

Statistical Reports

Ecology, 103(2), 2022, e03563
© 2021 by the Ecological Society of America.

Using machine learning to model nontraditional spatial dependence in occupancy data

NARMADHA M. MOHANKUMAR ¹ AND TREVOR J. HEFLEY 

Department of Statistics, Kansas State University, Manhattan, Kansas, USA

Citation: Mohankumar, N. M., and T. J. Hefley. 2022. Using machine learning to model nontraditional spatial dependence in occupancy data. *Ecology* 103(2):e03563. 10.1002/ecy.3563

Abstract. Spatial models for occupancy data are used to estimate and map the true presence of a species, which may depend on biotic and abiotic factors as well as spatial autocorrelation. Traditionally researchers have accounted for spatial autocorrelation in occupancy data by using a correlated normally distributed site-level random effect, which might be incapable of modeling nontraditional spatial dependence such as discontinuities and abrupt transitions. Machine learning approaches have the potential to model nontraditional spatial dependence, but these approaches do not account for observer errors such as false absences. By combining the flexibility of Bayesian hierarchical modeling and machine learning approaches, we present a general framework to model occupancy data that accounts for both traditional and nontraditional spatial dependence as well as false absences. We demonstrate our framework using six synthetic occupancy data sets and two real data sets. Our results demonstrate how to model both traditional and nontraditional spatial dependence in occupancy data, which enables a broader class of spatial occupancy models that can be used to improve predictive accuracy and model adequacy.

Key words: hierarchical Bayesian model; machine learning; occupancy model; presence–absence data; site occupancy; spatial dependence; zero-inflated binomial model.

INTRODUCTION

Many ecological studies collect occupancy data to understand the dynamics of species occurrence over space and time (e.g., Hepler et al. 2018, Joseph 2020). Occupancy data are collected by making replicated visits to sites and recording the presence or absence of at least one individual. During a site visit, individuals may go undetected even when present, resulting in the detection of no individuals (i.e., a false absence). Failure to account for false absences can have a significant impact on parameter estimates and predictions (Hoeting et al. 2000, MacKenzie et al. 2002, Tyre et al. 2003).

To facilitate the analysis of occupancy data that contain false absences, Hoeting et al. (2000), MacKenzie et al. (2002), and Tyre et al. (2003) introduced a zero-inflated Bernoulli model that specifies a distribution of the observed data given the true presence at a site. Using familiar notation for Bayesian hierarchical models, the conditional distribution of the data is

$$y_{ij} | z_i, p_{ij} \sim \begin{cases} \text{Bernoulli}(p_{ij}) & , z_i = 1 \\ 0 & , z_i = 0 \end{cases}, \quad (1)$$

where $y_{ij} = 1$ denotes the presence and detection of one or more individuals at the i^{th} site ($i = 1, 2, \dots, n$) during the j^{th} sampling period ($j = 1, 2, \dots, J_i$) and $y_{ij} = 0$ denotes that no individuals were detected. Detection of at least one individual depends on the probability p_{ij} . The z_i is the true presence ($z_i = 1$) or absence ($z_i = 0$) at the i^{th} site, which is assumed to be constant during all J_i sampling periods and modeled as

$$z_i | \psi_i \sim \text{Bernoulli}(\psi_i). \quad (2)$$

In Eq. 2, the probability of true presence, ψ_i , is modeled using an intercept term and q site-level covariates with the equation

$$g(\psi_i) = \mathbf{x}_i' \boldsymbol{\beta}, \quad (3)$$

where $g(\cdot)$ is an appropriate link function (e.g., logit or probit), $\mathbf{x}_i \equiv (1, x_1, x_2, \dots, x_q)'$, and $\boldsymbol{\beta} \equiv (\beta_0, \beta_1, \beta_2, \dots, \beta_q)'$. Within the vector $\boldsymbol{\beta}$, β_0 is the intercept parameter and $\beta_1, \beta_2, \dots, \beta_q$ are regression coefficients.

Manuscript received 2 May 2021; accepted 7 July 2021; final version received 7 October 2021. Corresponding Editor: Julia A. Jones.

¹ E-mail: meenu@ksu.edu

Since the introduction of the occupancy model in Eqs. 1–3, many extensions were developed to address model inadequacies. For example, to account for spatial dependence Johnson et al. (2013) added a correlated normally distributed site-level effect, η_i (i.e., $(\eta_1, \eta_2, \dots, \eta_n)' \sim N(0, \Sigma)$; see ch. 26 in Hooten and Hefley 2019) to Eq. 3 that resulted in

$$g(\psi_i) = \mathbf{x}_i' \boldsymbol{\beta} + \eta_i. \quad (4)$$

The approach by Johnson et al. (2013) has been effective in accounting for occupancy model inadequacies caused by traditional spatial dependence (e.g., Wright et al. 2019), which is assumed to have been generated from a correlated normally distributed random effect that imparts varying levels of smoothness on the spatial process. Discontinuities, abrupt transitions, and other “non-normal” spatial processes are common in ecological data, and the traditional spatial random effect may fail to capture such dynamics (e.g., Hefley et al. 2017). Unfortunately, ecologists lack alternative occupancy model specifications that would allow them to check for and, if needed, model nontraditional spatial dependence.

