Biometrics, 2025, 81(2), ujaf057
https://doi.org/10.1093/biomtc/ujaf057
Biometric Methodology Discussion Paper

Discussion on “Continuous-space occupancy models” by Wilson J. Wright and Mevin B. Hooten

Léa Pautrel[1,2,*], Marie-Pierre Etienne[3,4], Olivier Gimenez[1]

[1]CEFE, CNRS, EPHE, IRD, University of Montpellier, 34000 Montpellier, France, [2]OïkoLab, TerrOïko, 81540 Sorèze, France, [3]Ensai CREST—UMR 9194, CNRS, University of Rennes, 35000 Rennes, France, [4]IRMAR, Institut-Agro Rennes-Angers, University of Rennes, 35000 Rennes, France

[*]Corresponding author: Léa Pautrel, OïkoLab, TerrOïko, 2 place Dom Devic, 81540 Sorèze, France (lea.pautrel@terroiko.fr).

1 INTRODUCTION

Within the family of models designed to estimate species presence, known as species distribution models (SDMs), occupancy models account for imperfect detection. While there are multiple available options to perform SDMs in continuous space, this is not the case for occupancy models. Wright and Hooten (in press) propose a very interesting approach to fill this gap. They propose a novel approach using a clipped Gaussian process to infer species presence over continuous space, instead of the traditional site-based occupancy models. Their proposition has significant potential, and we examine it further from both practical and ecological perspectives.

2 MODEL FORMULATION

There are 2 major ecological benefits to the authors’ approach: the ability to estimate the actual proportion of area occupied (as opposed to the proportion of sites occupied) and the potential to address the issue of change-of-support, when the spatial scale is not aligned between different spatial data sources.

2.1 Continuous-space occupancy inference

The advantage of considering occupancy continuously in space, as the authors highlight, lies in the ability to estimate the proportion of area that is truly occupied, rather than relying on the proportion of occupied sites (discrete spatial units that depend on the site sampling process). The latter is inherently biased upward, because a site is considered occupied if even a small part of it is occupied. Moving beyond this discretization allows for a more objective measurement of the quantity of interest and could, for example, make it possible to compare studies conducted at different spatial resolutions. On a larger scale, this could help limit overestimation biases in occupied areas, regardless of the discretization scale used (Moat et al., 2018). This is particularly important because the area of occupancy is one of the criteria used by the International Union for the Conservation of Nature to establish the Red List of Ecosystems (Rodríguez et al., 2015), and this list is a cornerstone of many biodiversity conservation actions, a key foundation for prioritization and communication efforts.

However, discrete-space representations are not always irrelevant. In many cases, discretization is not arbitrary but biologically meaningful, particularly due to the structuring of habitats and populations. This is especially relevant for species that are highly specialized in certain habitats, where our ecological knowledge allows us to clearly identify and discretize these habitats. For example, colonial nesting birds are structured as metapopulations, making spatial discretization not only intuitive but also ecologically justified. Colony dynamics follow metapopulation processes, where patch occupancy fluctuates due to local extinctions and (re)colonization of suitable habitats (Hanski and Gilpin, 1991). Dynamic site occupancy models can effectively capture these fluctuations, and in this context, representing colonies as discrete spatial units is both appropriate and ecologically meaningful (Barbraud et al., 2003). Another example where discretizing space into sites is relevant is for amphibian species that are highly dependent on permanent ponds identified in the landscape (eg, Mazerolle et al. (2005)).

2.2 Change-of-support

In statistical ecology, addressing the change-of-support problem is a recurring challenge. This issue arises from scale mismatches between different data sources, which can be a temporal and/or a spatial mismatch. The proposed continuous-space occupancy model handles spatial scale mismatches: for example, species occurrences recorded as point locations must be aligned with the spatial scale of environmental covariates, often available at coarser resolutions. Additionally, spatial covariates from different sources may have inconsistent resolutions, which must also be accounted for. We believe that change-of-support is an important aspect to consider to ensure using reliable methods.

These challenges have become more prominent with the rise of integrated models, which combine multiple data streams and introduce spatial and temporal mismatches (Pacifici et al., 2019). In particular, integrated SDMs have received considerable attention in statistical ecology, making the change-of-support issue increasingly relevant in this field. Occupancy models, a specific type of SDM that accounts for imperfect detection (Kéry et al., 2010; Comte and Grenouillet, 2013), face the same challenges,

Received: February 10, 2025; Revised: February 27, 2025; Accepted: April 28, 2025

meaning solutions developed for SDMs may also apply to occupancy models.

Several types of spatial misalignment can occur:

1. Area-to-area: Species observations may be recorded at a coarser or finer resolution than the underlying ecological processes or environmental covariates, requiring upscaling or downscaling to a common resolution. For example, species distributions or abundances have been inferred from coarse-resolution observation data such as fisheries catch declarations or harvest records (Alglave, 2022; Fernández-López et al., 2025). Beyond species observation data, area-to-area mismatches also arise when integrating different environmental covariates, as they often vary in spatial resolution and require standardization.
2. Area-to-point: Environmental data often have coarser resolutions than species occurrences. A common case is land cover data from remote sensing, used to map habitats at broad scales (Niedballa et al., 2015). The resulting rasters misalign with point-based species records. This challenge is an ongoing concern in statistical ecology, as shown by Mourguiart et al. (2024), who compared methods to address spatial misalignment in SDMs.
3. Point-to-point: Even when both species and environmental data are point-based, their locations may not align. This issue concerns spatial precision rather than scale. While not a change-of-support problem, it is a spatial mismatch, more closely related to missing observations. It can be addressed through interpolation or imputation techniques (Finley et al., 2014).

There are a wide range of statistical methods to integrate data collected at different spatial scales (Gotway and Young, 2002; Madsen et al., 2008; Yarali and Rivaz, 2020). The occupancy model proposed by the authors offers a promising approach for handling area-to-area and area-to-point mismatches. Their method not only makes it possible to estimate occupancy continuously in space but also supports the integration of diverse spatial covariate data formats, as long as they are available at every point in space.

For instance, it could enable the use of vector geographic information system data, often available in public land-use databases, without requiring spatial aggregation or summarization metrics typically needed to upscale such data: If vector data of polygons that represent the forest cover is available, then this approach makes it possible to determine whether each given point falls within a forest, rather than relying on aggregated metrics such as the percentage of forest cover within a site. Conversely, this method supports the integration of coarse-resolution rasters available at large scales. However, users should remain cautious, as the resolution of the covariate raster is likely to influence the spatial precision of the resulting occupancy estimates, even if the integration is technically and mathematically feasible.

2.3 Further exploration and practical considerations

The authors propose a short simulation study that highlights the main advantage of their model: its ability to estimate the proportion of truly occupied area without the bias introduced by estimating the proportion of occupied sites. This is likely the main motivation behind its development. As potential users, we recognize the value of this approach and are eager to better understand the model's behaviour. We understand that addressing all potential questions about a new model in a single publication is not feasible and was not the authors' intent. However, we would like to see further exploration of the model, for instance, through additional simulations. Such work could serve as a valuable complementary publication and would certainly provide useful recommendations to ecologists.

Model assumptions. We would like to better understand the assumptions underlying this model and thus its potential applications. We assume that the model incorporates the classic assumptions of static single-species site-occupancy models, along with additional ones introduced by the equations used to transition to continuous space. For example, how is the closure assumption (animals do not move between sites) handled in this framework?

Sampling design. Complementary work could improve the sampling design recommendations for this model. What spatial covariate characteristics are needed to have an appropriate occupancy inference? What spatial structure diversity is recommended for a good estimation of the spatial covariance function? How reliable is this model with experimental designs with fewer observation points, which is often the case in practice?

Modelling choices. What is recommended in practice regarding the finite number of locations to have a reasonable approximation $(s_1, s_2, \ldots, s_D)$, with the approximation explained by the authors in equations (6) and (7)? Is it straightforward to define the buffer size for A_i? Should it depend on the species home range, or the sampling effort? How sensitive are the model outputs to these choices?

3 INFERENCE

Bayesian estimation of spatial models comes with a computational burden that generally increases with the number of observations. This challenge becomes even more pronounced when the spatial model includes a hidden layer that must be integrated out (Kang and Cressie, 2011; Banerjee and Fuentes, 2012). Although Integrated Nested Laplace Approximations (INLA) has become a widely popular approach for handling spatial random effects (Lindgren and Rue, 2015), this paper proposes an alternative method. The approach relies on approximating certain integrals using a discrete sum over surrogate points, selecting a prior to ensure partial conjugacy, and employing the Vecchia approximation, which draws inspiration from Markov processes. Specifically, it approximates the conditional distribution $z_d \mid z_1, \ldots, z_{d-1}$ by $z_d \mid z_{c(d)}$, where $c(d) \subset 1, \ldots, d-1$. This simplification of conditional dependencies results in a sparse precision matrix, allowing for more efficient posterior sampling. Practical implementation is also given significant attention by carefully organizing the D points and proposing an update strategy for spatial terms that marginalizes over the regression parameter $\boldsymbol{\beta}$.

