

DR. JÉRÔME GUÉLAT (Orcid ID : 0000-0003-1438-4378)

Article type : Research Article
Editor : Dr. Nick Isaac

Effects of spatial autocorrelation and imperfect detection on species distribution models

Jérôme Guélat^{1*} & Marc Kéry¹

24 January 2018

¹ Swiss Ornithological Institute, 6204 Sempach, Switzerland

* Correspondence author: jerome.guelat@vogelwarte.ch

Running title: Spatial autocorr. and detection error in SDMs

Summary

1. Species distribution models (SDMs) are widely used in ecology and related fields. They are frequently adopted to predict the expected occurrence (presence/absence) or abundance over large spatial scales, i.e., to produce a species distribution map. Two issues that almost universally affect these models are measurement errors (especially imperfect detection) and residual spatial autocorrelation.
2. We explored the effects of imperfect detection and autocorrelation in abundance models by simulating datasets which did or did not contain these two effects and

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.12983

This article is protected by copyright. All rights reserved.

analysing them with four different models that did or did not accommodate them.

Specifically, we used a Poisson GLM as a baseline, an N-mixture model accounting only for imperfect detection, and two N-mixture models that accounted for imperfect detection, but differed in their specification of spatial autocorrelation (CAR random effects vs. twodimensional splines).

3. In a case study, we then applied these models to Common Redstart (*Phoenicurus phoenicurus*) data from the second Swiss Breeding Bird Atlas (1993–1996) and validated them using an independent monitoring dataset.
4. We found that both imperfect detection and autocorrelation strongly affected the quality and the uncertainty of abundance maps, especially when they occurred together. Spatial N-mixture models were well able to estimate the true abundance maps. Explicit modeling of measurement error *and* spatial autocorrelation can thus greatly improve the quality of SDMs and should not be ignored when producing large-scale abundance or occurrence maps.

Keywords: abundance, CAR, conditional autoregressive model, distribution, generalized additive models, imperfect detection, measurement error, spatial autocorrelation, species distribution model

Introduction

Distribution and abundance, their variation in space and time, and the underlying causes for spatiotemporal patterns have always fascinated scientists in ecology and related fields (Krebs 2001). Beyond identifying mechanisms that govern variation in occurrence and abundance, there is often an interest in reliably predicting the distribution under certain

environmental or other conditions, for instance to forecast a process in time or to extrapolate it in space in the form of a map. Such predictions may help achieving a better understanding of biological mechanisms, are essential to represent distribution patterns, and can be useful for scientific conservation planning. For the latter two goals, it is typical to deal with large spatial domains (Stokes *et al.* 2010; Chakraborty *et al.* 2010; Martin *et al.* 2015). In addition to obtaining valid point estimates of occurrence probability or abundance, an honest assessment of the uncertainty associated with such predictions is an important, but often neglected part of the modeling of distribution and abundance (Barry & Elith 2006). Presenting not only an estimate, but also the uncertainty around it, is an essential ingredient for informed decision-making in the management based on such models (Cressie *et al.* 2009).

The statistical modeling of presence/absence (distribution) and abundance is a fairly mature field in ecological statistics. The typical parametric statistical model for a species distribution is some variant of a logistic regression (Hosmer *et al.* 2013). Similarly, the typical model for abundance comes in form of a Poisson regression (Poisson GLM) or one of its variants (Cameron & Trivedi 1998; Hilbe 2014).

Measurement error and spatial autocorrelation are two very common challenges in the modeling of distribution and abundance. They have not received much attention in applied studies, although they are likely present in most ecological data sets collected in the field (but see Chelgren *et al.* 2011; Johnson *et al.* 2013; Broms *et al.* 2015). By measurement error we mean errors of omission ("false absences") and of commission ("false presences") when modeling occurrence and under- or overcounting when modeling abundance (Kéry & Royle 2016). These errors arise from imperfect detection, from the multiple counting of the same individual or from species mis-identification.

Imperfect detection is especially ubiquitous in ecological datasets of distribution and abundance (Royle *et al.* 2007; Kéry & Schmidt 2008; Kellner & Swihart 2014). Spatial autocorrelation denotes the situation where the values of a variable are correlated for sites at nearby locations (Legendre 1993; Dormann 2007; Dormann *et al.* 2007). The value of the variable at one site can thus be partially predicted by the values at neighboring sites. Spatial autocorrelation in species distributions typically occurs because the preferred habitat is itself autocorrelated or because of biotic processes such as dispersal, conspecific attraction, competition with another species or other complex (e.g., source-sink) dynamics. *Residual* spatial autocorrelation is what is usually meant by "spatial autocorrelation" and denotes the situation where the residuals of a model are spatially structured *after* covariate effects have been accounted for. When modeling species distributions, the presence of such (residual) spatial autocorrelation is often an indication that an important covariate was not included in the model or that the model was misspecified in some other way.