We demonstrate a framework for occupancy data to model both traditional and nontraditional spatial dependence. Our framework takes a machine learning approach to model the site-level effect in Eq. 4 and can model both traditional and nontraditional spatial dependence. We illustrate this framework using six synthetic data sets containing traditional and nontraditional spatial dependence and then apply our approach to understand the spatial dynamics of Thomson's gazelle (*Eudorcas thomsonii*) in Tanzania and sugar gliders (*Petaurus breviceps*) in Tasmania.

MATERIALS AND METHODS

Occupancy data requirements

Our proposed modeling framework builds upon the occupancy model of MacKenzie et al. (2002) and Tyre et al. (2003) and therefore is intended for use with occupancy data that are collected with repeated site visits, during which the true presence or absence of individuals at a site does not change. In addition, we require that false-negative detections are the only observational error. However, our framework is adaptable to accommodate other types of occupancy data (see “Model extensions” in Appendix S1 for additional detail). For example, our framework can be adapted to account for false presence, which occurs when individuals are not present at a site but are recorded as occurring at a site.

Spatial occupancy model framework

Our proposed framework involves lifting the normal distributional assumption in the spatial component that

accounts for the spatial dependence. To accomplish this, we replace the site-level effect in Eq. 4 with

$$g(\psi_i) = \mathbf{x}_i' \boldsymbol{\beta} + f(\mathbf{s}_i). \quad (5)$$

Conceptually, this is an important change; the $f(\mathbf{s}_i)$ is an unknown spatially varying process that is a function, $f(\cdot)$, that depends on the coordinate vector, \mathbf{s}_i , of the i^{th} site. The function $f(\cdot)$ is always unknown and is approximated.

This change in perspective is common in the field of machine learning, in which the goal is to “learn” or approximate an underlying function using data (see ch. 5 in Hastie et al. 2009). This simple change in Eq. 5 expands the types of model specifications for the spatially varying process, $f(\mathbf{s}_i)$. For example, regression trees are used to learn about underlying functions that have discontinuities and abrupt transitions, and using regression trees to approximate $f(\mathbf{s}_i)$ could model nontraditional spatial dependence.

Many approaches from machine learning, such as support vector regression, neural networks, boosted regression trees, and Gaussian processes, could approximate $f(\cdot)$. These approaches have been widely used by ecologists to make predictions and inferences about species distributions from abundance and presence-absence data (e.g., De'ath and Fabricius 2000, Cutler et al. 2007, Elith et al. 2008, Golding and Purse 2016). However, machine learning approaches are not widely used to model occupancy data because of the issues associated with false absences. Furthermore, approximating the spatial dependence within the occupancy model using machine learning approaches requires custom programming and a level of technical knowledge that hinders widespread use. The existing approaches that blend machine learning approaches with occupancy models are approach specific (e.g., Hutchinson et al. 2011, Joseph, 2020), and therefore switching among the different types of approaches to approximate $f(\cdot)$ is a challenge. For example, switching from a neural network to a regression tree to approximate $f(\cdot)$ in Eq. 5 would require extensive retooling of computer code, therefore hindering model checking, comparisons, and selection.

Fortunately, Shaby and Fink (2012) developed a model-fitting algorithm based on Markov chain Monte Carlo (MCMC) that enables off-the-shelf software for machine learning approaches, such as those available in R (e.g., `rpart(...)`, `svm(...)`, `gam(...)`), to be embedded within hierarchical Bayesian models. Once the initial computer code is written for the occupancy model, switching among machine learning approaches to approximate $f(\cdot)$ requires modifying only a few lines of code. Details associated with model fitting are provided in Appendix S1 of the Supplementary Material.

Modeling spatial dependence

To identify the spatial dependence and evaluate model adequacy, we used a model selection and model checking approach. First, we used a wide variety of approaches to model spatial dependence and then used a measure of predictive accuracy to determine which approach most accurately models the spatial process. We supplemented this predictive approach with a measure of model adequacy (e.g., Wright et al. 2019).

Following Hooten and Hobbs (2015), we measured the predictive accuracy using $-2 \times \text{LPPD}$, where LPPD is the log posterior predictive density. The $-2 \times \text{LPPD}$ is similar to the information criterion used for model selection but uses out-of-sample data rather than in-sample data (Hooten and Hobbs 2015). As such, $-2 \times \text{LPPD}$ and the difference in $-2 \times \text{LPPD}$ among models can be interpreted similarly to the information criterion that attempts to approximate $-2 \times \text{LPPD}$ using in-sample data (e.g., Watanabe–Akaike information criteria). For example, if model A produced a $-2 \times \text{LPPD}$ score less than the $-2 \times \text{LPPD}$ score produced by model B, then model A has higher predictive accuracy. As a standard of comparison, we fitted an occupancy model that does not account for spatial dependence (i.e., Eq. 3; from this point forwards non-spatial occupancy model).