It would be interesting to see a comparison of this Markov approximation with the INLA approach in term of computational efficiency, but also in terms of impact of the different approximations on the posterior distribution accuracy. As INLA approximation relies partly on a Gaussian Random Markov field defined on a mesh, it might not be so different than the approximation of the conditional distribution.

4 SUGGESTIONS FOR FUTURE DEVELOPMENTS

4.1 Clipped occupancy process

An assumption in the current formulation of the model is that the latent occupancy process is clipped to produce binary presence/absence estimates. We wonder whether removing the clipping step in the latent continuous occupancy process could be worth exploring as an adaptation of the authors' model. Retaining $\tilde{z}(s)$ to represent the ecological process, instead of clipping it to 0 or 1 as $z(s)$, could provide a more nuanced measure of habitat suitability, offering a continuous gradient of occupancy potential rather than a strict presence/absence classification. Additionally, this approach might simplify inference, as the integral $\int_{A_i} z(s)ds$ would be replaced by $\int_{A_i} \tilde{z}(s)ds$, which remains a Gaussian process, preserving its properties and potentially improving computational efficiency.

4.2 Continuous space-time occupancy modeling

Beyond improving the spatial representation of occupancy, a natural progression is to extend this continuous approach into the temporal dimension. The authors closed their paper by suggesting the possibility of modeling occupancy in continuous space-time. Like them, we find this a fascinating research avenue; however, we also recognize the substantial complexity of constructing such models and the potential challenges in producing robust, reliable outcomes.

This paper focuses on the continuous-space question in occupancy modeling, yet continuous-time raises different issues. Continuous-time models in occupancy often refer to time-to-first-detection models (Henry et al., 2020; Halstead et al., 2021; Haines et al., 2023; Priyadarshani et al., 2024), in which the species observation data are represented as the time until first detection, hence a continuous response variable, rather than as binary detection/non-detection data. Another option to expand the author's model would be to analyze time-to-each-detection data, an approach that gained interest in recent years with the advent of sensors such as camera traps (Kellner et al., 2022; Haines et al., 2023; Rushing, 2023). However, in these examples the continuous-time component applies only to the detection submodel of the hierarchical occupancy model and does not affect the occupancy state. We found that such approaches may not improve occupancy estimates over discrete-time models in many cases (Pautrel et al., 2024). The challenges involved in extending a continuous-time approach to the occupancy sub-model of the hierarchical framework share aspects with those we expect when trying to develop a truly continuous spatio-temporal occupancy model.

A primary concern arises from animal movement. In classical discrete-space occupancy models, an assumption is that animals do not move between discretized space units. This is typically achieved by choosing sites larger than the species' home range. Continuous-space models generally assume that the occupancy state is constant over time. When both space and time are modeled as continuous processes, these assumptions no longer hold because of animal movements in space and time.

Therefore, it seems necessary to account for movement in some way. Two non-exclusive options are possible. The first, more complex approach, would be to add a deterministic component to the model to explicitly describe movement. The second, a simpler extension of the presented model, would keep movement implicit, while incorporating spatio-temporal variations in presence. This aligns with dynamic occupancy models that incorporate colonization-extinction processes. Most dynamic occupancy model operate in discrete time, and we can envision how the authors' model could be adapted to a dynamic occupancy model in discrete time, for instance, by considering a colonization-extinction process. However, for inference to be possible, dynamic models typically rely on the assumption that sites are discrete in space (MacKenzie et al., 2003).

Given these challenges, ecological diffusion models offer an appealing alternative. These models, which rely on partial differential equations to capture diffusion processes, have been used to understand phenomena such as species dispersal, disease spread, and invasion dynamics (Cangelosi and Hooten, 2009; Louvrier et al., 2020; Zamberletti et al., 2022). For instance, Hefley et al. (2017) introduced a mechanistic spatiotemporal model within a hierarchical Bayesian framework as an alternative to regression-based approaches like SDMs or occupancy models with covariates whose impact on the response variable (eg, species presence) is estimated via regression. Notably, such approaches benefit from a continuous space and time framework regarding the implementation efficiency (Hefley et al., 2017), in contrast to classical regression models, where continuous approaches can impose a significant computational burden and substantially increase fitting time, as observed in the authors' study.

4.3 Model implementation and availability

New models inevitably come with trade-offs. For example, in this case, despite the authors' optimizations for MCMC inference, their proposed model takes longer to fit than more classical discrete-space occupancy models. However, with fitting times of approximately 3 h for their simulation study and 6 h for their avian data application, we feel that it should still be manageable for researchers. While these computational costs may be a limitation, we personally think they do not outweigh the importance of implementation. Making new models available in accessible software would allow practitioners to explore their potential benefits using their own data and research questions.

ACKNOWLEDGMENTS

M.-P.E. and O.G. are co-senior authors. We thank the co-editor for inviting us to contribute with this discussion paper on the authors' work, and we congratulate Wilson J. Wright and Mevin B. Hooten on their insightful paper.

FUNDING

This work was supported by the Association Nationale de la Recherche et de la Technologie (PhD Grant CIFRE N°2022/0428).

CONFLICT OF INTEREST

None declared.

REFERENCES

Alglave, B. (2022). *Inferring Fish Spatio-Temporal Distribution and Identifying Essential Habitats : Tackling the Challenge of Preferential Sampling and Change of Support to Integrate Heterogeneous Data Sources*. PhD thesis, Agrocampus Ouest.

Banerjee, S. and Fuentes, M. (2012). Bayesian modeling for large spatial datasets. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4, 59–66.

Barbraud, C., Nichols, J. D., Hines, J. E. and Hafner, H. (2003). Estimating rates of local extinction and colonization in colonial species and an extension to the metapopulation and community levels. *Oikos*, 101, 113–126.

Cangelosi, A. R. and Hooten, M. B. (2009). Models for bounded systems with continuous dynamics. *Biometrics*, 65, 850–856.

Comte, L. and Grenouillet, G. (2013). Species distribution modelling and imperfect detection: comparing occupancy versus consensus methods. *Diversity and Distributions*, 19, 996–1007.

Fernández-López, J., Acevedo, P., Illanas, S., Blanco-Aguiar, J. A., Vicente, J. and Gimenez, O. (2025). Game target-group: implementing inhomogeneous Poisson point process to estimate animal abundance from harvest data. *Methods in Ecology and Evolution*, 16, 170–182.

Finley, A. O., Banerjee, S. and Cook, B. D. (2014). Bayesian hierarchical models for spatially misaligned data in R. *Methods in Ecology and Evolution*, 5, 514–523.

Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97, 632–648.

Haines, L. M., Altwegg, R. and Borchers, D. L. (2023). Exact likelihoods for N-mixture models with time-to-detection data. *Australian and New Zealand Journal of Statistics*, 65, 327–343.

Halstead, B. J., Rose, J. P. and Kleeman, P. M. (2021). Time-to-detection occupancy methods: Performance and utility for improving efficiency of surveys. *Ecological Applications*, 31, e2267.

Hanski, I. and Gilpin, M. (1991). Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society*, 42, 3–16.

Hefley, T. J., Hooten, M. B., Russell, R. E., Walsh, D. P. and Powell, J. A. (2017). When mechanism matters: Bayesian forecasting using models of ecological diffusion. *Ecology Letters*, 20, 640–650.

Henry, D. A. W., Lee, A. T. K. and Altwegg, R. (2020). Can time-to-detection models with fewer survey replicates provide a robust alternative to traditional site-occupancy models? *Methods in Ecology and Evolution*, 11, 643–655.

Kang, E. L. and Cressie, N. (2011). Bayesian inference for the spatial random effects model. *Journal of the American Statistical Association*, 106, 972–983.

Kellner, K. F., Parsons, A. W., Kays, R., Millspaugh, J. J. and Rota, C. T. (2022). A two-species occupancy model with a continuous-time detection process reveals spatial and temporal interactions. *Journal of Agricultural, Biological and Environmental Statistics*, 27, 321–338.

Kéry, M., Gardner, B. and Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, 37, 1851–1862.

Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63, 1–25.

Louvrier, J., Papaïx, J., Duchamp, C. and Gimenez, O. (2020). A mechanistic-statistical species distribution model to explain and forecast wolf (*Canis Lupus*) colonization in South-Eastern France. *Spatial Statistics*, 36, 100428.

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G. and Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84, 2200–2207.

Madsen, L., Ruppert, D. and Altman, N. S. (2008). Regression with spatially misaligned data. *Environmetrics*, 19, 453–467.

Mazerolle, M. J., Desrochers, A. and Rochefort, L. (2005). Landscape characteristics influence pond occupancy by frogs after accounting for detectability. *Ecological Applications*, 15, 824–834.

Moat, J., Bachman, S. P., Field, R. and Boyd, D. S. (2018). Refining area of occupancy to address the modifiable areal unit problem in ecology and conservation. *Conservation Biology*, 32, 1278–1289.