Failing to accommodate imperfect detection in distribution or abundance models leads to an underestimation of the true occurrence probability or of true abundance, respectively, and of their associated uncertainty. It may also lead to biased parameter estimates if some covariates affect both the ecological state (distribution or abundance) and its measurement (our data) (Kéry *et al.* 2010; Kéry 2011; Lahoz-Monfort *et al.* 2014). In spite of this, some researchers have recommended ignoring this challenge (Johnson 2008; Banks-Leite *et al.* 2014; Stephens *et al.* 2015). Hence, there is no consensus about the importance of accommodating this sort of measurement error in SDMs.

The effects of failing to accommodate spatial autocorrelation have been debated quite vigorously recently (Diniz-Filho *et al.* 2003; Dormann 2007; Beale *et al.* 2007, 2010;

Bini *et al.* 2009; Hawkins 2012; Thibaud *et al.* 2014). Researchers agree that failure to account for autocorrelation will lead to an overstatement of the precision of the estimates. However, in some cases an inversion of the slope estimate of the relationships was observed (Kühn 2007; Bini *et al.* 2009) and predictions therefore become very misleading. Hence, there is again no consensus about whether modeling residual spatial autocorrelation in SDMs is worthwhile.

Binomial N-mixture models represent a modeling framework which appears particularly interesting for the large-scale joint modeling of abundance and imperfect detection (Royle 2004; Dail & Madsen 2011). For occurrence, the analogous model is called a site-occupancy model (MacKenzie *et al.* 2002, 2003; Tyre *et al.* 2003). In both models, the information about distribution or abundance measurement error comes from repeated measurements of presence/absence (i.e., detection/non-detection observations) or of abundance (i.e., counts). Site-occupancy models for presence/absence data have been increasingly used as a powerful framework for SDMs (Altwegg *et al.* 2008; Kéry *et al.* 2010; Kéry 2011; Lahoz-Monfort *et al.* 2014; Guillera-Arroita *et al.* 2015). In contrast, for abundance modeling N-mixture models have been used much less commonly, although they could be adopted for many data sets collected in large-scale monitoring programs (Royle *et al.* 2007).

Many approaches have been developed to account for residual spatial autocorrelation (for recent overviews, see Dormann *et al.* 2007; Beale *et al.* 2010). Two methods appear particularly interesting, since they are easy to implement even for moderately large datasets, are particularly flexible, and have performed well in complex simulations (Beale *et al.* 2010): conditional autoregressive (or CAR) models (Besag *et al.* 1991), and geoadditive models (Kammann & Wand 2003) which use splines to model the

spatial structure.

In SDMs, there seems to be considerable uncertainty about the importance and the consequences of measurement errors and spatial autocorrelation. A formal assessment of models accommodating both, and a comparison with models that do not, has never been undertaken. We therefore investigated the consequences of imperfect detection and residual spatial autocorrelation on the performance of SDMs for estimating abundance using both simulations and real data.

Methods

Data simulation

To investigate the consequences of imperfect detection and spatial autocorrelation we simulated replicated datasets where we varied the following four factors: strength of spatial autocorrelation, detection probability, average abundance and sample size. We used R 3.1.0 (R Core Team 2014) to generate datasets using the following general model within a 50x50 cell landscape:

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta x_i + \gamma \rho_i$$

$$\rho_i \sim \text{MVN}(0, \sigma^2 e^{-\theta d_{ij}})$$

$$C_{it} \sim \text{Binomial}(N_i, p).$$

The first random process generates the true latent abundance at each site i , N_i , x_i is a continuous covariate, ρ_i is a spatially correlated random effect generated using an exponential correlation function, and γ is the strength of the autocorrelation. Further, σ^2 is the spatial variance, θ is the scale parameter governing the distance-dependent decay of the spatial correlation in the expected abundance and d_{ij} is the distance between sites

i and *j*. In other words the covariance matrix of the multivariate normal distribution determines the distance based strength of pairwise correlations in the landscape. The second random process is linked with the outcome of the first (i.e., is conditional on N_i) and produces the observed data, the counts C_{it} at site *i* during visit *t*. This second process describes the abundance measurement error and is governed by the per-visit detection probability *p*. By fixing some parameters at zero or at one we obtain important restricted versions of this model. For instance, by setting $\gamma = 0$ we obtain an N-mixture model without spatial correlation (Royle 2004) and by setting $p = 1$, we obtain the standard Poisson GLM, which is the basis for most abundance modeling in ecology.