In addition, we used Moran's I correlogram to check model adequacy. Moran's I has been used to detect traditional spatial dependence in the residuals of fitted occupancy models (Wright et al. 2019). However, if traditional approaches fail to capture spatial dependence, then Moran's I may identify such inadequacies.

Synthetic data examples

For our synthetic data examples, we show the probability of occupancy in Fig. 1, which includes the three scenarios of nontraditional spatial dependence and the three scenarios of traditional spatial dependence listed below.

- 1) Spatial dependence that has discontinuities and abrupt transitions generated by a step-wise function (nontraditional; Fig. 1a).
- 2) Spatial dependence forming a circle with the probability of occupancy being low in the center and smoothly increases toward the edges (nontraditional; Fig. 1b).
- 3) Spatial dependence defined by a cosine function (nontraditional; Fig. 1c).
- 4) Normally distributed random effect with a correlation matrix specified by a conditional autoregressive process (traditional; Fig. 1d).
- 5) Normally distributed random effect with a correlation matrix specified by an exponential covariance function (traditional; Fig. 1e).

- 6) Normally distributed random effect with a correlation matrix specified by a squared exponential covariance function (traditional; Fig. 1f).

For each scenario, we generated synthetic data using Eqs. 1, 2, and 5 on a unit square study area (i.e., $S = [0, 1] \times [0, 1]$). We divided the study area, S , into 900 grid cells (sites). We set the true values for the parameters to $p_{ij} = 0.5$ and $\beta_0 = 0$. We excluded covariates and regression coefficients in our synthetic data so that the spatial process was unobstructed when ψ_i is mapped onto S , which aids when visual and numerical comparisons are made among the machine learning approaches. From the 900 grid cells, we considered a random sample of $n = 200$ sites as the study area with $J_i = 3$ visits for model fitting.

We applied our spatial occupancy modeling framework to the six synthetic data sets and compared the performance of four embedded machine learning approaches, which included regression trees, support vector regression, a low-rank Gaussian process, and a Gaussian Markov random field. The low-rank Gaussian process and Gaussian Markov random field are approaches that model traditional spatial dependence for data sets with a large number of sites and have been used in models for occupancy data (Johnson et al. 2013, Heaton et al. 2019). The regression tree and support vector regression are nontraditional approaches and may be capable of modeling nontraditional types of spatial dependence.

We assessed the performance of each approach to model spatial dependence using $-2 \times \text{LPPD}$ calculated at 200 sites with $J_i = 3$ visits that were not used for model fitting (from this point forwards out-of-sample sites) and using Moran's I correlogram. In addition, we visually compared the true probability of occupancy (ψ_i) to the posterior mean of the probability of occupancy ($E(\psi_i | \mathbf{y})$; Fig. 2). Details associated with the synthetic data are provided in Appendix S2 of the Supplementary Material.

Thomson's gazelle data

We illustrate our spatial occupancy modeling framework using a data set from Hepler et al. (2018), who reported the presence and absence of Thomson's gazelle at 195 sites within Serengeti National Park, Tanzania (Fig. 3a). The sites were sampled using a network of motion-sensitive and thermally activated cameras. Images were classified by participants on the citizen science website Snapshot Serengeti. A site visit consisted of an 8-day period during the year 2012 (e.g., 1–8 January 2012). Each site was visited between 1 and 46 times (the mean number of visits was 29). Following Hepler et al. (2018), $y_{ij} = 1$ (from Eq. 1) was recorded if an image of at least one Thomson's gazelle was captured at the i^{th} site within the j^{th} 8-d window. A value of $y_{ij} = 0$ was recorded if the site was sampled, but no individuals were observed. Of the 195 sites, 141 had at least