Mourguiart, B., Chevalier, M., Marzloff, M., Caill-Milly, N., Mengersen, K. and Liquet, B. (2024). Dealing with area-to-point spatial misalignment in species distribution models. *Ecography*, 2024, e07104.

Niedballa, J., Sollmann, R., bin Mohamed, A., Bender, J. and Wilting, A. (2015). Defining habitat covariates in camera-trap based occupancy studies. *Scientific Reports*, 5, 17041.

Pacifici, K., Reich, B. J., Miller, D. A. W. and Pease, B. S. (2019). Resolving misaligned spatial data with integrated species distribution models. *Ecology*, 100, e02709.

Pautrel, L., Moulherat, S., Gimenez, O. and Etienne, M.-P. (2024). Analysing biodiversity observation data collected in continuous time: Should we use discrete- or continuous-time occupancy models? *Methods in Ecology and Evolution*, 15, 935–950.

Priyadarshani, D., Huynh, H.-D., Altwegg, R. and Hwang, W.-H. (2024). A unified framework for time-to-detection occupancy and abundance models. *Methods in Ecology and Evolution*, 15, 555–568.

Rodríguez, J. P., Keith, D. A., Rodríguez-Clark, K. M., Murray, N. J., Nicholson, E., Regan, T. J. et al. (2015). A practical guide to the application of the IUCN red list of ecosystems criteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140003.

Rushing, C. S. (2023). An ecologist's introduction to continuous-time multi-state models for capture-recapture data. *Journal of Animal Ecology*, 92, 936–944.

Wright, W. J. and Hooten, M. B. (2025). Continuous-space occupancy models. *Biometrics*, 81, 1–11.

Yarali, E. and Rivaz, F. (2020). Incorporating covariate information in the covariance structure of misaligned spatial data. *Environmetrics*, 31, e2623.

Zamberletti, P., Papaïx, J., Gabriel, E. and Opitz, T. (2022). Understanding complex spatial dynamics from mechanistic models through spatio-temporal point processes. *Ecography*, 2022, e05956.

Received: February 10, 2025; Revised: February 27, 2025; Accepted: April 28, 2025

Biometrics, 2025, 81(2), ujaf056
https://doi.org/10.1093/biomtc/ujaf056
Biometric Methodology Discussion Paper

Discussion on “Continuous-space occupancy models” by Wilson J. Wright and Mevin B. Hooten

Jeffrey W. Doser* and Krishna Pacifici

Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27606, United States

*Corresponding author: Jeffrey W. Doser, Department of Forestry and Environmental Resources, North Carolina State University, 2800 Faucette Drive, Raleigh, NC 27606, United States (jwdoser@ncsu.edu).

1 INTRODUCTION

Congratulations to Wilson J. Wright and Mevin B. Hooten for this insightful contribution and thank you to the *Biometrics* editors for the opportunity to discuss this paper. Spatial occupancy models are an increasingly common framework used to model species distributions while accounting for false negatives in data collection and residual spatial autocorrelation in the ecological process. Spatial autocorrelation is typically accommodated within an occupancy modeling framework through the use of discrete conditionally autoregressive terms (Johnson et al., 2013) or with continuous spatial processes (Doser et al., 2022) despite the observed data being collected within areal units. Wright and Hooten argue that such misalignment between the observed data and modeling of spatial structure in the ecological process can result in inferior inferences regarding the proportion of area occupied by a species of interest. The authors propose an elegant solution to this problem based on a clipped Gaussian process (De Oliveira, 2000) and change of support methods (Cressie, 1996) that they implement using an efficient Markov chain Monte Carlo (MCMC) algorithm.

In this discussion, we outline an alternative approach to address the change of support via a point process occupancy (PPO) model (Koshkina et al., 2017) that explicitly makes the connection between local density of individuals and detection probability (Royle and Nichols, 2003). This model presents a different viewpoint of what is meant by presence/absence (Gelfand, 2022). By comparing this approach to the Wright and Hooten model (hereafter WH model), we hope to more explicitly consider the interpretation of “occupancy” and how it can differ across modeling frameworks.

2 OCCUPANCY MODELING VIA POINT PROCESSES

Individual animals can be viewed as points distributed across space, which are naturally represented via point process models (Hefley and Hooten, 2016). Let $\mathcal{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n)$ denote the locations of n individuals within some study area A. The likelihood for a spatial point process can be written as

$$\mathcal{L}(\boldsymbol{\theta}; \mathcal{S}) = \exp\left\{ -\int_A \lambda(\boldsymbol{s}; \boldsymbol{\theta}) d\boldsymbol{s} \right\} \prod_{i=1}^{n} \lambda(\boldsymbol{s}_i; \boldsymbol{\theta}), \tag{1}$$

where $\lambda(\boldsymbol{s}; \boldsymbol{\theta})$ is an intensity function determining the distribution of individuals across space that depends on parameters $\boldsymbol{\theta}$. Two common choices for modeling $\lambda(\boldsymbol{s}; \boldsymbol{\theta})$ in ecology are the nonhomogeneous Poisson process (NHPP) and the log Gaussian Cox process (LGCP; Illian et al., 2008). For the simpler NHPP, the intensity function $\lambda(\boldsymbol{s}; \boldsymbol{\theta})$ is modeled according to

$$\log(\lambda(\boldsymbol{s}; \boldsymbol{\theta})) = \boldsymbol{x}^\top(\boldsymbol{s})\boldsymbol{\beta}, \tag{2}$$

where $\boldsymbol{\beta}$ represents the effects of a set of spatially referenced covariates $\boldsymbol{x}(\boldsymbol{s})$. The LGCP additionally incorporates a Gaussian process, $\mathrm{w}(\boldsymbol{s})$ into the log intensity function according to

$$\log(\lambda(\boldsymbol{s}; \boldsymbol{\theta})) = \boldsymbol{x}^\top(\boldsymbol{s})\boldsymbol{\beta} + \mathrm{w}(\boldsymbol{s}). \tag{3}$$

The most common form of data collection for occupancy models is where observers survey a set of areal units $j = 1, \ldots, J$, each with area $\mathcal{A}_j$, multiple times over $k = 1, \ldots, K_j$ repeat visits to the site. The integrated intensity function over area $\mathcal{A}_j$ is defined by

$$\bar{\lambda}_j = \int_{\mathcal{A}_j} \lambda(\boldsymbol{s}; \boldsymbol{\theta}) d\boldsymbol{s}. \tag{4}$$

Applying results from point process theory, the number of individuals N_j within area $\mathcal{A}_j$ is distributed as

$$N_j \sim \text{Poisson}(\bar{\lambda}_j). \tag{5}$$

Note that standard occupancy models (MacKenzie et al., 2002; hereafter STO models) require the “closure” assumption, which is equivalent to saying that the number of individuals within area $\mathcal{A}_j$ must remain greater than 0 or at 0 over all K_j visits. Here, we consider the more stringent assumption that N_j remains constant over each of the K_j visits in order to directly link the occupancy data collection process with the point process. This as-

Received: January 30, 2025; Revised: March 3, 2025; Accepted: April 23, 2025

sumption is equivalent to saying that individuals do not move to a different area over the time span of the repeat visits.

The occupancy of areal site j, z_j, is immediately defined from (5) such that $z_j = 1$ if and only if $N_j > 0$ and similarly $z_j = 0 \iff N_j = 0$. The probability of the species occupying site j, ψ_j, is defined as

$$\psi_j = P(z_j = 1) = P(N_j > 0) = 1 - P(N_j = 0) = 1 - e^{-\bar{\lambda}_j}. \tag{6}$$

By viewing occupancy as a process explicitly defined from a point process, it is then straightforward to link the detection probability of the species to the number of individuals present in the areal site (Royle and Nichols, 2003). Let $y_{j,k}$ denote the observed detection (1) or nondetection (0) of the species of interest at site j during visit k. The observation model for $y_{j,k}$ can be defined by

$$y_{j,k} \mid N_j \sim \text{Bernoulli}(p^*_{j,k}), \tag{7}$$

$$p^*_{j,k} = 1 - (1 - p_{j,k})^{N_j}, \tag{8}$$

$$\text{logit}(p_{j,k}) = \boldsymbol{v}^\top_{j,k}\boldsymbol{\alpha}, \tag{9}$$

where $p^*_{j,k}$ is the probability of detecting the species, $p_{j,k}$ is the probability of detecting an individual, and $\boldsymbol{\alpha}$ are effects of covariates $\boldsymbol{v}_{j,k}$. This PPO model effectively extends the PPO model of Koshkina et al. (2017) to explicitly account for the impacts of local abundance on detection probability (Royle and Nichols, 2003). The PPO model could be implemented in a Bayesian framework using Markov chain Monte Carlo and, similar to the WH model, leverage Nearest Neighbor Gaussian Processes (Datta et al., 2016) if $\lambda(\boldsymbol{s}; \boldsymbol{\theta})$ is modeled using an LGCP.

