Spatial models to account for variation in observer effort in bird atlases

Andrew M. Wilson1, Daniel W. Brauning2, Caitlin Carey3, Robert S. Mulvihill4

1Environmental Studies Department, Gettysburg College, Gettysburg, PA, USA
2Wildlife Management Bureau, Pennsylvania Game Commission, Harrisburg, PA, USA
3Conservation Management Institute, Virginia Tech, Blacksburg, VA, USA
4National Aviary, Allegheny Commons West, Pittsburgh, PA, USA

Correspondence
Andrew M. Wilson, Environmental Studies Department, Gettysburg College, Gettysburg, PA, USA.
Email: awilson@gettysburg.edu

Abstract
To assess the importance of variation in observer effort between and within bird atlas projects and demonstrate the use of relatively simple conditional autoregressive (CAR) models for analyzing grid-based atlas data with varying effort. Pennsylvania and West Virginia, United States of America. We used varying proportions of randomly selected training data to assess whether variations in observer effort can be accounted for using CAR models and whether such models would still be useful for atlases with incomplete data. We then evaluated whether the application of these models influenced our assessment of distribution change between two atlas projects separated by twenty years (Pennsylvania), and tested our modeling methodology on a state bird atlas with incomplete coverage (West Virginia). Conditional Autoregressive models which included observer effort and landscape covariates were able to make robust predictions of species distributions in cases of sparse data coverage. Further, we found that CAR models without landscape covariates performed favorably. These models also account for variation in observer effort between atlas projects and can have a profound effect on the overall assessment of distribution change. Accounting for variation in observer effort in atlas projects is critically important. CAR models provide a useful modeling framework for accounting for variation in observer effort in bird atlas data because they are relatively simple to apply, and quick to run.

KEYWORDS
bird atlas, conditional autoregressive, observer effort, Pennsylvania, spatial model, West Virginia

INTRODUCTION

1.1 Use of atlases

Grid-based biological atlases, especially of birds, have become increasingly popular ways of documenting species’ status and distributions since the first large-scale efforts were initiated in the 1960s (Gibbons, Donald, Bauer, Fornasari, & Dawson, 2007). Expanding networks of amateur surveyors have enabled the completion of bird atlas projects covering geographic scales ranging from counties, through states, countries, and even continents (e.g., Hagemeijer & Blair, 1997). Increasingly, bird and other biological atlas data are being utilized to investigate large-scale environmental issues, notably pertaining to climate change (e.g., Gillings, Balmer, & Fuller, 2015; Huntley, Altwegg, Barnard, Collingham, & Hole, 2012; Matthews, Iverson, Prasad, & Peters, 2011; Thomas & Lennon, 1999; Zuckerberg, Woods, & Porter, 2009) and species conservation (e.g., Araújo, Thuiller, Williams, & Reginster, 2005; Boakes et al., 2010; van der Hoek et al., 2015). Atlas
data are especially useful for developing climate envelope models, which can predict future distributions under climate change scenarios, thereby informing large-scale and long-term conservation plans (e.g., Beale, Baker, Brewer, & Lennon, 2013; Coetzee, Robertson, Erasmus, Van Rensburg, & Thuiller, 2009; Virkkala & Lehikoinen, 2014).

1.2 | Variation in effort and false negatives

Because most bird atlas projects rely on citizen scientists to complete the majority of the field surveys, field methods are designed to promote mass participation with the aim of achieving comprehensive spatial coverage (Greenwood, 2007). Inevitably, this leads to trade-offs between data quality and coverage (Robertson, Cumming, & Erasmus, 2010; Szabo, Butchart, Possingham, & Garnett, 2012), such as the adoption of somewhat flexible field protocols which often do not impose standardization of survey effort. This lack of structure results in spatial variation in observer effort, with the highest effort often expended in areas with habitats or bird communities of most interest to amateur surveyors (e.g., Szabo, Davy, Hooper, & Astheimer, 2007; Tulloch, Mustín, Possingham, Szabo, & Wilson, 2013). Further, accessibility may constrain spatial coverage, such that areas at a greater distance from centers of human populations often receive the lowest effort, especially if there is a lack of accessible roads or paths (McCarthy, Fletcher, Rota, & Hutto, 2012; Syfert, Smith, & Coomes, 2013). Hence, spatial bias can result in significant taxonomic bias, that is, under-representation of certain species or species groups (Robertson et al., 2010).