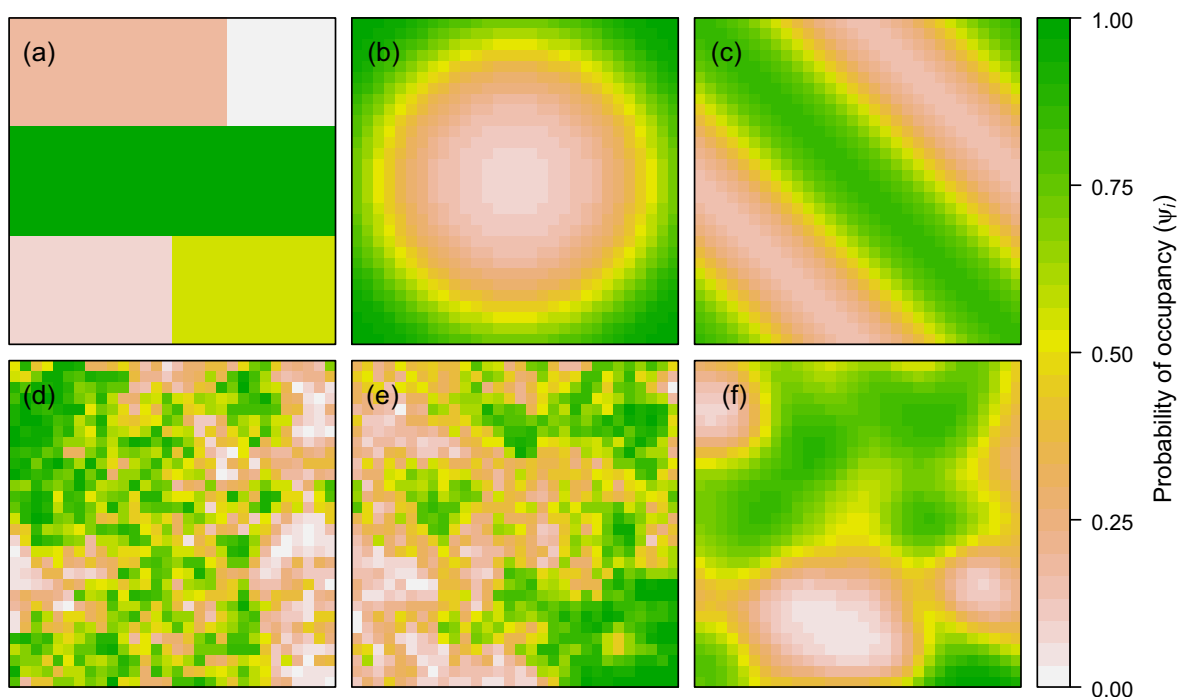


FIG. 1. Synthetic data examples showing the probability of occupancy (ψ_i in Eq. 5) at 900 potential sites (pixels) for six scenarios of traditional and nontraditional spatial dependence. The nontraditional scenarios include spatial dependence having discontinuities and abrupt transitions (panel a), forming a circle (panel b), and defined by a cosine function (panel c). The traditional scenarios include spatial dependence generated from a normally distributed random effect with a correlation matrix specified using a conditional autoregressive process (panel d), an exponential covariance function (panel e), and a squared exponential covariance function (panel f).

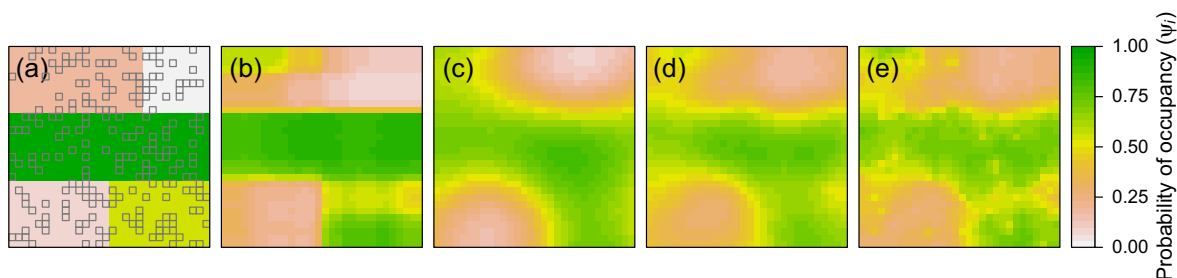


FIG. 2. The probability of occupancy from scenario 1 of the synthetic data example (panel a) and the posterior mean of the probability of occupancy ($E(\psi_i | \mathbf{y})$) obtained by fitting spatial occupancy models that included an embedded regression tree (panel b), a support vector regression (panel c), a low-rank Gaussian process (panel d), and a Gaussian Markov random field (panel e). The gray squares in panel a are the locations of the 200 sampled sites used for model fitting.

one detection. We use 100 randomly selected sites for model fitting and reserve the remaining 95 sites to calculate $-2 \times \text{LPPD}$.

Similar to our synthetic data example, we applied our spatial occupancy modeling framework by embedding four machine learning approaches, which include regression trees, support vector regression, a low-rank Gaussian process, and a Gaussian Markov random field. We excluded site-level covariates in our data example to illustrate our approach's ability to model multiple processes that generate spatial dependence (e.g., missing

site-level covariates and spatial autocorrelation) and to illustrate the ability of our method to serve as a "spatial interpolator" for occupancy data (i.e., similar to indicator or binomial kriging, but accounting for false absences). However, as with traditional occupancy models, we can easily include site-level covariates into our spatial occupancy models. Details associated with the data example are provided in Appendix S3 of the Supplementary Material. The data used in our data examples are available from the Dryad Digital Repository (Mohankumar and Hefley 2021).