3 COMPARISON TO THE WH MODEL

The WH model distinctly differs from the PPO model. The PPO model considers occupancy as solely a discrete concept whose value implicitly depends upon the size of the area $\mathcal{A}_j$ over which occupancy is being defined. As the size of $\mathcal{A}_j$ increases, ψ_j increases toward one. This concept that occupancy probability is scale-dependent is commonplace in the ecological literature (Efford and Dawson, 2012). Unlike the STO model ([1] and [2] in Wright and Hooten), the PPO model outlined here allows for occupancy to be defined at different scales via the integrated intensity function and the deterministic relationship between occupancy and the underlying point process (Koshkina et al., 2017).

The WH model considers occupancy as a process in continuous space. Analogous to the discussion in Gelfand and Shirota (2019), the WH model defines occupancy as a Bernoulli trial at any given location $\boldsymbol{s}$ as opposed to the probability that the number of individuals within some area around location $\boldsymbol{s}$ is greater than 0. In this framework, "occupancy" of an areal unit $\mathcal{A}$ would correspond to a block average of all locations in $\mathcal{A}$, or equivalently, the proportion of the point locations $\boldsymbol{s} \in \mathcal{A}$ where occupancy is one. This quantity is what Wright and Hooten use to relate detection probability to the continuous occupancy surface (i.e., [5] in Wright and Hooten), cogently arguing that detection probability should increase as this proportion becomes closer to one. This is an important realization to consider when applying this model and interpreting the resulting occupancy surface, particularly given the arguably more common interpretation of occupancy as being defined only for discrete units (Lele et al., 2013). To conceptualize this, suppose the expected abundance of individuals increases within areal unit $\mathcal{A}_j$ but the increases only occur within a subset of the unit that is already occupied. In this case, occupancy probability as defined by the PPO model would increase since occupancy probability by definition increases with expected abundance. However, occupancy probability as defined by the WH model would remain the same since the proportion of area occupied does not change.

Despite the differences, the approaches are similar in that they both attempt to link detection–nondetection data collected at an areal unit to an ecological process occurring across continuous space. Furthermore, the WH model and PPO model both explicitly address heterogeneity in detection probability that is not accounted for in the STO model. In the PPO model, detection probability of the species within an areal unit increases as the abundance of the site increases (8). Similarly, in the WH model, detection probability of the species within an areal unit increases as the proportion of the site that is occupied increases (i.e., [5] in Wright and Hooten). A key limitation of the STO model is that it does not account for abundance-related heterogeneity in detection probability, which can in certain situations lead to bias (Dorazio, 2007). Importantly, both the WH model and PPO model require any covariates on occupancy be available at each spatial location $\boldsymbol{s}$ in the study region, which may pose a significant limitation for practitioners interested in implementing these frameworks when important habitat features for the species of interest are not available via remote sensing products.

4 THE CLOSURE ASSUMPTION

The STO model requires making the assumption that the true occupancy state of an areal site remains constant over the time span of the repeat surveys done at the site (ie, the "closure" assumption). Given that occupancy is defined across continuous space in the WH model, does the WH model require closure across the entire continuous domain? In other words, for all $\boldsymbol{s} \in A$, does the model require $z(\boldsymbol{s})$ to remain constant across the repeated visits? Or rather does the model require that only $\max_{\boldsymbol{s} \in \mathcal{A}} z(\boldsymbol{s})$ remain constant over the repeated visits? To separately estimate occupancy and detection, we would expect only the latter to be a necessary assumption. However, the reliance of detection probability on the block average occupancy ([5] in Wright and Hooten) across the areal unit indicates that if this block-level average were to change over the repeat visits, bias may be induced in detection probability and ultimately the occupancy surface. Similarly, in the PPO model outlined in Section 2, detection probability is directly related to abundance in the areal site, and thus any change in abundance (and not just a change from $N_j = 0$ to $N_j > 0$ or vice versa) would likely render bias in the estimated occupancy probabilities. Note the ecological implications of this "bias" may simply result in a shift in interpretation of the underlying estimates (Kendall and White, 2009). Nevertheless, further assessment of violations of the closure assumption on the WH model could be fruitful in helping identify its use and interpretation by practitioners.

5 CONCLUDING REMARKS

The different interpretations of occupancy between the WH model and the PPO model outlined here may lead to the question of which viewpoint of occupancy is "correct"? We do not believe this is a useful question and instead argue that both viewpoints can provide useful information on species distributions. The most suitable framework for a given application likely depends on the characteristics of the species of interest and study design. For example, the WH model provides an intuitive way to model plant cover (Wright, 2024), while the PPO model may be helpful in linking interpretations of animal occupancy to animal movement, which are often described using point processes (eg, Fieberg et al., 2021). Crucially, we believe it is more important for ecologists using different occupancy modeling frameworks to clearly define what is meant by "occupancy" in a given analysis, how the analysis framework influences this interpretation, and the impacts such a framework and its assumptions have on the underlying inferences that can be drawn.

In summary, the continuous spatial occupancy model presented by Wright and Hooten is an important step forward in the growing literature on spatially-explicit species distribution models. We again congratulate the authors for their insightful contribution and look forward to future advances in this area.

FUNDING

None declared.

CONFLICT OF INTEREST

None declared.

REFERENCES

Cressie, N. A. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, 3, 159–180.

Datta, A., Banerjee, S., Finley, A. O. and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111, 800–812.

De Oliveira, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Computational Statistics & Data Analysis*, 34, 299–314.

Dorazio, R. M. (2007). On the choice of statistical models for estimating occurrence and extinction from animal surveys. *Ecology*, 88, 2773–2782.

Doser, J. W., Finley, A. O., Kéry, M. and Zipkin, E. F. (2022). spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. *Methods in Ecology and Evolution*, 13, 1670–1678.

Efford, M. G. and Dawson, D. K. (2012). Occupancy in continuous habitat. *Ecosphere*, 3, 1–15.

Fieberg, J., Signer, J., Smith, B. and Avgar, T. (2021). A "How to" guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, 90, 1027–1043.

Gelfand, A. E. (2022). Spatial modeling for the distribution of species in plant communities. *Spatial Statistics*, 50, 100582.

Gelfand, A. E. and Shirota, S. (2019). Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. *Ecological Monographs*, 89, e01372.

Hefley, T. J. and Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports*, 1, 87–97.

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, New York: John Wiley & Sons.

Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C. and Pond, B. A. (2013). Spatial occupancy models for large data sets. *Ecology*, 94, 801–808.

Kendall, W. L. and White, G. C. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *Journal of Applied Ecology*, 46, 1182–1188.

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M. and Stone, L. (2017). Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution*, 8, 420–430.

Lele, S. R., Merrill, E. H., Keim, J. and Boyce, M. S. (2013). Selection, use, choice and occupancy: clarifying concepts in resource selection studies. *Journal of Animal Ecology*, 82, 1183–1191.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A. and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.

Royle, J. A. and Nichols, J. D. (2003). Estimating abundance from repeated presence–absence data or point counts. *Ecology*, 84, 777–790.

Wright, W. J. (2024). *Advances in Bayesian Spatial Statistics for Ecology and Environmental Science*, *PhD thesis*. Fort Collins, Colorado: Colorado State University.

Received: January 30, 2025; Revised: March 3, 2025; Accepted: April 23, 2025

Biometrics, 2025, 81(2), ujaf058
https://doi.org/10.1093/biomtc/ujaf058
Biometric Methodology Discussion Paper

Rejoinder to the discussion on "Continuous-space occupancy models"

Wilson J. Wright [1,*] and Mevin B. Hooten [2]

[1]Department of Statistics, University of Missouri, Columbia, MO 65211, United States, [2]Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX 78705, United States

[*]Corresponding author: Wilson J. Wright, Department of Statistics, University of Missouri, Columbia, MO 65211, United States (wilson.wright@missouri.edu).

ABSTRACT

The discussions of our paper consider some assumptions of continuous-space occupancy models, alternative approaches, and directions for future research. In this short rejoinder, we expand on some of these ideas and provide additional comments.

KEYWORDS: animal movement; change of spatial support; closure assumption; spatial statistics; species distributions.

1 INTRODUCTION

Occupancy modeling has become an invaluable approach for mapping species distributions, while accounting for imperfect detection. In Wright and Hooten (2025), we considered species occupancy to be an ecological process with continuous spatial support. However, the data for occupancy studies are typically collected over areal survey units (called "sites"). Our approach accounts for this disconnect by directly modeling the corresponding change of support (eg, Cressie, 1996; Gelfand et al., 2001; Gotway and Young, 2002) between the species occupancy process and observed data. We showed how ignoring the change of support can bias inferences for the proportion of continuous area occupied by the species.

We thank the discussants for reading our article and providing insightful comments, questions, and suggestions. In this rejoinder, we elaborate on points brought up in the discussion papers and consider additional directions for future research. We begin by summarizing the 2 discussion papers: We refer to Doser and Pacifici (2025) as DP and Pautrel et al. (2025) as PEG throughout this rejoinder. DP describe an alternative change of support model that connects species occupancy at areal survey units to an underlying point process in continuous space. They also discuss how their approach compares to the continuous-space occupancy model we proposed. PEG discuss many aspects of our proposed model and suggest a variety of directions for future research. Their suggestions include alternative model implementations that could be considered as well as ideas for continuous space-time occupancy models. We again want to thank all of the discussants—we appreciate their contributions to this important topic in spatial occupancy models.