When repeat atlases projects are used to assess shifts in species’ distributions, changes in survey effort can result in biased measures of changes in range margins (Kujala, Vepsäläinen, Zuckerberg, & Brommer, 2013). It is especially important, therefore, that estimates of changes in distribution between atlas periods include an assessment of changes in survey effort. Many published bird atlases have not accounted for variation in survey effort when producing estimated changes in range size, and indeed, not all bird atlases have adequately collected data to measure survey effort. The application of species distributions models (SDM) is one way of accounting for biases due to variable survey effort. Species distribution models are algorithms that “identify a mathematical or logical function linking species’ occurrences and a set of predictors” (Kamino et al., 2012). A multitude of different SDMs have been developed and applied to ecological data (e.g., Alizpurua, Paquet, Brotons, & Titeux, 2015; Comte & Grenouillet, 2013; Elith, Kearney, & Phillips, 2010; Rocchini et al., 2011), but accounting for false negatives is a persistent challenge (Chefaoui & Lobo, 2008; Kéry, 2011; Rocchini et al., 2011). A failure to detect a species in a given area could be because the habitat is not suitable, or it could merely be due to insufficient survey effort or inappropriate survey protocols. Developing survey protocols to maximize the likelihood of detection across multiple species is challenging, and the probability of detecting all species—given their presence—is almost always less than one (Kéry, 2011). Occupancy models (MacKenzie et al., 2006) are increasingly used to account for nonperfect detection in biological atlas data (e.g., Broms, Johnson, Altwegg, & Conquest, 2014). These models are based on capture–recapture theory and hence rely on repeated site visits to model both species occurrence and detectability (MacKenzie et al., 2006). Unfortunately, data captured by bird atlas projects are often highly unstructured and may include opportunistic data, and highly variable sequences of survey effort between blocks, and over multiple years (e.g., Wilson, Brauning, & Mulvihill, 2012). Further, data capture during first generation atlases was often not sophisticated enough to retain a permanent record of the visit history, or even visit years, to each block. Hence, application of occupancy models is not always feasible, and such models can be computationally challenging (Broms et al., 2014; Kéry, Gardner, & Monnerat, 2010), which is a major hurdle if practitioners with limited resources are to apply them to large numbers of species.

1.3 | Aims

In this study, we use data from the Atlas of Breeding Birds in Pennsylvania, henceforth the “PBBA I” (Brauning, 1992; fieldwork 1983–1989), and the Second Atlas of Breeding Birds in Pennsylvania, henceforth the “PBBA II” (Wilson et al., 2012; fieldwork 2004–2009), to explore the use of a SDM approach to account for variation in observer effort, both within and between repeated atlas projects. We apply the methods to data from the West Virginia Breeding Bird Atlas II project, henceforth WVBBA II (unpublished; fieldwork 2009–2014), to demonstrate that relatively simple conditional autoregressive (CAR) models offer a useful framework for modeling species distributions using atlas data, by incorporating information on broad-scale landscape features and observer effort. Throughout this study, we use the term “occupancy probability” as shorthand for “occupancy and detection probability.” Our models predict species occurrences based on hypothetical but realistic levels of survey effort per block and hence we do not explicitly model detection probabilities.

Specifically, we test the following hypotheses:

1. Spatial (CAR) models are effective at predicting species’ occupancy probabilities for bird atlases with incomplete data
2. Spatial (CAR) models are superior to nonspatial models at predicting species’ occupancy probabilities
3. Accounting for variation in observer effort improves model fit

2 | METHODS

2.1 | Data

The PBBA I established a survey grid based on U.S. Geological Survey (USGS) 7.5-min topographic quadrangle maps; dividing each USGS quadrangle into six atlas “blocks,” the bottom-right of which being designated as a “priority” block. This resulted in 4,937 atlas blocks, including 787 priority blocks, of 3.75’ longitude and 2.5’ latitude (approximately 5.2 × 4.62 km) across the state of Pennsylvania (Brauning, 1992). Priority blocks were targeted for more thorough coverage than...
other blocks. Fieldwork was completed by approximately 2,000 volunteers in the years 1983–1989. The PBBA II (2004–2009) was a repeat effort with very minor changes to field survey protocols. Although the median change in block effort between atlas projects was an increase of 5 hr (mean = 8.68, SD = 33.67), there was not a uniform increase in effort across all blocks. Not only were the changes in effort not evenly distributed across the state (Wilson et al., 2012), but effort was lower than PBBA I in 37% of blocks. It should be noted that details of survey effort in each block (times and duration of individual visits) in PBBA I are longer available, but the database does have a record of total effort hours.