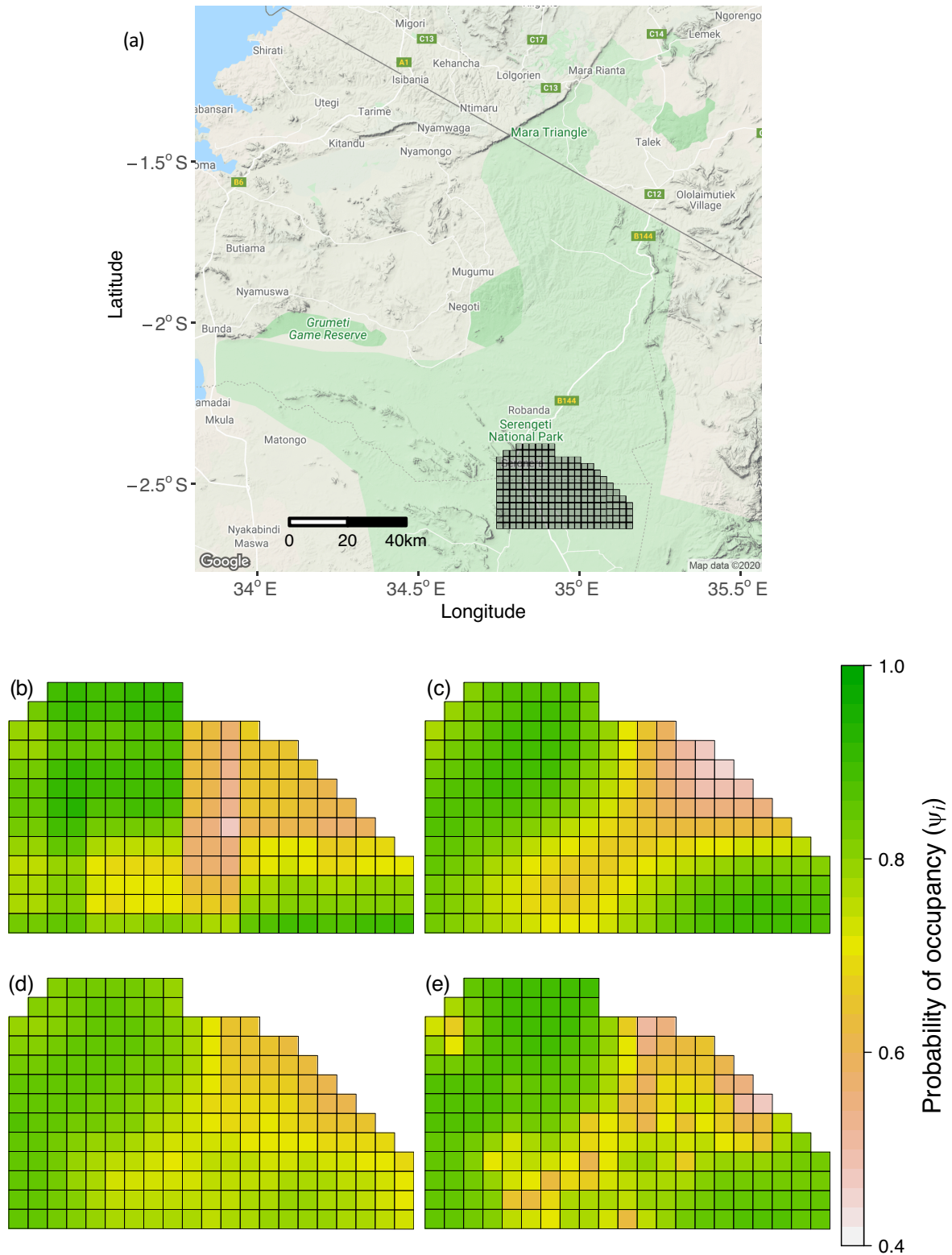


FIG. 3. Thomson's gazelle data from Hepler et al. (2018) collected at 195 sites within Serengeti National Park, Tanzania (panel a) and the posterior mean of the probability of occupancy ($E(\psi_i | y)$; panels b–e). Panels (b–e) show $E(\psi_i | y)$ obtained by fitting spatial occupancy models that included an embedded regression tree (panel b), a support vector regression (panel c), a low-rank Gaussian process (panel d), and a Gaussian Markov random field (panel e).

Sugar glider data

We illustrate our modeling framework using a second data set from Allen et al. (2018), who reported the presence and absence of sugar gliders. The data were collected during four or five site visits made to 100 sites in the Southern Forest region of Tasmania (Fig. 4a). Of the 100 sites, 79 had at least one sugar glider detected.