2 SPECIES OCCUPANCY AND INDIVIDUAL MOVEMENT

Both DP and PEG make the point that movement of individual animals is related to species occupancy and can impact inferences from statistical models. In particular, animal movement is directly tied to the closure assumption in classical occupancy models (see Section 3). The statistical analysis of animal movement data has risen in popularity (Hooten et al., 2017) and it is natural to consider these processes when analyzing ecological data in space (eg, Glennie et al., 2021; McClintock et al., 2022). The connection to individual animal movement makes it more realistic to define species occupancy in continuous space. An individual's home range is typically defined as the spatial area where that individual lives and uses to carry out its normal activities over a defined period of time (Burt, 1943). This idea is well-studied in the field of ecology and there are a variety of methods to estimate individual home ranges using animal movement data (eg, Christ et al., 2008; Wilson et al., 2010, Silva et al., 2022). While occupancy studies collect species-level data and not individual-level data, we can directly connect the home range idea to species-level occupancy.

Consider a study region $\mathcal{S}$ that contains M total individuals and suppose we know the home range $\mathcal{R}_m \subseteq \mathcal{S}$ for all individuals $m = 1, \ldots, M$. We consider a point location occupied if it is within the home range of any individual. Therefore, the total area within the study region that is occupied by a species is simply the union of the individual home ranges $\cup\mathcal{R}_m$. This idea defines species occupancy for any point location $\mathbf{s} \in \mathcal{S}$ such that $z(\mathbf{s}) = 1$ if $\mathbf{s} \in \cup\mathcal{R}_m$ and is 0 otherwise. This connection to animal movement reinforces why it is important to define occupancy in continuous space. With that said, discrete-space occupancy models are also useful and PEG provide examples of

Received: April 11, 2025; Accepted: April 24, 2025

where discrete-space occupancy models may be preferred. However, there are many cases where it is more realistic to model occupancy in continuous space than in discrete survey units.

Hefley and Hooten (2016) presented species distribution processes and data in a hierarchical framework. They described species distributions and their various forms of data (including occupancy data) in a natural spatio-temporal setting and showed how they can be modeled using point processes (among other techniques). The point process model described by DP also connects occupancy to the locations of individual animals and falls into the class of models described by Hefley and Hooten (2016). The DP approach takes a "snapshot" perspective of species occupancy in the sense that it assumes individuals do not move between areal survey units for the duration of the study. Consequently, their point process model defines species occupancy at areal units but not at each point location **s**. These differences highlight another important characteristic of our proposed continuous-space occupancy model. Our model defines species occupancy in such a way that it does not depend on the sites or how surveys are conducted. Instead, it is solely based on the underlying ecological process describing how individuals use space in the study region. We agree with PEG that defining occupancy so that it is independent of the survey units is beneficial and could be useful for comparing inferences from studies with different sampling designs.

3 CLOSURE ASSUMPTION

In classical occupancy models, the sites are assumed to be closed and there are many papers that discuss this topic (eg, MacKenzie et al., 2002, 2017; Kendall and White, 2009; DiRenzo et al., 2022; Valente et al., 2024). Both DP and PEG raised questions about how this assumption should be interpreted in continuous-space models. The connections between species occupancy and individual movement are also useful for considering these questions. First, however, it is important to note that when mapping distributions of mobile species, standard occupancy models are typically interpreted as providing inference about species "use" of sites (MacKenzie et al., 2004; Nichols et al., 2008; Valente et al., 2024). This reflects the fact that the species may not be located within a site at every moment in time and therefore not available for detection. Thus, the terms "occupancy" and "use" are often used interchangeably.

How inferences are affected by violations of the closure assumption due to the movement of individuals can depend upon the duration of surveys, timing of surveys, size of home ranges, and movement speed. For instance, Valente et al. (2024) simulated trajectories of individuals and the corresponding standard occupancy data. Their results showed that spacing the repeat surveys far apart relative to the movement characteristics of individuals led to unbiased inferences for seasonal occupancy (use). This is generally consistent with other research focusing on scenarios when the species may be unavailable at a site due to individual movement (eg, Kendall and White, 2009). Specifically, if availability at a site is independent across the repeated visits, availability becomes a component of the detection process and does not necessarily bias inferences for occupancy (Kendall and White, 2009; DiRenzo et al., 2022). If the probability of being available within a site is very low, then effective detection probabilities are low and this can lead to biased inferences (DiRenzo et al., 2022).

We expect these considerations for standard occupancy models to be helpful when considering closure in our continuous-space approach. Following our definition of occupancy in Section 2, for mobile species, we are interested in the collective use $\cup\mathcal{R}_m$ within a study region. For our continuous-space model, we assume that species occupancy does not change which corresponds to static home ranges for the duration of the study. During some surveys of occupied sites, the species will be unavailable for detection due to all individuals being outside the site boundary. As in standard occupancy models, we expect the availability to be accommodated by the detection process assuming that visits are spaced sufficiently far apparent that the locations of individuals are approximately independent among different visits. In fact, our proposed detection model related detection probabilities to the within-site occupancy proportion. A positive relationship between detection probabilities and within-site occupancy proportions could be interpreted in the context of individual availability—an individual is more likely to be available for detection if a larger portion of its home range $\mathcal{R}_m$ is included within the site area $\mathcal{A}_i$. This is a similar concept to the relationship between detectability and abundance described by Royle and Nichols (2003).

As with standard occupancy models, the closure considerations depend on the characteristics of individual home ranges as well as study design considerations like timing of surveys, duration of surveys, and size of sites. We echo PEG that it would be beneficial to further explore how these considerations affect inferences for continuous-space models. Such studies could specifically explore the impacts of individual movement using analogous simulations to those performed by Valente et al. (2024). An additional consideration is the functional form of the detection model in our approach. There could be alternative specifications that better capture the relationship between detection and within-site occupancy proportions. Additionally, there could be other sources of heterogeneity in detection probabilities, such as within-site abundance. The point process model proposed by DP highlights this idea (see also Royle and Nichols, 2003). In general, assessments of model fit should be conducted and can help guide how models should be modified to better meet assumptions. Goodness-of-fit tests (MacKenzie and Bailey, 2004; Wright et al., 2016) and residual diagnostics (Warton et al., 2017; Wright et al., 2019) exist for standard occupancy models—these approaches could be adapted for continuous-space models. In particular, including spatial variability in detection probabilities may be another way to account for heterogeneity in detection probabilities (Wright et al., 2019). This may be needed to capture additional variability in detection probabilities that could be related to abundance, for instance.

4 CONCLUSION

It is an exciting time for research in spatial ecology. New data types and a broadening set of statistical modeling approaches are being explored and developed that help us learn about the natural world. As models increase in complexity to achieve more realism using larger and more varied data sets, our computational needs increase. We are also seeing a shift away from automated black-box software toward community-sourced open software for implementing such models. Research teams in spatial ecology are becoming more diverse and now often include statisticians and data scientists who are adept at developing code for fitting statistical models in a way that leverages parallel environments and cloud computing resources, while using faster and more stable programming languages.

We've noticed a series of expand-contract phases associated with methodological developments in ecological statistics over the years. For example, we sought to develop integrated likelihoods to fit early forms of ecological models but then as computing power increased and hierarchical modeling rose in prominence, we saw a wide variety of new model forms proposed that take advantage of MCMC computing, for example. After a decade of that, many researchers have started seeking more compact integrated likelihood forms again so they could stabilize computing and take advantage of newer techniques such as Hamiltonian Monte Carlo.

We see continuous-space occupancy models as part of an "expand phase" in the field. Admittedly, in their current form, these newer models require more computing resources than their predecessors, as both DP and PEG pointed out. We are encouraged by newer computational techniques and ways to treat the implementation in a sequence of computing stages, for example, that will facilitate the adoption of these models and allow ecologists to take advantage of this newer technology to answer important spatial ecological questions (eg, Hooten et al., 2021; McCaslin et al., 2021). For example, such methods are already being applied to accelerate the implementation of capture-recapture models (Hooten et al., 2023; 2024) and may also be helpful for fitting spatial occupancy models in the future.

FUNDING

None declared.

CONFLICT OF INTEREST

None declared.

REFERENCES

Burt, W. H. (1943). Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy*, 24, 346–352.

Christ, A., Hoef, J. V. and Zimmerman, D. L. (2008). An animal movement model incorporating home range and habitat selection. *Environmental and Ecological Statistics*, 15, 27–38.

Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographic Systems*, 3, 159–180.

DiRenzo, G. V., Miller, D. A. W. and Grant, E. H. C. (2022). Ignoring species availability biases occupancy estimates in single-scale occupancy models. *Methods Ecol. Evol.*, 13, 1790–1804.

Doser, J. W. and Pacifici, K. (2025). Discussion on "Continuous-space occupancy models" by Wilson J. Wright and Mevin B. Hooten. *Biometrics*, 81, 1–3.