We generated datasets under three scenarios: The first scenario assumed perfect detection and no spatial autocorrelation. We thus generated the data by setting $p = 1$ and $\gamma = 0$, which resulted in identical counts for every visit. The second scenario assumed imperfect detection, but no spatial autocorrelation. We thus generated the data by setting $\gamma = 0$ and using two different values for *p* (0.3 and 0.8). The third scenario assumed both imperfect detection (with $p = \{0.3, 0.8\}$) and spatial autocorrelation of varying strength (with $\gamma = \{0.5, 1.5\}$).

In addition, for each scenario we simulated two sample sizes by randomly choosing 200 or 1,000 among the 2,500 cells of our simulated 50x50 cell landscape. We also chose two abundance classes for *N*, "rare" or "common", corresponding to an average abundance of around 2,500 and 30,000 individuals for the whole lattice, respectively (the total abundance is a stochastic Poisson realization). We simulated three visits for each sampled site. In summary, this design resulted in a total of 28 cases for which we simulated 100 datasets each. See Appendix 1 for a graphical depiction of our simulations and the Data Accessibility section for the accompanying R code.

Data analysis

To evaluate the effects of spatial autocorrelation and imperfect detection on the quality of abundance predictions, we used four models which differed in whether or not, and how, they accommodated imperfect detection and spatial autocorrelation (Table 1).

Model 1: To provide a baseline, we used as our first model a standard Poisson GLM that completely ignores imperfect detection and spatial autocorrelation. This model is widely used in the literature to make abundance predictions (Potts & Elith 2006; Zuur *et al.* 2009). We fitted the following model to the maximum count (C_i) at site i :

$$C_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta x_i$$

where x is the covariate used in the data generating process.

Model 2: Our second model was an N-mixture model (Royle 2004) which accounts for imperfect detection but not for spatial autocorrelation.

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta x_i$$

$$C_{it} \sim \text{Binomial}(N_i, p)$$

where N_i is the true abundance at site i , C_{it} are the counts for site i and replicate t , and p is a constant detection probability.

Model 3: This is an extension of the N-mixture model that includes spatial autocorrelation in the form of an intrinsic conditionally autoregressive (CAR) model (Besag *et al.* 1991):

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta x_i + \delta_i$$

$$\delta_i \sim \text{Normal}\left(\frac{\sum_{j=1}^m \delta_j}{m}, \frac{\tau^2}{m}\right)$$

$$C_{it} \sim \text{Binomial}(N_i, p)$$

where δ_i is the conditional autoregressive random effect for site i , m is the number of neighbors of site i , δ_j are the values of δ for the neighbors of site i and τ^2 is the spatial variance parameter. Since we simulated our data on a 50*50 cell lattice, we defined a neighborhood by the eight cells touching cell i (except for edge cells which have fewer neighbors). Spatial autocorrelation is thus modeled as a discrete process. We refer to this model as the "CAR model".

Model 4: This was another extension of the N-mixture model that accounts for spatial autocorrelation using semi-parametric geoaddivitive models (Kammann & Wand 2003).

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta x_i + f(s_i)$$

$$C_{it} \sim \text{Binomial}(N_i, p)$$

Here, s_i is the location of site i (identified by its coordinates), and $f()$ is a 2-dimensional cubic spline function. We used 25 knots for the spline, regularly spaced in the 50*50 cells lattice. This model can be written as a generalized linear mixed model (Ruppert *et al.* 2003; Crainiceanu *et al.* 2005; Péron *et al.* 2011) and we refer to it as "the splines model".

In contrast to the CAR model, spatial autocorrelation is modeled as a continuous process.

We call the first two models *nonspatial* and the last two *spatial* since the latter two models account for spatial autocorrelation. We fitted all models with Bayesian inference using the BUGS language in program JAGS (Plummer 2003) except for model 3, for which we used OpenBUGS (Lunn *et al.* 2009), because JAGS has no inbuilt CAR functionality. We ran three chains with 10,000 iterations each and discarded the first 5,000 iterations as a burnin. Convergence was assessed using the Brook-Gelman-Rubin

statistic (R-hat). If the R-hat of at least one of the main structural parameters was greater than 1.1, we generated another dataset using the same simulation scenario and refitted the model and did so until we had accumulated a sample size of 100 for all 28 cases.

Validation

First, we fitted all four models to each of the 2,800 simulated datasets and obtained predictions of the expected response ($\hat{\lambda}$) and the associated 95% credible intervals (CRI) for the entire 2,500 cell-lattice landscape (see Fig. 1 for example). Since we knew the true expected abundance λ for each cell, we computed three validation statistics on the whole lattice for each dataset/model combination.

First was the Pearson correlation coefficient r , which measures the association between the truth and the predictions. A perfect correlation ($r = 1$) doesn't imply perfect predictions; all predictions may simply be biased by the same relative amount. Hence r measures the degree to which a model gets the patterns right.