As this data set has a relatively small number of sites, we used 75 randomly selected sites for model fitting and reserve the remaining 25 sites to calculate $-2 \times \text{LPPD}$. We used the same modeling approaches for this example as we did in the Thomson's gazelle example. Details associated with the data example are provided in Appendix S4 of the Supplementary Material. The data used in our data examples are available from

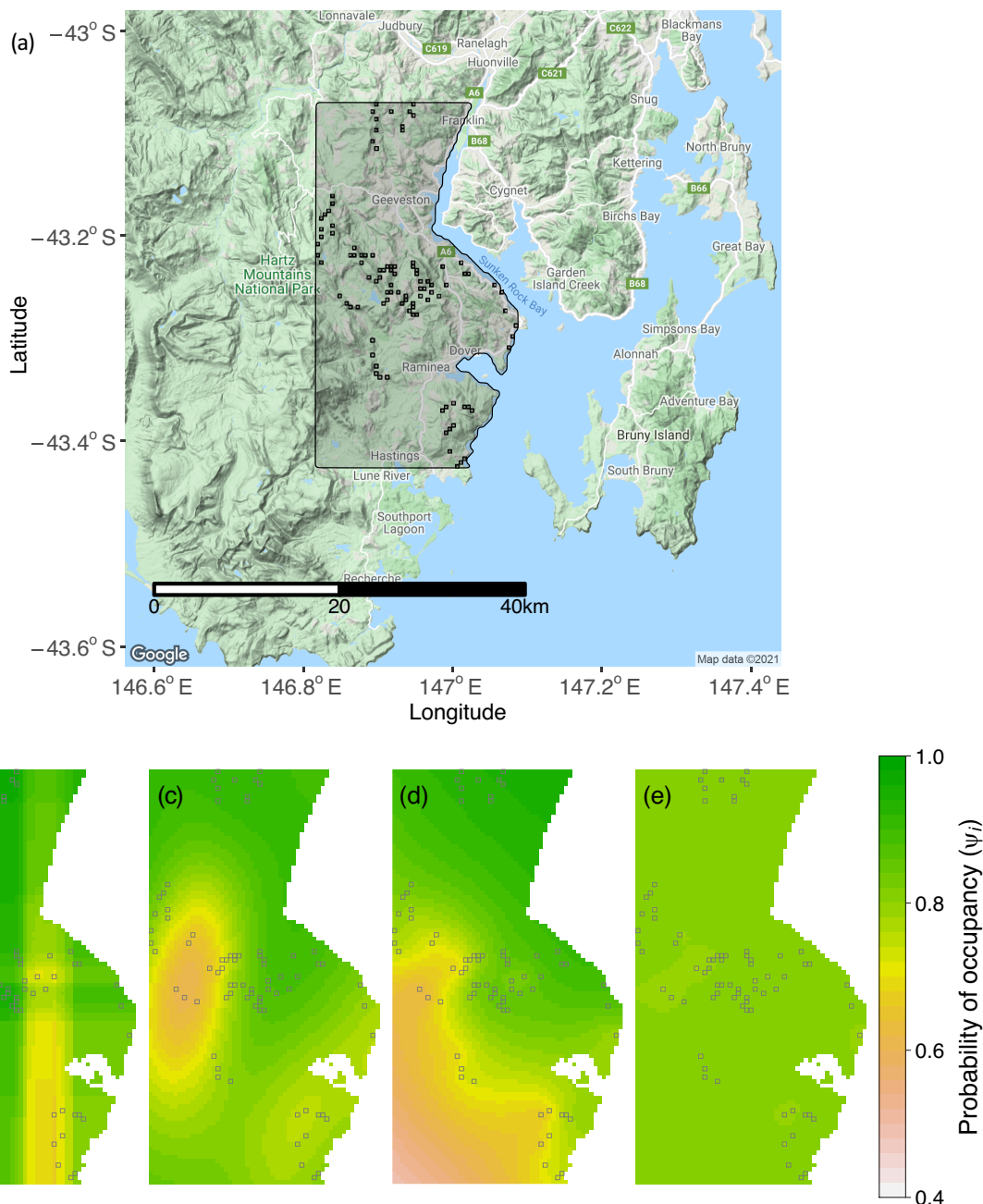


FIG. 4. Sugar glider data from Allen et al. (2018) collected at 100 sites in the Southern Forest region of Tasmania and the posterior mean of the probability of occupancy ($E(\psi_i | \mathbf{y})$; panels b–e). Panels (b–e) show $E(\psi_i | \mathbf{y})$ obtained by fitting spatial occupancy models that included an embedded regression tree (panel b), a support vector regression (panel c), a low-rank Gaussian process (panel d), and a Gaussian Markov random field (panel e).

the Dryad Digital Repository (Mohankumar and Hefley 2021).

RESULTS

Synthetic data examples

In scenario 1, the occupancy model with an embedded regression tree performed best because the other embedded machine learning approaches did not capture the abrupt transition created by the step-wise spatial process (Fig. 2). The $-2\times\text{LPPD}$ was 348.5, 377.2, 377.5, and 384.0 for the embedded regression tree, support vector regression, low-rank Gaussian process, and Gaussian Markov random field, respectively. For comparison, the $-2\times\text{LPPD}$ obtained from the non-spatial occupancy model was 433.1. Similarly, for scenario 1, the comparison of the Moran's I between the occupancy models suggested that spatial dependence must be accounted for using a regression tree; all other approaches resulted in lingering spatial dependence (see Appendix S5: Fig. S6).