Gelfand, A. E., Zhu, L. and Carlin, B. P. (2001). On the change of support problem for spatio-temporal data. *Biostatistics*, 2, 31–45.

Glennie, R., Buckland, S. T., Langrock, R., Gerrodette, T., Ballance, L., Chivers, S. et al. (2021). Incorporating animal movement into distance sampling. *Journal of the American Statistical Association*, 116, 107–115.

Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97, 632–648.

Hefley, T. J. and Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports*, 1, 87–97.

Hooten, M. B., Johnson, D. S. and Brost, B. M. (2021). Making recursive Bayesian inference accessible. *The American Statistician*, 75, 185–194.

Hooten, M. B., Johnson, D. S., McClintock, B. T. and Morales, J. M. (2017). *Animal Movement: Statistical Models for Telemetry Data*. Boca Raton, FL: CRC press.

Hooten, M. B., Schwob, M. R., Johnson, D. S. and Ivan, J. S. (2023). Multistage hierarchical capture–recapture models. *Environmetrics*, 34, e2799.

Hooten, M. B., Schwob, M. R., Johnson, D. S. and Ivan, J. S. (2024). Geostatistical capture–recapture models. *Spatial Statistics*, 59, 100817.

Kendall, W. L. and White, G. C. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *J. Appl. Ecol.*, 46, 1182–1188.

MacKenzie, D. I. and Bailey, L. L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, 9, 300–318.

MacKenzie, D. I., Bailey, L. L. and Nichols, J. D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, 73, 546–555.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J. and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. and Hines, J. E. (2017). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. San Diego, CA: Elsevier.

McCaslin, H. M., Feuka, A. B. and Hooten, M. B. (2021). Hierarchical computing for hierarchical models in ecology. *Methods in Ecology and Evolution*, 12, 245–254.

McClintock, B. T., Abrahms, B., Chandler, R. B., Conn, P. B., Converse, S. J., Emmet, R. L. et al. (2022). An integrated path for spatial capture–recapture and animal movement modeling. *Ecology*, 103, e3473.

Nichols, J. D., Bailey, L. L., O'Connell, A. F., Talancy, N. W., Campbell Grant, E. H., Gilbert, A. T. et al. (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, 45, 1321–1329.

Pautrel, L., Etienne, M.-P. and Gimenez, O. (2025). Discussion on "Continuous-space occupancy models" by Wilson J. Wright and Mevin B. Hooten. *Biometrics*, 81, 1–4.

Royle, J. A. and Nichols, J. D. (2003). Estimating abundance from repeated presence–absence data or point counts. *Ecology*, 84, 777–790.

Silva, I., Fleming, C. H., Noonan, M. J., Alston, J., Folta, C., Fagan, W. F. et al. (2022). Autocorrelation-informed home range estimation: A review and practical guide. *Methods Ecol. Evol.*, 13, 534–544.

Valente, J. J., Jirinec, V. and Leu, M. (2024). Thinking beyond the closure assumption: Designing surveys for estimating biological truth with occupancy models. *Methods in Ecology and Evolution*, 15, 2289–2300.

Warton, D. I., Stoklosa, J., Guillera-Arroita, G., MacKenzie, D. I. and Welsh, A. H. (2017). Graphical diagnostics for occupancy models with imperfect detection. *Methods in Ecology and Evolution*, 8, 408–419.

Wilson, R. R., Hooten, M. B., Strobel, B. N. and Shivik, J. A. (2010). Accounting for individuals, uncertainty, and multiscale clustering in core area estimation. *The Journal of Wildlife Management*, 74, 1343–1352.

Wright, W. J. and Hooten, M. B. (2025). Continuous-space occupancy models. *Biometrics*, 81, 1–11.

Wright, W. J., Irvine, K. M. and Higgs, M. D. (2019). Identifying occupancy model inadequacies: can residuals separately assess detection and presence? *Ecology*, 100, e02703.

Wright, W. J., Irvine, K. M. and Rodhouse, T. J. (2016). A goodness-of-fit test for occupancy models with correlated within-season revisits. *Ecology and Evolution*, 6, 5404–5415.

Received: April 11, 2025; Accepted: April 24, 2025

Biometrics, 2025, 81(2), ujaf055
https://doi.org/10.1093/biomtc/ujaf055
Biometric Methodology Discussion Paper

Continuous-space occupancy models

Wilson J. Wright [1,*,†] and Mevin B. Hooten [2]

[1]Department of Statistics, Colorado State University, Fort Collins, CO 80523, United States, [2]Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX 78705, United States

[*]Corresponding author: Wilson J. Wright, Department of Statistics, Colorado State University, 102 Statistics Building, 1877 Campus Delivery, Fort Collins, CO 80523, U.S. (wilson.wright@colostate.edu).

[†]Current affiliation: Department of Statistics, University of Missouri, Columbia, MO 65211, U.S.

ABSTRACT

Occupancy models are used to infer species distributions over large spatial extents while accounting for imperfect detection. Current approaches, however, are unable to model species occurrence over continuous spatial domains while accounting for the discrete spatial domain of the observed data. We develop a new class of spatial occupancy models that embeds a change of spatial support between the observed data and occurrence process. We use a clipped Gaussian process to represent species occurrence in continuous space, which can provide inferences at a finer resolution than the observed occupancy data. Our approach is beneficial because it allows for more realistic models of species occurrence, can account for species occurring in only a portion of a surveyed site, and can relate detection probabilities to these within-site occurrence proportions. We show how our model can be fit using Bayesian methods and develop a computationally efficient MCMC algorithm. In particular, we rely on a Vecchia approximation to implement the spatial Gaussian process describing species occurrence and develop a surrogate data approach for jointly updating the spatial terms and spatial covariance parameters. We demonstrate our model using simulated data and compare our approach to alternative spatial occupancy models. We also use our model to analyze ovenbird occurrence data collected in New Hampshire, USA.

KEYWORDS: Bayesian statistics; change of spatial support; hierarchical model; nearest neighbor Gaussian process; spatial statistics; species distributions.

1 INTRODUCTION

Mapping the distributions of wildlife species is a fundamental component of many ecological studies and wildlife monitoring programs. Occupancy models (Hoeting et al., 2000; MacKenzie et al., 2002; 2018) have become an invaluable approach for modeling species distributions because they account for detection errors that are prevalent in many ecological surveys. This is imperative for obtaining unbiased inferences about species occurrence and how it relates to predictor variables of interest. Additionally, occupancy models are widely applied because they can be used to analyze data from multiple types of surveys, are applicable to a variety of different taxa, and are particularly useful when monitoring species over large spatial extents (MacKenzie et al., 2002; Noon et al., 2012).

Species distributions are the result of inherently spatial processes, and there are multiple approaches for modeling spatial dependence in occupancy data (Gelfand and Shirota, 2019; Hefley and Hooten, 2016; Latimer et al., 2006). For instance, spatial dependence in site-level occupancy probabilities can be modeled using conditionally autoregressive terms (in discrete space; Johnson et al., 2013; Broms et al., 2014) or with spatial terms (in continuous space) modeled as a Gaussian process (Ovaskainen et al., 2016; Wright et al., 2021; Doser et al., 2023). These alternative approaches make different assumptions about the spatial support of the process of interest—highlighting a challenge for modeling the spatial structure in occupancy data. While species occupancy is typically viewed as arising from a continuous spatial process (Hooten et al., 2003; Efford and Dawson, 2012), the observed data are collected during surveys of areal units (MacKenzie et al., 2002, 2018). Current spatial occupancy models are unable to account for the change of spatial support between the occupancy and observation processes.

Change of support methods provide a way to make inferences for spatial units that have a support that differs from that of the observed data (Cressie, 1996; Gelfand et al., 2001; Gotway and Young, 2002). These methods are currently available for continuous data (Cressie, 1996; Gelfand et al., 2001), count data (Bradley et al., 2016), and binary data that have been aggregated to areal units (Walker et al., 2021; 2020). Ignoring a change in spatial support can result in biased predictions (Cressie, 1996) and biased inferences for regression coefficients (Walker et al., 2020). Previous studies have used Poisson point processes to handle changes of support when modeling species distributions (Koshkina et al., 2017; Pacifici et al., 2019). However, these methods focus on integrating different types of data, which are observed at areal units of different sizes (Koshkina et al., 2017; Pacifici et al., 2019), rather than explicitly modeling species occupancy in continuous space. We develop a new framework that treats occupancy as a binary process in continuous space and learn about this process using data observed at areal survey units. This provides an approach for modeling a change in spatial support for a new type of data compared to previous spatial models.