Blocks for the WVBBA II (2009–2014), developed for the WVBBA I (1984–1989; Buckelew & Hall, 1994), were delineated in the same way as the PBBA blocks. This resulted in 2,653 atlas blocks across the state of West Virginia, 469 of which were designated as priority blocks. Due to a lower population density, and consequently smaller volunteer pool, block coverage—in 2009–2014—for the WVBBA II was incomplete in comparison with the PBBA II. No bird records were submitted for 580 atlas blocks (21.9%), and observer effort exceeded 1 hr in less than half of blocks (48.2%). However, coverage of the 469 priority blocks in WVBBA II was comprehensive, with a mean of 19.6 hr (SE = 0.65) of survey effort expended per block. Like the Pennsylvania atlas, spatial distributions of observer effort in the WVBBA II were not uniform, consisting of fairly comprehensive coverage in the Allegheny Mountains and lower elevation “panhandle” (eastern half of the state), but very patchy coverage across much of the Allegheny Plateau (western half of the state; Figure 2).

2.2 | Study species

We tested our hypotheses on Pennsylvania atlas data for six species: Ruffed Grouse (*Bonasa umbellus*), American Kestrel (*Falco sparverius*), Carolina Wren (*Thryothorus ludovicianus*) (Figure 1), Ovenbird (*Seiurus*...
auropalli
cerulean Warbler (Dendroica cerulea), and Henlow’s Sparrow (Ammol
terrible representative sample across abundance and habitat gradients. Among these, six species (Table 1) are localized species (Henlow’s Sparrow, found in 4.6% of blocks), species with statewide distributions (Ovenbird, American Kestrel) and range restricted species (Carolina Wren, absent from most of the northern one-third of the state); forest obligates (Ruffed Grouse, Ovenbird, Cerulean Warbler), a grassland obligate (Henlow’s Sparrow), a generalist (Carolina Wren); increasing species (Ovenbird, Carolina Wren) and declining species (Ruffed Grouse, American Kestrel, Cerulean Warbler, Henlow’s Sparrow).

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Mean value (SE)</th>
<th>Global Moran’s I</th>
<th>z-score</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>62.47 (0.37)</td>
<td>0.77</td>
<td>53.47</td>
<td>a</td>
</tr>
<tr>
<td>Deciduous</td>
<td>50.87 (0.39)</td>
<td>0.846</td>
<td>58.82</td>
<td>a</td>
</tr>
<tr>
<td>Mixed</td>
<td>11.08 (0.21)</td>
<td>0.82</td>
<td>57.09</td>
<td>a</td>
</tr>
<tr>
<td>Conifer</td>
<td>0.51 (0.027)</td>
<td>0.66</td>
<td>46.36</td>
<td>a</td>
</tr>
<tr>
<td>Core</td>
<td>34.97 (0.35)</td>
<td>0.752</td>
<td>52.4</td>
<td>a</td>
</tr>
<tr>
<td>Edge</td>
<td>27.51 (0.14)</td>
<td>0.648</td>
<td>45.06</td>
<td>a</td>
</tr>
<tr>
<td>Young forest</td>
<td>0.18 (0.006)</td>
<td>0.548</td>
<td>38.71</td>
<td>a</td>
</tr>
<tr>
<td>% Water</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1.77 (0.087)</td>
<td>0.034</td>
<td>21.4</td>
<td>a</td>
</tr>
<tr>
<td>% Wetland</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1.29 (0.041)</td>
<td>0.893</td>
<td>62.13</td>
<td>a</td>
</tr>
<tr>
<td>Emergent</td>
<td>1.07 (0.037)</td>
<td>0.892</td>
<td>62.07</td>
<td>a</td>
</tr>
<tr>
<td>Woody</td>
<td>0.22 (0.009)</td>
<td>0.893</td>
<td>62.1</td>
<td>a</td>
</tr>
<tr>
<td>% Farmland</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>23.52 (0.29)</td>
<td>0.716</td>
<td>49.77</td>
<td>a</td>
</tr>
<tr>
<td>Grassland</td>
<td>13.66 (0.16)</td>
<td>0.664</td>
<td>46.16</td>
<td>a</td>
</tr>
<tr>
<td>Row crop</td>
<td>9.72 (0.17)</td>
<td>0.736</td>
<td>51.21</td>
<td>a</td>
</tr>
<tr>
<td>% Developed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>10.9 (0.21)</td>
<td>0.757</td>
<td>52.73</td>
<td>a</td>
</tr>
<tr>
<td>Open (e.g., lawn)</td>
<td>6.03 (0.069)</td>
<td>0.689</td>
<td>47.97</td>
<td>a</td>
</tr>
<tr>
<td>Low density</td>
<td>3.04 (0.084)</td>
<td>0.724</td>
<td>50.39</td>
<td>a</td>
</tr>
<tr>
<td>Medium density</td>
<td>1.24 (0.051)</td>
<td>0.665</td>
<td>46.44</td>
<td>a</td>
</tr>
<tr>
<td>High density</td>
<td>0.6 (0.044)</td>
<td>0.673</td>
<td>47.91</td>
<td>a</td>
</tr>
<tr>
<td>Medium + high</td>
<td>1.84 (0.088)</td>
<td>0.695</td>
<td>48.63</td>
<td>a</td>
</tr>
<tr>
<td>% Reclaimed strip mines</td>
<td>0.37 (0.021)</td>
<td>0.514</td>
<td>38.74</td>
<td>ab</td>
</tr>
<tr>
<td>Stream density (m/km²)</td>
<td>1.072 (4.97)</td>
<td>0.490</td>
<td>34.11</td>
<td>c</td>
</tr>
<tr>
<td>River density (m/km²)</td>
<td>88.1 (5.92)</td>
<td>0.131</td>
<td>11.34</td>
<td>c</td>
</tr>
<tr>
<td>Forested stream density (m/km²)</td>
<td>728.8 (4.76)</td>
<td>0.642</td>
<td>44.66</td>
<td>ac</td>
</tr>
<tr>
<td>Mean elevation (m)</td>
<td>377.5 (2.19)</td>
<td>0.923</td>
<td>64.19</td>
<td>d</td>
</tr>
<tr>
<td>Elevation range (m)</td>
<td>165.5 (1.37)</td>
<td>0.637</td>
<td>44.34</td>
<td>d</td>
</tr>
</tbody>
</table>