Detailed results for scenarios 2–6 are presented in Appendix S5 of the Supplementary Material. For example, in scenario 2, the spatial dependence forms a circle with the probability of occupancy being low in the center and smoothly increases toward the edge of the circle (Fig. 1b). For scenario 2, we expected and found that the embedded support vector regression performed best (see Appendix S5). This was expected because this machine learning approach is best suited to learn about smoothly varying deterministic functions. In total, the results from the scenarios clearly demonstrated that, if the spatial process is a discontinuous step function, then the approaches used to model traditional spatial dependence are not adequate, and the approaches such as regression trees should be used. If the spatial dependence is traditional, the differences among the approaches are less distinct; nevertheless, in general, support vector regression performs better for smoothly varying processes (see Appendix S5).

Spatial occupancy dynamics of Thomson's gazelle

Across the four embedded machine learning approaches, the probability of occupancy at a site ranged from 0.45 to 0.95 (Fig. 3b–e). Generally, the probability of occupancy was high across the entire study area. However, there was a distinct band running from the southwest to the northeast of the study area where the probability of occupancy was much lower (Fig. 3b–e).

The measure of predictive accuracy, $-2\times\text{LPPD}$, was 669.4, 668.8, 671.0, and 668.7 for embedded regression trees, support vector regression, a low-rank Gaussian process, and a Gaussian Markov random field, respectively. For comparison, the $-2\times\text{LPPD}$ obtained from the non-spatial occupancy model was 676.7. Comparison of the Moran's I between the non-spatial and spatial occupancy

models suggested that accounting for spatial dependence improves model adequacy, although the utility of Moran's I is questionable because the differences among approaches are trivial, which may be due to the small number of sites (see Appendix S3: Fig. S1; Carrijo and da Silva, 2017). In total, the $-2\times\text{LPPD}$ and Moran's I suggested that spatial dependence should be accounted for in the model. However, Moran's I and $-2\times\text{LPPD}$ suggested that the differences among machine learning approaches are less distinct, therefore it is unclear if the spatial dependence is traditional or nontraditional.

Spatial occupancy dynamics of sugar gliders

For the sugar glider data example, the probability of occupancy at a site ranged from 0.48 to 0.97 (Fig. 4b–e) across the four embedded machine learning approaches. The probability of occupancy was generally high across the entire study area; however, there was an area in the eastern and southeastern portion of the study area where the probability of occupancy was relatively low (i.e., $\psi_i < 0.60$), and there were clear visual differences in the probability of occupancy among the four machine learning approaches (Fig. 4b–e). The measure of predictive accuracy, $-2\times\text{LPPD}$, was 78.2, 80.4, 79.6, and 78.9 for embedded regression trees, support vector regression, a low-rank Gaussian process, and a Gaussian Markov random field, respectively. For comparison, the $-2\times\text{LPPD}$ obtained from the non-spatial occupancy model was 80.3. Similar to Thomson's gazelle example, the comparison of the Moran's I between the occupancy models suggested that accounting for spatial dependence improves model adequacy (see Appendix S4: Fig. S1). In total, the $-2\times\text{LPPD}$ and Moran's I suggested that the spatial process (i.e., $f(\cdot)$ in Eq. 5) is best modeled using a regression tree. Using Moran's I and $-2\times\text{LPPD}$ as evidence, the results suggested that the spatial dependence is nontraditional.

DISCUSSION

The use of occupancy models has increased rapidly since the early 2000s. Occupancy data are inherently spatial but, unfortunately, only a limited number of approaches exists to model the spatial process (e.g., Hoeting et al. 2000, Johnson et al. 2013). This lack of spatial modeling options for occupancy data is in contrast with species distribution models (SDM) that predict the spatial distribution of a species using statistical and machine learning approaches applied to presence-only, count, and presence-absence data. There is a bewildering number of approaches within the SDM literature that are used to model the spatial process. Unfortunately, many of the SDM approaches do not account for contamination in the response variable (e.g., false absences). Understandably, ecologists may feel forced to choose between SDM approaches that do not account for contamination in the response variable (e.g.,

regression trees) and approaches that do, but with a lack of spatial modeling (e.g., occupancy models).

The crux for ecologists planning to use our framework is to determine which machine learning approaches are likely to capture the spatial process, which will require a level of familiarity with the properties of a wide range of machine learning approaches. We recommend James et al. (2013) for a gentle introduction and Hastie et al. (2009) and Murphy (2012) for more advanced and broad presentations. Within the ecological literature, there are also several excellent guides to machine learning approaches (e.g., De'ath and Fabricius 2000, Cutler et al. 2007, Elith et al. 2008).