Received: April 15, 2024; Revised: November 5, 2024; Accepted: April 24, 2025

Our continuous-space occupancy framework is beneficial because it allows for more realistic spatial models of species occurrence while still properly accounting for the discrete spatial support of the observed data. This provides multiple advantages compared to previously developed spatial occupancy models. For example, our approach allows inferences from areal data to be downscaled, accounting for the fact that areal survey units may only be partially occupied by a species. The observation component of our model can relate these within-unit occupancy proportions to the probability of detecting a species during surveys. This is not possible using standard occupancy models because occupancy is defined as a binary random variable at the level of the survey unit. Another benefit of our model is that it allows for improved inferences about the proportion of area occupied by a species. While this quantity is often of interest, the phrase "proportion of area occupied" can be misleading because modeling species occurrence at the site-level, as done in standard approaches, only permits inferences about the proportion of *sites* occupied (Efford and Dawson, 2012). Our continuous-space occupancy framework directly addresses this limitation and provides inferences for the proportion of area occupied in continuous space.

We implement our model using Bayesian methods and develop a computationally efficient Markov chain Monte Carlo (MCMC) algorithm for fitting our model. We assume that the binary spatial process for occupancy arises from clipping a latent continuous field that is modeled using a Gaussian process (De Oliveira, 2000, 2020). To improve the computational efficiency of fitting this model over potentially large spatial extents and at many point locations, we approximate the latent Gaussian process using a nearest neighbor Gaussian process (NNGP; Datta et al., 2016). The NNGP approximation makes implementation of spatial models for extremely large datasets possible and helps facilitate Bayesian computation because the necessary calculations associated with the spatial terms are much faster (Datta et al., 2016; Finley et al., 2019). However, MCMC can still be slow to mix for the spatial terms (Finley et al., 2019) and spatial covariance parameters (Murray and Adams, 2010). To address these challenges, we developed an elliptical slice sampler (Murray et al., 2010) to update the spatial terms after marginalizing over the regression coefficients. The conventional elliptical slice sampler assumes that the spatial covariance parameters are fixed. We relax this assumption by modifying the surrogate data slice sampler proposed by Murray and Adams (2010) to be more compatible with NNGPs. The surrogate data slice sampler updates the spatial terms and spatial covariance parameters jointly. This approach is sufficiently general that it could be applied to other spatial models with latent NNGPs.

The remainder is organized as follows. In Section 2, we describe occupancy data and the standard models used to analyze these data. Our new model for occupancy in continuous space is presented in Section 3, and the details of the MCMC algorithm we used to fit this model appear in Section 4. In Section 5, we illustrate our approach using a simulated example and perform a simulation study to compare our continuous-space occupancy model to alternative spatial occupancy models. In Section 6, we analyze ovenbird occurrence data collected in the Hubbard Brook Experimental Forest, New Hampshire, USA. Section 7 discusses future directions that build upon this research. Additional details about our MCMC algorithm, NNGP calculations, and alternative spatial occupancy models are provided in the Web Appendices, which are available in the online Supplementary Material.

2 OCCUPANCY DATA AND STANDARD ANALYSES

We begin with an overview of the typical data available for conventional occupancy analyses (see also MacKenzie et al., 2002, 2018). Data on species occupancy are collected at areal survey units called "sites." We let $\mathcal{A}_i \subseteq \mathcal{S}$ for $\mathcal{S} \subset \mathbb{R}^2$ denote the region defining site i for $i = 1, \ldots, n$. Binary detection/non-detection data are collected at each site i during visits j for $j = 1, \ldots, J_i$ where J_i denotes the total number of visits to site i. In standard occupancy models, the binary data y_{ij} are modeled as

$$y_{ij} \sim \begin{cases} \mathrm{I}(y_{ij} = 0), & z_i = 0 \\ \mathrm{Bernoulli}(p_{ij}), & z_i = 1 \end{cases}, \tag{1}$$

where z_i denotes a partially observed binary random variable for whether the species is present (1) or not (0) at site i, and p_{ij} is the probability of detecting the species during visit j to site i. The detection probabilities can be modeled as a function of predictor variables using a generalized linear model framework. When the species is not present at a site (ie, $z_i = 0$), we assume that there are no false positive detections and $y_{ij} = 0$ for all j with probability 1. However, this assumption can be relaxed, and there are approaches for modeling false positive detections (eg, Chambert et al., 2015; Ruiz-Gutierrez et al., 2016).

Standard approaches also model the occupancy process at the site-level and therefore assume a discrete spatial domain. That is, occupancy at each site is modeled as

$$z_i \sim \mathrm{Bernoulli}(\psi_i), \tag{2}$$

where ψ_i denotes the probability of occupancy at site i, which is modeled as a function of spatial predictor variables. A site is considered occupied if the species occurs anywhere within the site area $\mathcal{A}_i$, and we address this in our continuous-space model in Section 3. Spatial occupancy models allow ψ_i to have additional spatial structure. This spatial structure can be included using a conditional autoregressive term (Johnson et al., 2013; Broms et al., 2014) or with spatial effects modeled using a continuous Gaussian process based on the locations of the site centroids (Ovaskainen et al., 2016; Wright et al., 2021; Doser et al., 2023).

The discretization of the spatial domain imposed by defining sites is generally arbitrary and does not necessarily have an ecological interpretation. Additionally, occupancy is a process in continuous space (Efford and Dawson, 2012), and it is possible for a species to occur in only a portion of a site (eg, Figure 1). How well standard models can approximate the underlying spatial occupancy process depends on the resolution of the discretization defined by the sites. However, the proportion of *sites* occupied will always be higher than the proportion of the *study area* that is occupied when considering observations over a regular grid (eg, Figure 1; see also Efford and Dawson, 2012). We define an alternative approach for analyzing species

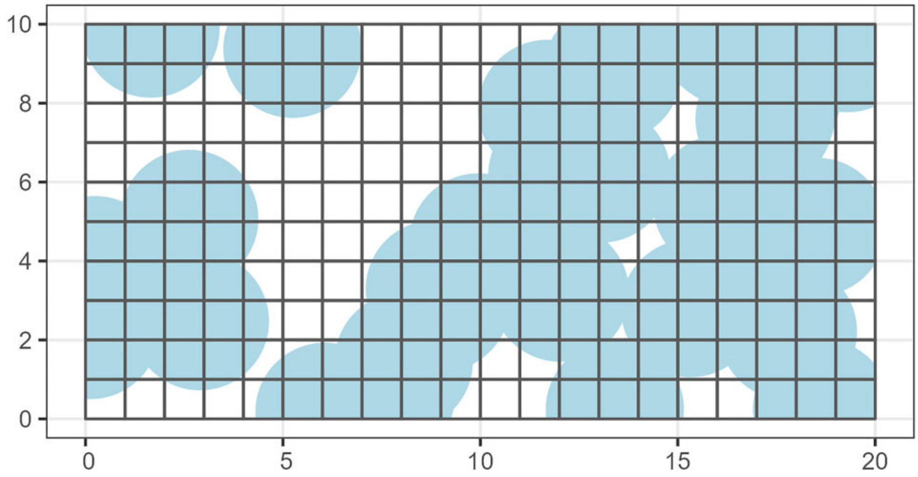

FIGURE 1 Hypothetical example where a regular grid defines sites throughout the study region and species occurrence is shown by the shaded regions. The occupancy process is defined for continuous space even though the detection data are collected at areal sites that discretize the spatial domain. In this example, 72.3% of the study region is occupied, but the species occurs in 87.5% of the sites.

occupancy data that models the occupancy process on a continuous spatial domain. The primary challenge for this new approach is that the observation data are still measured at areal sites and we must account for the resulting change of support (eg, Cressie, 1996; Gelfand et al., 2001; Gotway and Young, 2002) between the occupancy process and the observed data.

3 MODEL

We model species occurrence for spatial locations $\mathbf{s} \in \mathcal{S} \subseteq \mathbb{R}^2$ as a clipped Gaussian process (De Oliveira, 2000, 2020) defined over continuous space. We let $z(\mathbf{s})$ denote the binary spatial process for whether the species occurs (1) at location $\mathbf{s}$ or not (0). Note that we define species occurrence $z(\mathbf{s})$ for any point location $\mathbf{s}$, which differs from the site-level occupancy z_i that is used in standard occupancy models (1) and (2). We assume that this binary spatial process $z(\mathbf{s})$ arises from clipping a latent continuous process $\widetilde{z}(\mathbf{s})$ such that $z(\mathbf{s}) = \mathrm{I}(\widetilde{z}(\mathbf{s}) \geq 0)$ and

$$\widetilde{z}(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + \eta(\mathbf{s}), \tag{3}$$

where $\mathbf{x}(\mathbf{s})$ is a vector of spatially indexed predictor variables with corresponding coefficients $\boldsymbol{\beta}$ and $\eta(\mathbf{s})$ is a spatial Gaussian process with mean zero and spatial covariance function K_η. We assume the spatial predictor variables $\mathbf{x}(\mathbf{s})$ are available at any point location $\mathbf{s} \in \mathcal{S}$, although, if not, they could be modeled as a separate Gaussian process.