Our second statistic was the relative bias:

$$\frac{\frac{1}{n} \sum_{i=1}^n (\hat{\lambda}_i - \lambda_i)}{\bar{\lambda}_i}$$

where the summation is over $n=2,500$ cells in the lattice, $\hat{\lambda}_i$ is the predicted and λ_i the true abundance in cell i , and $\bar{\lambda}_i$ is the landscape-wide arithmetic mean of true abundance. We did this latter standardization to express bias relative to the average abundance in the landscape, which facilitated comparison of models with different abundance levels.

Third, we computed the coverage rate as the percentage of the cells for which the true value λ_i was inside of the predicted 95% CRI.

Case study

We fitted the same models to a dataset comprising counts of the Common Redstart (*Phoenicurus phoenicurus*) from the Swiss Breeding Bird Atlas 1993-1996 (Schmid *et al.* 1998), where 2,716 one km squares were surveyed three times during 15 April - 30 June. Unfortunately, only the total number of territories detected is available. Hence, to obtain the required replicated counts, we assumed the total count was N and simulated three counts as binomial random numbers with detection probability of 0.5 (based on our own unpublished analyses).

We used the following standardized covariates (obtained from the Federal Offices of Topography and of Statistics) in all models: elevation (linear and quadratic effects), forest cover, pasture cover, orchard cover, slope (linear and quadratic effects) and northness (linear and quadratic effects). For model 4, we used 100 knots regularly spaced across Switzerland to compute the spline. We ran three chains with 20,000 iterations (30,000 for model 3) and discarded the first 10,000 iterations (20,000 for model 3) as a burnin. Convergence was assessed visually and using the Brook-Gelman-Rubin statistic (\hat{R}).

For validation, we drew the predicted counts \hat{N}_i from a Poisson distribution with expectation $\hat{\lambda}_i$ for each of the about 42,000 square kilometers of Switzerland and compared them to data from a different monitoring program which started in 1999 and which has an identical data collection protocol, the Swiss Breeding Bird Survey MHB (Schmid *et al.* 2002). At each site, we averaged the counts between 1999 and 2006 to yield data for 546 square kilometers randomly distributed across Switzerland. The quality of the abundance predictions was evaluated using the same statistics as above, except that we used absolute rather than relative bias (no division by $\bar{\lambda}_i$).

Results

Simulation study

We highlight the key results only and provide the full results as supplementary information (Appendix 2).

Pearson correlation coefficient: Correlation was high for all models when fit to data without spatial autocorrelation (Fig 2 a, b). Hence, in the absence of spatial autocorrelation, a model that ignores imperfect detection (i.e., the Poisson GLM) predicted spatial *patterns* in abundance about as well as did the models that accounted for it (but see the section on bias below).

However, when spatial autocorrelation was added in the data, the correlation decreased sharply for the two nonspatial models (models 1 and 2). In contrast, models 3 and 4 performed equally well in terms of pattern recognition (Fig. 2 c, d). Interestingly, the two spatial models (3 and 4) could not completely account for a strong spatial autocorrelation (Fig. 2 d, e). They appeared to oversmooth since they lack a parameter for the *strength* of the autocorrelation. Nevertheless, they still greatly outperformed the two nonspatial models. With increased sample size, the correlation remained the same for the two nonspatial models, but it strongly increased for the two spatial models. Model 3 performed slightly better than model 4 (Fig. 2 d, e). The abundance level did not seem to affect the correlation (Appendix 2).

Relative bias: Predictions appeared unbiased as long as there was no spatial autocorrelation in the data and detection probability was high ($p = 0.8$ per survey, which

translates to $1 - (1 - 0.8)^3 = 0.96$ for all surveys combined) (Fig. 3 a). With low detection probability ($p = 0.3$), predictions were strongly biased low under the Poisson GLM (model 1) which assumes perfect detection implicitly (Fig. 3 b).

For a rare species with low spatial autocorrelation we found hardly any differences between N-mixture models without spatial autocorrelation (model 2) and the two models with it (models 3 and 4) (Fig. 3 c). In contrast, for a common species, predictions produced by model 2 were also biased low (Fig. 3 d).

When high magnitude of spatial autocorrelation was combined with low detection probability in the data, the benefits of accounting for imperfect detection became minimal unless we also accounted for spatial autocorrelation (models 1 and 2 yielded predictions with a similar negative bias). Even the predictions of model 4 had a slight negative bias and only model 3 produced unbiased predictions (Fig. 3 e). All models yielded relatively unbiased predictions when detection probability and spatial autocorrelation were high (Fig. 3 f). Effect of sample size on the bias was negligible (Appendix 2).