aNational Land Cover Data 2006 (Fry et al., 2011).  
bAbandoned Mine Land Inventory (PA DEP 2009).  
cNetworked Streams of Pennsylvania (ERRI 1998).  
dNational Elevation Dataset, USGS.
We used the models developed for the PBBA II to produce predicted probabilities of block occupancy for 136 species (i.e., those that were found in at least 20 blocks) in WVBBA II.

2.3 | Data analysis

The predicted probability of occurrence by block was modeled using Hierarchical Logistic Regression in WinBUGS for each species (Lunn, Thomas, Best, & Spiegelhalter, 2000). The model took the form:

\[
\text{logit}(p_i) = \alpha + \left( \sum_{k=1}^{S} b_k x_{ik} \right) + \gamma_i + \delta_i + \varepsilon
\]

where \( p_i \) is the predicted probability of occurrence in block \( i \), \( \alpha \) is the intercept, \( b_k \) is the parameter estimate for \( s \) landscape covariates \( x \), \( \gamma_i \) is a correction factor to account for observer effort (i.e., deviation from a specified "standard"), \( \delta_i \) is the spatial effect, and \( \varepsilon \) is random error.

Landscape covariates were selected by modeling recorded presence/absence using a stepwise (by AIC) logistic model in program R (package MASS; Venables & Ripley, 2002). Between 8 and 12 candidate models were tested for each species. Candidate models were selected based on expert opinion and exploratory analysis of habitat associations (Wilson et al., 2012). There were 26 candidate landscape covariates (Table 2). Land cover data were from Landsat ETM+ derived data (c. 2005; Fry et al., 2011). The extent of reclaimed surface mine grassland was estimated by intersecting areas identified in the Abandoned Mine Inventory data for 2009 (PA DEP, 2009) with grassland and herbaceous land cover types.

Effort effects were modeled after Link and Sauer (2007), using the formula:

\[
f(\gamma) = \exp[D(\gamma/B) - 1]/C
\]

where \( B \) is the standardized number of hours (e.g., mean or median), and \( C \) and \( D \) are estimated parameters that determine the shape of the relationship between hours and probability of detection. This formulation enables estimation of effort effects that range from linear (\( C = 0, D = 1 \)), to diminishing returns (\( C = 0, 0 < D < 1 \)), and diminishing returns with an asymptote (\( C < 0, D > 0 \)). This function provides a multiplier; such that if \( B = 30 \), the multiplier would be \( >1 \) for blocks that receive \(<30\) hr of survey effort—increasing the predicted probability of occurrence in blocks with low effort. In the above example, the multiplier would equal \( 1 \) for survey effort of \( 30 \) hr.

