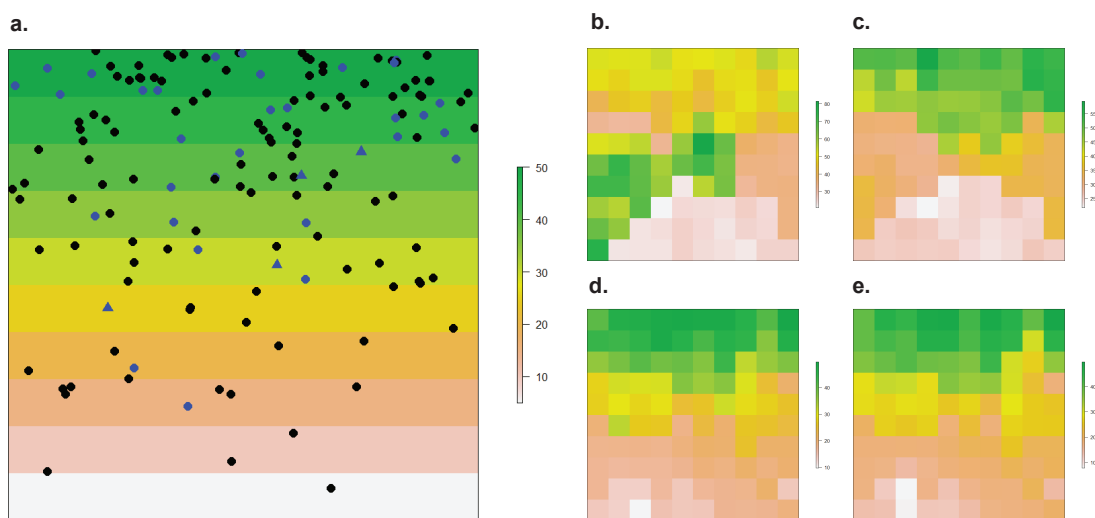


Figure 4: **Accounting for mixed-certainty, mixed-accuracy sightings.** (A) An example simulated dataset over 50 years; 3 valid sightings are recorded per year (black), 20% of which are recorded as “uncertain” (blue circles), while 4 erroneous sightings are recorded total (blue triangles), with randomly assigned dates. Spatial OLE ($k = 7$, $N = 12$) without quality control (B) is highly prone to misinterpretation, versus spatial OLE with quality control (C), which noticeably improves predictions but also limits data available for use. Solow & Beet’s model ($k = 5$, $N = 10$) includes mixed certainty data and performs noticeably better, regardless of selecting model 1 (D) or model 2 (E).

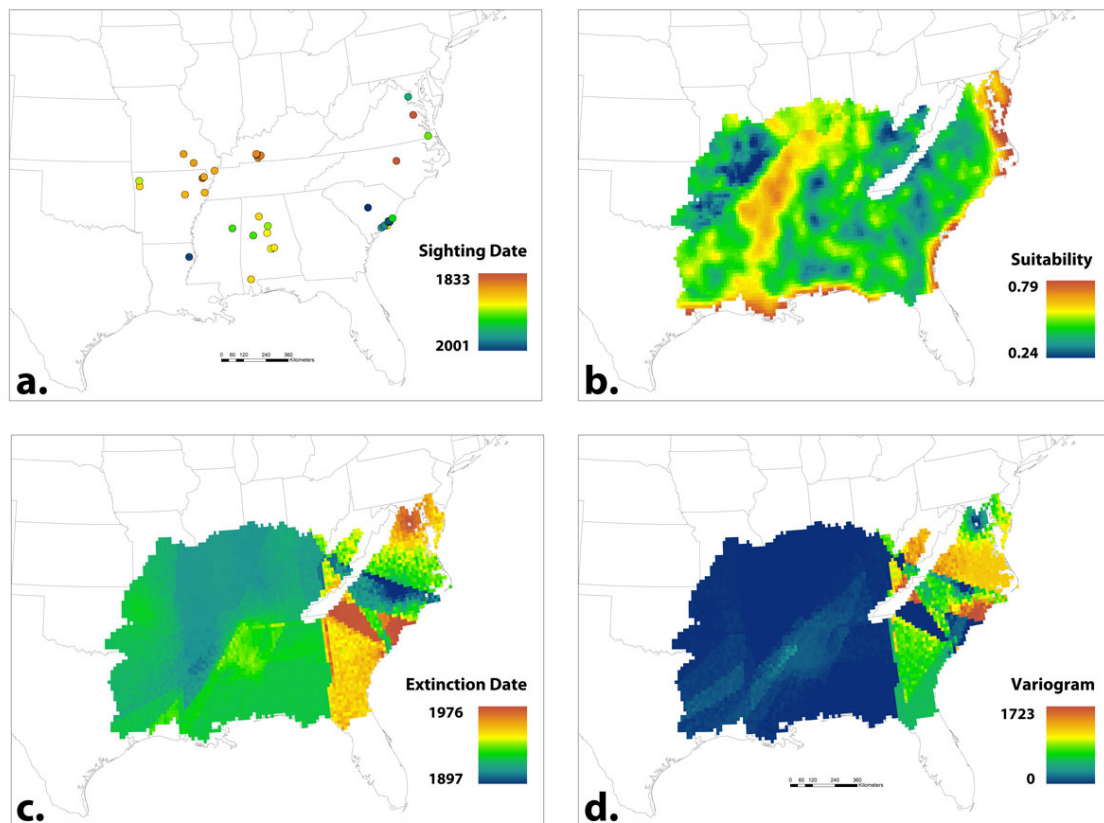


Figure 5: **Mapping Bachman's Warbler with limited occurrence data.** (A) Sighting records ($n = 118$). (B) MaxEnt ecological niche model. (C) Extinction date reconstructed using Solow & Beet's model 2, on recommended best practices settings. (D) Variance in Solow & Beet estimates (higher variance means more uncertain estimates).

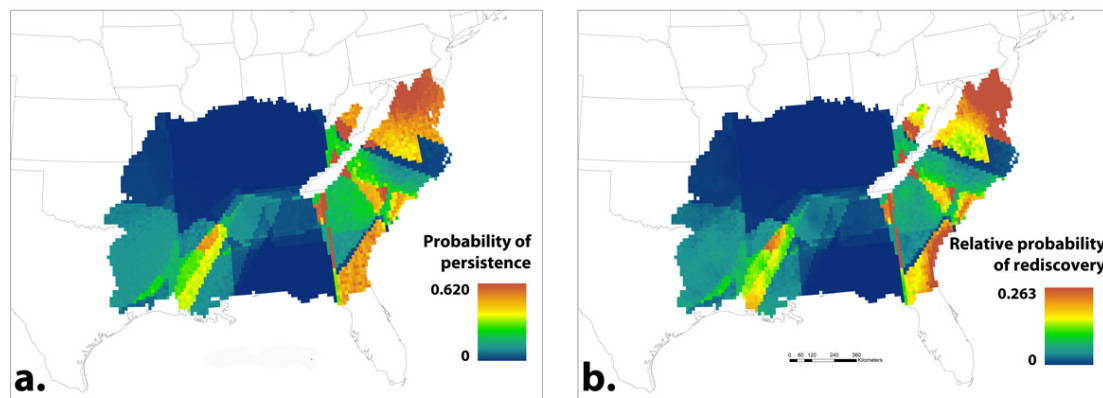


Figure 6: **Where are the odds of rediscovery highest?** (A) Probability of persistence based on Bayesian hypothesis testing in Solow & Beet model (extinction date only). (B) Combined probability of rediscovery (product of probability of persistence from EDE and environmental suitability from ENM).