Coverage: -Credible interval coverage was very low when imperfect detection was present in the data but unaccounted for (model 1) (Fig. 4 a). With spatial autocorrelation in the data, the nonspatial N-mixture model (model 2) also strongly underestimated the prediction uncertainty (Fig. 4 b, c, d). Model 4 yielded an improved coverage rate, but did not quite achieve the nominal 95%. In contrast, model 3 seemed to overestimate the prediction uncertainty and coverage was higher than the nominal 95% (Fig. 4 b). With increased strength of spatial autocorrelation, the nonspatial N-mixture model (model 2) had about the same coverage as the Poisson GLM. Model 4 did better but didn't quite

achieve 95%. The coverage of model 3 was still slightly higher than the nominal 95% (Fig. 4 c).

With spatial autocorrelation in the data, high detection probability yielded a lower coverage rate for model 2 than did low detection probability (Fig. 4 b, d). The coverage rates of the other models were less affected by the magnitude of detection probability. Interestingly, for the large sample size, coverage rates were lower than the for the small sample size for all models except for model 3 (Fig. 4 b, e).

Case study

Nonspatial models tended to produce more homogeneous maps with less fine-grained patterns (Fig. 5 a, b). Maps under the two nonspatial models were similar because of the relatively high per-visit detection probability of the Common Redstart, which translates into 0.88 for the three surveys combined. The two spatial models produced much patchier maps (Fig. 5 c, d). According to Swiss bird experts (e.g., our esteemed colleagues Hans Schmid, Peter Knaus, Niklaus Zbinden), the spatial models produced the most satisfactory maps in terms of spatial patterns.

Validation also showed that the two spatial models produced better maps than did the two nonspatial models (Pearson correlation coefficient: model 1: 0.33, model 2: 0.34, model 3: 0.68, model 4: 0.72). The bias of maps was similar for all models, but slightly lower for model 3 (model 1: -0.29, model 2: -0.24, model 3: -0.17, model 4: -0.23). Coverage rates were close to nominal for all models (model 1: 0.93, model 2: 0.93, model 3: 0.98, model 4: 0.96).

Prediction uncertainty for $\hat{\lambda}_i$ was very small for the nonspatial models and for the model 4, but extremely large for model 3 (Fig. 6).

Discussion

We used simulation and real data to compare spatial predictions of abundance under four models that did or did not account for two common characteristics of abundance data: measurement error (in the form of imperfect detection) and residual spatial autocorrelation. Simulations showed that both imperfect detection and spatial autocorrelation can dramatically affect the quality of the abundance maps and their uncertainty, especially when they occur together. An explicit modeling of both mechanisms in spatial N-mixture models with CAR or spline spatial processes significantly improved the maps and, at the same time, yielded a more honest assessment of the estimation uncertainty (Cressie *et al.* 2009). However, more care should be taken with the interpretation of the model coefficients since covariates may be collinear with the spatial term (Hodges & Reich 2010; Hanks *et al.* 2015).

This may be the first study of the combined consequences of spatial autocorrelation and imperfect detection on the quality of spatial predictions. Most previous studies have investigated the effects of these two processes separately. They focused on the accuracy of regression parameters, not on the quality of the predictions, i.e., *not* on the resulting species distribution map, although the map is often what we are most interested in. For example, the studies on the consequences of imperfect detection mostly focused on the bias of the model coefficients (e.g., Kéry *et al.* 2010; Lahoz-Monfort *et al.* 2014). Similarly, most studies on the effects of spatial autocorrelation simply compared the coefficients and the uncertainty assessments of spatial vs. non-spatial models, but did not directly look at the maps produced (Diniz-Filho *et al.* 2003; Dormann 2007; Bini *et al.* 2009; Beale *et al.* 2010). Our study shows that spatial autocorrelation and

imperfect detection can both have severe consequences for the quality of species distribution maps.

Most of the following discussion is based on our results for our abundance models, but presumably measurement error and spatial autocorrelation affect occurrence models in an exactly analogous way.

Large scale abundance modeling

We fitted two abundance models that accounted both for spatial autocorrelation and imperfect detection and hence provided reliable predictions. We were surprised to be able to fit these fairly complex spatial models and to produce such abundance maps even for relatively large datasets and with very reasonable computation time on a desktop computer. Such large-scale maps are particularly needed in conservation science (e.g., Stokes *et al.* 2010; Martin *et al.* 2015). We think that it is good news that they can actually be fit in practice using off-the-shelf Bayesian modeling software (i.e., OpenBUGS, JAGS), which make sophisticated model fitting accessible to many ecologists. It would also be possible to increase the size of the datasets by using approximations (i.e., dimension reduction) to model the spatial autocorrelation (see Latimer *et al.* 2009; Chakraborty *et al.* 2010; Johnson *et al.* 2013, although only the latter also accounts for measurement error explicitly). However, these may presumably no longer be fit in BUGS, which means that they will be much less accessible to ecologists unless they are implemented in specific software or R packages.