Spatial effects were included using a Gaussian CAR model, which accounts for spatial autocorrelation in lattice data, such as gridded atlas blocks. Spatially explicit models are “expected to yield better predictions” (Bahn et al., 2006) than nonspatial models. In ecological terms, such models incorporate important spatial information that may relate to unknown effects of bird population dynamics and underlying environmental variation. Spatial autocorrelation was assessed using global Moran’s I test statistic using ESRI’s Spatial Statistics Tool. A Moran’s I value close to zero indicates spatial randomness while a positive value (up to 1) indicates positive spatial autocorrelation.

Statistical significance of Moran’s I was tested using z-tests (Z score is based on the Randomization Null Hypothesis computation). The computation of spatial autocorrelation was based on Queen contiguity and Euclidean distance (Anselin, 2005).

Models were implemented using Markov chain Monte Carlo (MCMC) simulation in WinBUGS. Vague normal prior distributions (0, 0.0001) were used to begin the MCMC sampling. Models were fitted with 5,000 iterations following a minimum 1,000 sample burn-in. Model convergence was checked by examining trace plots for all parameters (Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2012). Most models showed convergence after a few hundred iterations, but models applied to the sparser WVBBA II species data required up to 6,000 iterations to reach convergence.

To cross-validate our models and test our hypothesis that spatial models are effective at providing predictions for bird atlases with incomplete data, we ran five models for each of the six test species (PBBA II data):

1. all—using data from 100% of atlas blocks
2. 25% random—models trained using 25% of blocks, and tested on the remaining 75%
3. 50% random—models trained using 50% of blocks, and tested on the remaining 50%
4. 75% random—models trained using 75% of blocks, and tested on the remaining 25%
5. Priority blocks—models trained using only priority blocks, which comprise 16.5% of all blocks, and tested on the remaining 83.5%

Model accuracy was evaluated for test data using the area under the receiver operating characteristic (ROC) curve, commonly denoted as area under the curve (AUC). Area under the curve values of 0.5 imply that model accuracy is no better than random, while AUCs of 0.8 or more are considered good, and values of 0.9 or more are considered excellent (Brotons, Herrando, Estrada, Pedrocchi, & Martin, 2008). The AUC for test data was calculated in package pROC of program R (Robin et al., 2011). However, because AUC has received some criticism (e.g., Lobo, Jiménez-Valverde, & Real, 2008), we included an additional measure of model accuracy: the Point Biserial Correlation coefficient (CORR), which is a special case of Pearson’s correlation coefficient that measures the relationship between a continuous and a binary variable, as recommended by Kraemer (2006).

We tested our hypothesis that spatial models are superior to nonspatial models at predicting the presence/absence using four different nested models for each species:

4a. full (landscape characteristics + spatial effects + effort)
4b. landscape characteristics + spatial effects
4c. spatial effects + effort
4d. landscape characteristics + effort

Models 4a through 4d used the 75% random block data for model training (see model 4). Models were compared using Deviance Information Criterion (DIC).
Our third hypothesis, that incorporating observer effort effects improves model fit, was implicitly tested by comparing models 4a with 4c, and 4b with 4d. We also investigated the relationship between effort and predicted occupancy by running models with varying "standard" amounts of effort (B), ranging from 2 hr up to 50 hr, per block; including models with standard effort of 14 (the median block effort in PBBA II) and 21 hr (mean block effort). Effort was modeled as a function of two estimated parameters (after Link & Sauer, 2007), which allowed for a variety of relationships between field survey hours expended and the probability of detecting a species in each block. Hence, the modeled relationship was not necessarily linear; reflecting diminishing returns with increasing effort.

Finally, we demonstrated the extent to which changes in effort substantially affect bird atlas results, by making predictions using a "standard" 40 hr of effort across all blocks in both the PBBA I and PBBA II. A priori, we expected that because overall effort was lower in PBBA I than PBBA II, the effects of modeling presence/absence under a scenario of high effort would increase the number of predicted block occupied in the PBBA I more so than for the PBBA II. Our expectation was therefore that decreases in the number of occupied blocks between the atlases would otherwise have been underestimated, while increases would be overestimated. The models estimating change in the number of occupied blocks included all the data (model 1), and spatial and effort parameters, but did not include landscape covariates (e.g., model 4c) henceforth denoted as model 1c. We chose these reduced models because land cover data were not available for the period of the PBBA I. For further justification of this approach, see the comparison of models 4a through 4d in results. We ran this model for species found in at least 40 blocks during both atlases: 151 species or 88% of the 172 species confirmed to have bred in both PBBA I and PBBA II.