Recently, the hierarchical modeling framework commonly used in ecology has been expanded to include some types of machine learning approaches such as neural networks (Wikle 2019, Joseph 2020). Our study builds upon this previous work and expands the types of spatial models ecologists can use for data that fit within the occupancy model framework. Although our work is focused on spatial dependence among the true presence at a site, the approach is easily generalizable. For example, Eq. 5 implies a linear effect of the site-level covariates (i.e., $\mathbf{x}_i'\beta$). Shaby and Fink (2012) showed how machine learning approaches can be used to capture nonlinear and unknown relationships between covariates and the probability of occupancy, therefore alleviating the linear assumption in Eq. 5. Furthermore, many studies that use occupancy models perform covariate selection using model selection techniques (e.g., Hooten and Hobbs 2015). While model selection techniques work for a small number of covariates, machine learning approaches may be superior when there is a large number of covariates. Another important generalization is that the machine learning approaches can be embedded to model the probability of detection as a function of predictor variables such as Julian date and observer effort (e.g., similar to the use of cubic splines used by Johnston et al. 2018). To facilitate these extensions, we explain in Appendix S1 how to generalize our framework for other popular ecological models, which is a direct application of the work by Shaby and Fink (2012).

ACKNOWLEDGMENTS

We thank Drs. Staci Hepler, Michael Anderson, and Robert Erhardt for assistance with the Thomson's gazelle data. We thank Dr. Drew Tyre, Dr. Daniel Fink, an anonymous reviewer, and Dr. Julia Jones for comments that improved our manuscript. Funding for this project was provided by the National Science Foundation via grant DEB 1754491.

LITERATURE CITED

- Allen, M., M. H. Webb, F. Alves, R. Heinsohn, and D. Stojanovic. 2018. Occupancy patterns of the introduced, predatory sugar glider in Tasmanian forests. *Austral Ecology* 43:470–475.
- Carrijo, T. B., and A. R. da Silva. 2017. Modified Moran's I for small samples. *Geographical Analysis* 49:451–467.
- Cutler, D. R., T. C. Jr Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. *Ecology* 88:2783–2792.
- De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178–3192.
- Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802–813.
- Golding, N., and B. V. Purse. 2016. Fast and flexible Bayesian species distribution modelling using Gaussian processes. *Methods in Ecology and Evolution* 7:598–608.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, Berlin, Germany.
- Heaton, M. J., et al. 2019. A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics* 24:398–425.
- Hefley, T. J., M. B. Hooten, E. M. Hanks, R. E. Russell, and D. P. Walsh. 2017. Dynamic spatio-temporal models for spatial data. *Spatial Statistics* 20:206–220.
- Hepler, S. A., R. Erhardt, and T. M. Anderson. 2018. Identifying drivers of spatial variation in occupancy with limited replication camera trap data. *Ecology* 99:2152–2158.
- Hoeting, J. A., M. Leecaster, and D. Bowden. 2000. An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics* 5:102–114.
- Hooten, M. B., and T. J. Hefley. 2019. *Bringing Bayesian models to life*. CRC Press, Boca Raton, Florida, USA.
- Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. *Ecological Monographs* 85:3–28.
- Hutchinson, R. A., L.-P. Liu, and T. G. Dietterich. 2011. Incorporating boosted regression trees into ecological latent variable models. In *AAAI* 11:1343–1348.
- James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An introduction to statistical learning*. Springer, Berlin, Germany.
- Johnson, D. S., P. B. Conn, M. B. Hooten, J. C. Ray, and B. A. Pond. 2013. Spatial occupancy models for large data sets. *Ecology* 94:801–808.
- Johnston, A., D. Fink, W. M. Hochachka, and S. Kelling. 2018. Estimates of observer expertise improve species distributions from citizen science data. *Methods in Ecology and Evolution* 9:88–97.
- Joseph, M. B. 2020. Neural hierarchical models of ecological populations. *Ecology Letters* 23:734–747.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- Mohankumar, N. M., and T. J. Hefley. 2021. Using machine learning to model nontraditional spatial dependence in occupancy data. *Dryad Digital Repository*. <https://doi.org/10.5061/dryad.4xgxd259g>.
- Murphy, K. P. 2012. *Machine learning: a probabilistic perspective*. MIT Press, Cambridge, Massachusetts, USA.
- Shaby, B. A., and D. Fink. 2012. Embedding black-box regression techniques into hierarchical Bayesian models. *Journal of Statistical Computation and Simulation* 82:1753–1766.
- Tyre, A. J., B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, and H. P. Possingham. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications* 13:1790–1801.

- Wikle, C. K. 2019. Comparison of deep neural networks and deep hierarchical models for spatio-temporal data. *Journal of Agricultural, Biological and Environmental Statistics* 24:175–203.
- Wright, W. J., K. M. Irvine, and M. D. Higgs. 2019. Identifying occupancy model inadequacies: can residuals separately assess detection and presence? *Ecology* 100:e02703.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1002/ecy.3563/supinfo>

OPEN RESEARCH

The Thomson's gazelle data and sugar glider data used in our data example are available in the Dryad Digital Repository (Mohankumar and Hefley 2021) <https://doi.org/10.5061/dryad.4xgxd259g>. A tutorial showing how to implement our statistical model is provided in Appendix S1. Annotated computer code that can be used to reproduce all results and figures associated with the simulation experiment and data examples are provided in Appendices S2, S3, and S4. In addition to the appendix files, the annotated R codes are provided within DataS1.zip.