The two spatial models yielded very similar abundance predictions, even if their structure is different. However we found large differences when comparing their

uncertainty estimates. The CAR model (model 3) provided 95% CRIs that were too wide and their coverage rate was very close to 1. This pattern was confirmed in the case study where the uncertainties were so large that the predictions were hard to interpret. For the spline model (model 4), the uncertainty was slightly underestimated. It would be interesting to investigate the influence of knot number and position on this result. Computation time can also be important in practice, especially in a simulation experiment. The spline model (model 4) was much faster to fit than the CAR model (model 3), and convergence was easier to achieve.

Effects of imperfect detection and spatial autocorrelation on species distribution maps

Arguably, imperfect detection occurs in almost every field-based dataset on distribution and abundance. Over the last decade, several methods have been developed to account for this kind of measurement error both when modeling occurrence or abundance data. These methods are also applicable for large spatial domains (MacKenzie *et al.* 2002, 2003; Tyre *et al.* 2003; Royle 2004; Royle *et al.* 2007; Kéry & Royle 2016). If researchers neglect imperfect detection in their models of distribution and abundance, they may get maps which are biased low with respect to the true presence/ absence and abundance, i.e., they will only map *relative* distribution or *relative* abundance. That is, the mapped quantity is the product of occurrence probability or abundance, respectively, and detection probability (Kéry *et al.* 2010; Kéry 2011; Lahoz-Monfort *et al.* 2014; Guillera-Arroita *et al.* 2014). This may look like a small problem if we want to use the predictions in a conservation context since we are on the conservative side. Note, however, that our simulations represented the simplest-case scenario, where imperfect detection was constant across the landscape. The consequences of ignoring imperfect detection will be

more severe if a covariate has an effect on both distribution or abundance and on detection probability, especially if these effects have a different sign (Kéry 2008). In this case a model not accounting for imperfect detection may for instance say that some habitat covariate has no effect on the occurrence or on the abundance of a species, where in fact the effect is simply "wiped out" in the observed data by the measurement error which goes the opposite direction (Kéry *et al.* 2010; Lahoz-Monfort *et al.* 2014). One can of course improve the predictions by using data coming from multiple visits. However, if repeated measurements are available then it is hard to argue against using a model that explicitly accounts for measurement errors.

Quite likely, modeling of most spatial ecological dataset results in at least some residual spatial autocorrelation. This may be due to an important environmental covariate that is not included in the model or to biotic processes such as competition, dispersal, or conspecific attraction that are not accounted for (the latter being rarely done in SDMs). As we have seen in the simulations, even massively increasing the sample size did not fix this problem hence, the problem will not go away with "big data" (Fig. 2). The predictions will be wrong since an important mechanism is omitted in the model. Some authors have claimed that spatial autocorrelation is not a big problem in ecology and especially in SDMs (Hawkins 2012; Thibaud *et al.* 2014). Our results contradict this claim and illustrate how a species distribution map can be severely erroneous if spatial autocorrelation is not properly modeled (see Fig. 1 and Fig. 6). Let's imagine that some conservation measures would be taken based on the predicted abundance maps for the Common Redstart. The true abundance hotspots (e.g., in central and southern Switzerland) would be totally missed by a nonspatial model.

In our simulations, the relative bias was exacerbated by the simultaneous effects

of low detection probability and strong spatial autocorrelation. The modeling of only one of the two did not seem to help much, since the resulting maps were often equally biased as were the ones coming from a simple Poisson GLM which ignores both measurement error and autocorrelation. Only simultaneous modeling of both mechanisms improved the quality of the distribution maps substantially.

Recommendations

It may not always be possible or required to model imperfect detection and spatial autocorrelation. However, we recommend accounting for imperfect detection *if* one has access to replicated detection/nondetection observations or replicated counts, even if such repeated measurements are only available from some of the sites. Concerning spatial autocorrelation, we recommend performing adequate model assessments, especially looking for spatial trends in the residuals using tools like variogram or spatial correlogram. Several packages in R (R Core Team 2014) are available to perform such tasks (e.g., `gstat` and `ncf`). If the residuals are spatially structured, then we advise looking for other important covariates or trying to add some biotic realism in some way in order to properly remove residual spatial autocorrelation. Unfortunately, this may not always be possible, either because the important covariates and biotic mechanisms are unknown or because it would be too expensive to obtain this information. In this case the explicit modeling of spatial autocorrelation will vastly improve the quality of the abundance predictions.