To assess changes in block occupancy between atlas efforts, we used a relative change measures:

\[
\text{Recorded change in block occupancy} = \log \left( \frac{\text{recorded block occupancy PBBA II}}{\text{recorded block occupancy PBBA I}} \right)
\]

and:

\[
\text{Predicted change in block occupancy} = \log \left( \frac{\text{predicted block occupancy PBBA II}}{\text{predicted block occupancy PBBA I}} \right)
\]

3 | RESULTS

Data from the two Pennsylvania atlases show that there is a strong relationship between changes in observer effort (hours of effort) and changes in the number of species detected within each block (Figure 3). For blocks where there was a reduction in effort, there was typically a corollary reduction in the number of species observed; while, conversely, the number of species detected usually increased in blocks where effort increased between atlas periods. However, the relationship between changes in effort and changes in observed species richness was not linear, but shows saturation such that increases in effort of more than 40 hr do not continue to accrue these effort effects.

Underlying the relationship between changes in effort and number of species detected is the fact that the probability of detection for each species is a function of hours of effort expended in each block. Models that corrected PBBA II data for effort show that increased effort hours would significantly increase block detections for all species tested (Figure 4). None of the models reached an asymptote within 50 hr of survey effort, and there were notable differences among species. For Ovenbird, the predicted number of occupied blocks increased slowly with increased survey effort, suggesting that this species was likely to be detected—where present—even with a limited amount of observer effort. In contrast, the predicted number of occupied blocks for Ruffed Grouse markedly increased with increasing effort up to, and beyond, 50 hr. This suggests that Ruffed Grouse were likely substantially underreported in PBBA II, given that mean block effort was 21 hr. Models for the four other study species revealed relationships that fell somewhere between the extremes exhibited by Ruffed Grouse and Ovenbird.

Spatial autocorrelation in PBBA II block occupancy data was highly significant (Z-test, p-value < .001) for all six study species (Global Moran’s I, mean = 0.33; Table 1). Spatial autocorrelation of landscape covariates between atlas blocks was also highly significant (Z-test, p-value < .0001; Table 2), with Moran’s I averaging 0.68 across the 26 covariates tested. Across study species, CAR models were consistently better (higher AUC and CORR) than nonspatial models at predicting the probability of detection in validation (test) blocks (compare models 4a through 4c with 4d, Table 3). However, CAR models that included
effort effects but did not include landscape covariates (i.e., model 4c) were either as good as, or better, than the “Full” model (i.e., model 4a) across all six species.

CAR models with landscape and effort effects proved to be good at predicting block occupancy (i.e., high AUC and CORR) for all six test species. Point Biserial Correlation coefficients between observed data and predicted probabilities (training data) were highest for all six species when 75% of block data was used (Figure 5; Table 3). Models based on 75% of block data resulted in the highest AUCs: between 0.898 (American Kestrel) and 0.973 (Carolina Wren). The poorest performing model (American Kestrel, priority blocks) had an AUC of 0.814, still suggesting a “good” model. Maps of predicted block occupancy showed that all models, even those based on only 16.5% and 25% of block data, represented (qualitatively) very reasonable approximations of recorded distributions (see Figures 6 and S1.1 thru S1.5).

Without correcting for effort, the mean relative (recorded) change in block occupancy between the PBBA I and PBBA II among 151 bird species was +6.2% (SE = 1.55), whereas when the overall increase in effort was accounted for (using model 1c), the mean relative predicted change was +3.5% (SE = 1.57). Moreover, by correcting for effort, the predicted change in block occupancy was reduced for 127 of 151 species (either increases were predicted to be lower than recorded, or decreases were predicted to be greater). Without correcting for effort, 109 of 151 species showed an increase in block occupancy, whereas when correcting for effort, only 84 of 151 species showed an increase.

Maps of predicted probability of occupancy for the WVBBBA II demonstrated that the methods developed for Pennsylvania were especially effective in producing more complete estimates of species distributions for the relatively data sparse WVBBBA II (e.g., Figure 7). For example, although the Carolina Wren was documented in just 38.9%
of atlas blocks in the WVBBA II, it was detected in 91% of priority blocks, and when modeled, had a predicted block occupancy of 90.4 (95% credible interval 88.7–92.1%). Across all 136 WVBBA II study species, recorded occupancy rates in priority blocks were on average 4.6 times higher than in nonpriority blocks (Figure 8). Further, the relationship between priority and all block (both priority and nonpriority) detection rates was not linear, with under-detection especially pronounced in moderately widespread species (those found in approx. 25%–75% of priority blocks), as opposed to localized and ubiquitous species (Figure 8). Our models produced predicted block occupancy rates that were very close to those from priority blocks for all species, with the relationship between actual and predicted very close to the identity line (Figure 8).