Finally, we'd like to emphasize the importance of using several validation measures which assess different aspects of a model prediction. As seen above the Pearson correlation coefficient is not affected if all predictions are biased in the same

direction and by the same relative amount. This can happen if detection is imperfect but constant in space and we do not account for it. Focusing only on the Pearson correlation coefficient would show a good predictive ability of our model even though all predictions are biased down. This can be a serious problem, for example if the model is used to plan some conservation actions.

Supporting information

Appendix 1: Graphical overview of the scenarios used in the simulation study

Appendix 2: Detailed simulation results

Acknowledgements

This work was funded by the Swiss National Science Foundation (grant number 31003A_146125 to M.K. and Michael Schaub). We thank Michael Schaub, Gilberto Pasinelli and two anonymous reviewers for useful comments on this manuscript.

Authors' Contributions

J.G. and M.K. conceived the ideas and designed the methodology. J.G. designed the simulations and analysed the data. J.G. and M.K. wrote the paper.

Data Accessibility

R code to generate the simulated datasets:

<https://doi.org/10.6084/m9.figshare.5849841.v1>

R and BUGS code for the four models used in the simulations:

This article is protected by copyright. All rights reserved.

<https://doi.org/10.6084/m9.figshare.5849862.v2>

R and BUGS code for the Common Redstart case study:

<https://doi.org/10.6084/m9.figshare.5849868.v1>

Common Redstart dataset: <https://doi.org/10.6084/m9.figshare.5722432>

Coordinates of knots used for model 4: <https://doi.org/10.6084/m9.figshare.5827653.v1>

References

- Altwegg, R., Wheeler, M. & Erni, B. (2008). Climate and the range dynamics of species with imperfect detection. *Biology Letters*, **4**, 581–584.
- Banks-Leite, C., Pardini, R., Boscolo, D., Cassano, C.R., Püttker, T., Barros, C.S. & Barlow, J. (2014). Assessing the utility of statistical adjustments for imperfect detection in tropical conservation science. *Journal of Applied Ecology*, **51**, 849–859.
- Barry, S. & Elith, J. (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.
- Beale, C.M., Lennon, J.J., Elston, D.A., Brewer, M.J. & Yearsley, J.M. (2007). Red herrings remain in geographical ecology: a reply to Hawkins et al. (2007). *Ecography*, **30**, 845–847.
- Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J. & Elston, D.A. (2010). Regression analysis of spatial data. *Ecology Letters*, **13**, 246–264.
- Besag, J., York, J. & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1–20.
- Bini, L.M. et al. (2009). Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography*, **32**, 193–204.
- Broms, K.M., Hooten, M.B., Johnson, D.S., Altwegg, R. & Conquest, L.L. (2015). Dynamic occupancy models for explicit colonization processes. *Ecology*, **In Press**.
- Cameron, A.C. & Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, New York.
- Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M. & Silander Jr, J.A. (2010). Modeling large scale species abundance with latent spatial processes. *The Annals of Applied Statistics*, **4**, 1403–1429.
- Chelgren, N.D., Adams, M.J., Bailey, L.L. & Bury, R.B. (2011). Using multilevel spatial

models to understand salamander site occupancy patterns after wildfire. *Ecology*, **92**, 408–421.

- Crainiceanu, C.M., Ruppert, D. & Wand, M.P. (2005). Bayesian Analysis for Penalized Spline Regression Using WinBUGS. *Journal of Statistical Software*, **14**, 1–24.
- Cressie, N., Calder, C.A., Clark, J.S., Ver Hoef, J.M. & Wikle, C.K. (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553–570.
- Dail, D. & Madsen, L. (2011). Models for Estimating Abundance from Repeated Counts of an Open Metapopulation. *Biometrics*, **67**, 577–587.
- Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003). Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53–64.
- Dormann, C.F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129–138.
- Dormann, C.F. et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R. & Wintle, B.A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, **4**, 276–292.
- Guillera-Arroita, G., Lahoz-Monfort, J.J., MacKenzie, D.I., Wintle, B.A. & McCarthy, M.A. (2014). Ignoring imperfect detection in biological surveys is dangerous: a response to ‘fitting and interpreting occupancy models’. *PLOS ONE*, **9**, e99571.
- Hanks, E.M., Schliep, E.M., Hooten, M.B. & Hoeting, J.A. (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, **26**, 243–254.
- Hawkins, B.A. (2012). Eight (and a half) deadly sins of spatial analysis. *Journal of Biogeography*, **39**, 1–9.
- Hilbe, J.M. (2014). *Modeling Count Data*. Cambridge University Press, New York.
- Hodges, J.S. & Reich, B.J. (2010). Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love. *The American Statistician*, **64**, 325–334.
- Hosmer, D.W., Lemeshow, S. & Sturdivant, R.X. (2013). *Applied Logistic Regression*. Wiley.
- Johnson, D.H. (2008). In Defense of Indices: The Case of Bird Surveys. *Journal of Wildlife Management*, **72**, 857–868.
- Johnson, D.S., Conn, P.B., Hooten, M.B., Ray, J.C. & Pond, B.A. (2013). Spatial occupancy

- models for large data sets. *Ecology*, **94**, 801–808.
- Kammann, E.E. & Wand, M.P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **52**, 1–18.
- Kellner, K.F. & Swihart, R.K. (2014). Accounting for Imperfect Detection in Ecology: A Quantitative Review. *PLOS ONE*, **9**, e111436.
- Kéry, M. (2008). Estimating Abundance From Bird Counts: Binomial Mixture Models Uncover Complex Covariate Relationships. *The Auk*, **125**, 336–345.
- Kéry, M. (2011). Towards the modelling of true species distributions. *Journal of Biogeography*, **38**, 617–618.
- Kéry, M., Gardner, B. & Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, **37**, 1851–1862.
- Kéry, M. & Royle, J.A. (2016). *Applied hierarchical modeling in ecology (Volume 1)*. Academic Press, Waltham.
- Kéry, M. & Schmidt, B.R. (2008). Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, **9**, 207–216.
- Krebs, C.J. (2001). *Ecology*. Pearson.
- Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, **13**, 66–69.
- Lahoz-Monfort, J.J., Guillera-Arroita, G. & Wintle, B.A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504–515.
- Latimer, A.M., Banerjee, S., Sang, H., Mosher, E.S. & Silander, J.A. (2009). Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern United States. *Ecology Letters*, **12**, 144–154.
- Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659–1673.
- Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, **28**, 3049–3067.
- MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Franklin, A.B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, **84**, 2200–2207.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

- Martin, J., Edwards, H.H., Fonnesebeck, C.J., Koslovsky, S.M., Harmak, C.W. & Dane, T.M. (2015). Combining information for monitoring at large spatial scales: First statewide abundance estimate of the Florida manatee. *Biological Conservation*, **186**, 44–51.
- Péron, G., Ferrand, Y., Gossmann, F., Bastat, C., Guenezan, M. & Gimenez, O. (2011). Nonparametric spatial regression of survival probability: visualization of population sinks in Eurasian woodcock. *Ecology*, **92**, 1672–1679.
- Potts, J.M. & Elith, J. (2006). Comparing species abundance models. *Ecological Modelling*, **199**, 153–163.
- R Core Team. (2014). R: A Language and Environment for Statistical Computing.
- Royle, J.A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, **60**, 108–115.
- Royle, J.A., Kéry, M., Gautier, R. & Schmid, H. (2007). Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs*, **77**, 465–481.
- Ruppert, D., Wand, M.P. & Carroll, R.J. (2003). *Semiparametric Regression*. Cambridge University Press, New York.
- Schmid, H., Luder, R., Naef-Daenzer, B., Graf, R. & Zbinden, N. (1998). *Schweizer Brutvogelatlas. Verbreitung der Brutvögel in der Schweiz und im Fürstentum Liechtenstein 1993–1996*. Swiss Ornithological Institute, Sempach.
- Schmid, H., Zbinden, N. & Keller, V. (2002). *Überwachung der Bestandsentwicklung häufiger Brutvögel in der Schweiz*. Swiss Ornithological Institute, Sempach.
- Stephens, P.A., Pettoirelli, N., Barlow, J., Whittingham, M.J. & Cadotte, M.W. (2015). Management by proxy? The use of indices in applied ecology. *Journal of Applied Ecology*, **52**, 1–6.
- Stokes, E.J., Strindberg, S., Bakabana, P.C., Elkan, P.W., Iyenguet, F.C., Madzoké, B., Malanda, G.A.F., Mowawa, B.S., Moukoubou, C., Ouakabadio, F.K. & Rainey, H.J. (2010). Monitoring great ape and elephant abundance at large spatial scales: measuring effectiveness of a conservation landscape. *PLOS ONE*, **5**, e10294.
- Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C. & Guisan, A. (2014). Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution*, **5**, 947–955.
- Tyre, A.J., Tenhumberg, B., Field, S.A., Nieljalke, D., Parris, K.M. & Possingham, H.P. (2003). Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications*, **13**, 1790–1801.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009). *Mixed Effects*

Tables

Table 1: Overview of the characteristics of the four models used to analyse the simulated data. Model 1 is a standard Poisson GLM, model 2 is a standard binomial N-mixture model, model 3 is a binomial N-mixture model with a CAR random effect to account for residual spatial autocorrelation, and model 4 is a binomial N-mixture model with 2-dimensional splines to account for residual spatial autocorrelation.

	Model 1	Model 2	Model 3	Model 4
Imperfect detection	-	+	+	+
Spatial autocorrelation	-	-	+ (CAR)	+ (Splines)