A map of predicted change in block occupancy of the Carolina Wren in PA between the PABBA I and PABBA II (Figure 9) suggests that most of the isolated instances of apparent loss of block occupancy were likely the result of decreased observer effort in some blocks. The predicted probability of changes in block occupancy provides a clearer map of likely range expansion of this species than the recorded occupancy, even though close to 100% block coverage was achieved in both PABBAI and PABBA II.

4 | DISCUSSION

Our results suggest that CAR models incorporating coarse landscape and effort effects are successful at predicting species’ occupancy probabilities in bird atlas blocks with little to no observer effort. Further, as the landscape covariates added rather little (if any) predictive power for our test species, models incorporating only spatial and effort effects may provide adequate models that circumvent a considerable amount of GIS-based analysis required to extract landscape covariates, and the subsequent model selection required to identify the best predictors. However, because our model testing was limited to only six species in a relatively homogenous state (all of Pennsylvania is within the Temperate and Broadleaf Mixed Forest Biome; Olsen et al., 2001), we caution against assuming that our findings would apply to all species and regions. The reason that our CAR models had high predictive power, even when landscape covariates were not included, was likely due to the fact that landscape covariates were highly spatially autocorrelated between adjacent atlas blocks; hence, the spatial component of the model accounted for large-scale patterns in land cover.

The model testing based on various percentages of training data suggests that our CAR models would be applicable to bird atlas projects with incomplete coverage. Even for very sparse data (e.g., 25% training data for Henslow’s Sparrow, see Fig. S1.5), our predicted probability of occupancy map provided a good approximation of actual species’ distributions. The main failing of our models was an under-prediction of isolated block occurrences that were outside of the species’ core range within the state (e.g., Fig. S1.5). However, it is likely that for many species, isolated block occurrences away from the species’ core ranges represented small and temporarily erratic populations. Hence, if bird atlas data are to be utilized for conservation planning, correctly demarcating core species’ ranges is critical (Rondinini, Wilson, Boitani, Grantham, & Possingham, 2006).

By correcting for survey effort, the number of species assessed that expanded their range (block occupancy) rather than show a range contraction between PBBA I and PBBA II changed sufficiently to put an entirely different complexion on atlas findings. Recorded data suggested that species showing increased block occupancy outnumbered those showing decreased block occupancy by more than two to one (2.59:1), but after correcting for effort, the ratio was much closer to parity (1.25:1). The potential effects of not correcting for survey effort to evaluate range shifts have been documented by others (Kujala et al., 2013). Our analysis supports the need for SDMs that incorporate variation in observer effort (MacKenzie et al., 2006) to correctly measure range shifts.

While our methods show that spatial models can account for variation in observer effort, there are some limitations to our analysis. While the number of effort hours is correlated with the number of species detected in an atlas block, there are several other factors that could influence the probability that any given species is detected, including the efficiency and level of prior experience of observers, the number of individual visits within and between years, the diel distribution of survey effort, and the spatial distribution of effort within a given block. Observer effort may also be influenced by habitat diversity, with more effort required to survey blocks with diverse habitats.

Another potential limitation to our spatial models is the likely presence of anisotropy—that is, directional dependent spatial relationships. The Valley and Ridge Physiographic Province of south-central Pennsylvania and much of West Virginia has a pronounced southwest to northeast topography, a result of the weathering of belts of rocks from repeated by folding and faulting (Fenneman, 1938). This topography has a direct impact on land use, with farmland and human development...
dominating the valleys, and forests on the ridges, and hence valleys and ridges have markedly different habitats (Wilson et al., 2012).

Although we included some model selection in our analysis, the multitude of candidate models that can be developed from just a handful of environmental covariates can be daunting, especially when dealing with data from atlas projects that include tens to hundreds of species. Other studies have found that broad land use types, elevation, and (for large extents) latitude and longitude explain a large
proportion of the variance in species’ distributions (Storch, Konvicka, Benes, Martinkova, & Gaston, 2003). However, our finding that simple spatial models—even those without landscape level covariates—perform well when making predictions based solely on atlas data emphasizes the importance of incorporating spatial autocorrelation into the analysis of atlas data.

While there are many methods available to predict species distributions, including sophisticated methods to account for imperfect detection (e.g., Sadoti, Zuckerberg, Jarzyna, & Porter, 2013), bird atlas projects are often constrained by limited analytical capabilities (i.e., restricted funds to employ data analysts), and a tight deadline to complete analysis for (potentially) 100s of species.
In light of those constraints, the relatively simple models used in this study offer a practical alternative. Our models for WVBBA II data typically converged in less than 10 min on a standard desktop computer (Intel Core i7 processor with 3.6 GHz CPU and 16 GB RAM). The rapidity with which these models can be applied would allow for the testing of several competing models, for each species (hence 100s or 1000s of models total) within a relatively short time-frame.

There has been much discussion about the relative merits of accounting for imperfect detection using occupancy-detection models (Guillera-Arroita, 2017). Some studies have shown that occupancy-detection models perform better for species that are difficult to detect, but that gains are, at best, modest for more easily detected species (Comte & Grenouillet, 2013; Rota, Fletcher, Evans, & Hutto, 2010). Hence, for analysis of atlas data where the main aim is to extrapolate species distributions from incomplete surveys (e.g., WVBBA II), our approach may be sufficient for readily detected species. For less readily detected species, a more sophisticated approach may be necessary, but in those cases, sample sizes (number of block detections) may be prohibitively small, anyway.

### 4.1 Recommendations

Our results suggest that bird atlas data with incomplete block coverage, or uneven effort, can still provide valuable data on species’ distributions and distribution change. Relatively simple CAR models provide a usefully modeling framework with which to account for missing data and biases in survey effort. To apply SDM

---

**FIGURE 9**  Recorded (left) and predicted (right) distribution of the Carolina Wren in the PABBA I (top) and PABBA II (middle), and change between atlas periods (bottom). Results from model 1 (using all available data, see text)
approaches that account for spatial variation in survey effort, it is critical that effort is comprehensively and accurately quantified. Volunteer/surveyor effort hours is now documented by most bird atlas projects, but other ways of measuring effort, such as distance travelled through the sampling unit (Robertson et al., 2010), or species accumulation lists (Moreno & Halfter, 2000). With the increasing use of online data capture for atlas projects (Robertson et al., 2010), the requirement to include a measure of survey effort with each data submission is a simple addition to online data capture forms.

Our analysis of the PABBA II and WVBBA II data revealed that total effort hours may not be sufficient for producing effort-corrected SDMs for species that are active at specific times of the days, most notably nocturnal species. While the time of day of observations was required, along with overall effort hours in the online data submission portal for PABBA II, we suspect that nocturnal effort hours were under-reported (Wilson et al., 2012). We therefore suggest that the importance of parsing daytime and nocturnal hours is emphasized in the future atlas efforts, through communication with surveyors and through careful development of recording forms/online portals to document effort hours accordingly. This would also allow for the application of occupancy-detection models, which may be especially useful for scarce or difficult to detect species (Guillera-Arroita, 2017).

It is not possible to state a broadly applicable minimum requirement for survey effort and block coverage from our analysis. The minimum requirement would depend to some extent of habitat heterogeneity, species richness, and species’ densities. However, the CAR models that we have employed work best when unsurveyed blocks are adjacent to blocks with data—hence, large tracts of unsurveyed blocks should be avoided. We suggest that our methods be applied to different regions, and atlases with a variety of grid sizes and coverage, to assess their general applicability. Finally, we encourage data analysts to report the CPU time required to run SDMs as a matter of course, thereby enabling managers of bird atlases to adequately budget for data analysis following data collection.

ACKNOWLEDGMENTS

Both the Pennsylvania and West Virginia Atlas projects were completed by large numbers of volunteer fieldworkers and coordinators, who we thank for their tireless dedication. PABBA II was funded principally by the Pennsylvania Game Commission with support from grants administered by the U.S. Fish and Wildlife Service, Wildlife and Sport Fish Restoration Program. WVBBA II was sponsored by the West Virginia Division of Natural Resources, Wildlife Resources Section. We are most grateful to two anonymous referees who provided very useful and insightful comments, which greatly improved our manuscript.

CONFLICT OF INTEREST

None declared.

REFERENCES


Anselin, L. (2005). GeoDa. 0.95i ed. Release Notes. Urbana Champaign, IL: Spatial Analysis Laboratory (SAL), Department of Agricultural and Consumer Economics, University of Illinois, USA.


**SUPPORTING INFORMATION**

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Wilson AM, Brauning DW, Carey C, Mulvihill RS. Spatial models to account for variation in observer effort in bird atlases. *Ecol Evol*. 2017;00:1–13. https://doi.org/10.1002/ece3.3201