The detection/non-detection data y_{ij} are recorded at sites $i = 1, \ldots, n$ and visits $j = 1, \ldots, J_i$. We still assume there are no false positive detections, and thus the species can only be detected if it occupies at least a portion of the site. Therefore, we model the detection data y_{ij} as

$$y_{ij} \sim \begin{cases} \mathrm{I}(y_{ij} = 0), & \max_{\mathbf{s} \in \mathcal{A}_i} z(\mathbf{s}) = 0 \\ \text{Bernoulli}(p_{ij}), & \max_{\mathbf{s} \in \mathcal{A}_i} z(\mathbf{s}) = 1, \end{cases} \tag{4}$$

$$\Phi^{-1}(p_{ij}) = \mathbf{w}_{ij}'\boldsymbol{\alpha} + \gamma |\mathcal{A}_i|^{-1} \int_{\mathcal{A}_i} z(\mathbf{s}) d\mathbf{s}, \tag{5}$$

where the probit link function $\Phi(\cdot)$ denotes the cumulative distribution function for a standard normal random variable, $\mathbf{w}_{ij}$ is a vector of predictor variables related to detection, and $|\mathcal{A}_i|^{-1} \int_{\mathcal{A}_i} z(\mathbf{s}) d\mathbf{s}$ is the proportion of the site area where the species occurs. The occupancy proportion within a site is related to the probability of detection with corresponding parameter γ. In general, detection probabilities should increase as the proportion of the site that is occupied increases. The observation process of our model is similar to that of the standard occupancy model defined in Section 2 except that we explicitly define site-level occupancy as a function of the continuous occupancy process, and we allow the detection probability to depend on the proportion of the site that is occupied. This observation process also provides more flexibility than standard models because we do not need to assume that the site regions $\mathcal{A}_i$ are mutually disjoint (eg, see Section 6).

Our model accounts for the change of support between the observation and occupancy processes. This allows for inferences about occupancy to be downscaled to continuous space even though the observed data are measured at the site-level. In other words, the observed data are recorded at areal sites, but we are still able to obtain inferences for occupancy in continuous space—we do not need to assume the entire site is occupied when a species is detected there. Modeling occupancy in continuous space is beneficial because it allows for more realistic inferences about this ecological process. Additionally, it allows us to relate the detection process to the proportion of a site that is occupied. This relationship is intuitive but cannot be incorporated into traditional occupancy models.

4 PRIORS AND IMPLEMENTATION

We implement our model using Bayesian methods and assume normal prior distributions for $\boldsymbol{\alpha}$, γ, and $\boldsymbol{\beta}$. Using a Gibbs sampling approach, these priors facilitate conjugate updates for many of the parameters (see Web Appendix A). The main challenges for implementing this model are associated with updating the latent spatial terms $\widetilde{z}(\mathbf{s})$ and the spatial covariance parameters $\boldsymbol{\theta}$. We describe how we address these challenges and provide the details for our MCMC algorithm throughout the rest of this section and in Web Appendix A. All analyses for our simulated (Section 5) and real data examples (Section 6) were conducted in R (R Core Team, 2022). We also used the `Rcpp` (Eddelbuettel and Balamuta, 2018) and `RcppArmadillo` packages (Eddelbuettel and Sanderson, 2014) to code our MCMC algorithm.

4.1 Numerical quadrature

While our occupancy model is defined for continuous space, we consider only a finite number of locations to approximate $\widetilde{z}(\mathbf{s})$. We let $\widetilde{\mathbf{z}} \equiv (\widetilde{z}(\mathbf{s}_1), \ldots, \widetilde{z}(\mathbf{s}_D))'$ define the vector used to implement this approximation. The locations $\mathbf{s}_d$ for $d = 1, \ldots, D$ are chosen to cover all sites in the study region, and we provide more details on choosing these locations later in this section. Similarly, we define design matrix $\mathbf{X} \equiv (\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_D))'$ and vector of spatial terms $\boldsymbol{\eta} \equiv (\eta(\mathbf{s}_1), \ldots, \eta(\mathbf{s}_D))'$ at the same point locations. The finite-dimensional occupancy process can now be modeled as $\widetilde{\mathbf{z}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$ and $\boldsymbol{\Sigma}_\eta$ are defined by the spatial covariance function K_η. Additionally, we use numerical quadrature to approximate the functions of $z(\mathbf{s})$ in (4) and (5) as

$$\max_{\mathbf{s} \in \mathcal{A}_i} z(\mathbf{s}) \approx \max_{\mathbf{s}_d \in \mathcal{A}_i} \mathrm{I}(\widetilde{z}(\mathbf{s}_d) \geq 0) \tag{6}$$

and

$$|\mathcal{A}_i|^{-1} \int_{\mathcal{A}_i} z(\mathbf{s}) d\mathbf{s} \approx D_i^{-1} \sum_{\mathbf{s}_d \in \mathcal{A}_i} \mathrm{I}(\widetilde{z}(\mathbf{s}_d) \geq 0), \tag{7}$$

where D_i is the total number of points $\mathbf{s}_d$ contained in $\mathcal{A}_i$. If the sites define a regular grid over the study region, then we define point locations $\mathbf{s}_d$ such that each site contains the same number of points. The errors associated with both of these approximations can be made arbitrarily small by making D sufficiently large.

4.2 Nearest neighbor Gaussian process

As with many spatial models, increasing the number of points D used to approximate the latent spatial process can result in this model becoming computationally challenging to implement. We utilize a nearest neighbor Gaussian process (NNGP) to approximate $\widetilde{z}(\mathbf{s})$, which results in the finite-dimensional distribution of $\widetilde{\mathbf{z}}$ having a sparse precision matrix (Datta et al., 2016). We provide a general description of NNGPs in what follows, and more details about constructing the resulting sparse precision matrix can be found in Web Appendix B. The NNGP approach relies on a Vecchia approximation (Vecchia, 1988) of the joint distribution of a spatial process. Using standard factorization properties, the joint distribution of the spatial terms from our model can be written as

$$[\widetilde{\mathbf{z}}] = [\widetilde{z}_1][\widetilde{z}_2 \mid \widetilde{z}_1] \cdots [\widetilde{z}_D \mid \widetilde{z}_1, \ldots, \widetilde{z}_{D-1}], \tag{8}$$

where we use $[\cdot]$ to denote a probability density function (Gelfand and Smith, 1990). The Vecchia approximation defines the conditional distributions in the factorization (8) to only depend on a set of nearest neighbors selected from the previous observations. Consequently, we can approximate the joint distribution of the spatial terms as

$$[\widetilde{\mathbf{z}}] \approx \prod_{d=1}^{D} [\widetilde{z}_d \mid \widetilde{\mathbf{z}}_{c(d)}], \tag{9}$$

where $c(d) \subseteq \{1, \ldots, d-1\}$ defines a set of nearest neighbors among the previous terms ($c(1)$ is the null set) and $\widetilde{\mathbf{z}}_{c(d)}$ denotes a vector containing the spatial terms in that set.

The NNGP approximation requires decisions about how the D points are ordered and how the neighbors $c(d)$ are selected. We use the "maxmin" ordering proposed by Guinness (2018) and select the m nearest neighbors based on the spatial distances to the previous points based on this ordering (Vecchia, 1988; Datta et al., 2016). The maxmin ordering chooses the next location to be the one that maximizes the minimum distance to previous points and can substantially improve the accuracy of NNGP approximations (Guinness, 2018). Additionally, for at least some points, this ordering will include neighbors that have large distances from the point of interest, which can provide more information about parameters in the spatial covariance function (Stein et al., 2004). The exact maxmin ordering can be computationally demanding to calculate, and thus we approximate this ordering (see Web Appendix B).

4.3 Updating the spatial terms

The NNGP approximation allows for improved computation of the joint density of the spatial terms that can be utilized in our MCMC algorithm. However, updating the spatial terms $\widetilde{\mathbf{z}}$ can still be challenging because they are highly correlated and, consequently, MCMC chains can converge very slowly (Datta et al., 2016; Finley et al., 2019). When the data likelihood is Gaussian, marginalizing over the spatial random terms can improve the mixing of MCMC (Shi et al., 2017; Finley et al., 2019). This approach is not straightforward for our model because the likelihood defined in (4) and (5) becomes challenging to evaluate. We consider an alternative approach that marginalizes over the coefficients $\boldsymbol{\beta}$ and updates the spatial terms jointly using an elliptical slice sampler (Murray et al., 2010) or a surrogate data slice sampler (Murray and Adams, 2010).

We describe our MCMC algorithm when the spatial covariance parameters are fixed (ie, $\boldsymbol{\Sigma}_\eta$ is known) and then generalize this algorithm to allow the spatial covariance parameters to be modeled as well. Assuming the prior distribution $\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, integrating $\boldsymbol{\beta}$ from (3) implies

$$\widetilde{\mathbf{z}} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\mu}_\beta, \mathbf{X}\boldsymbol{\Sigma}_\beta\mathbf{X}' + \boldsymbol{\Sigma}_\eta), \tag{10}$$

for the finite-dimensional locations. We denote the mean and variance in (10) by $\boldsymbol{\mu}_{\widetilde{z}}$ and $\boldsymbol{\Sigma}_{\widetilde{z}}$, respectively. The marginal distribution for $\widetilde{\mathbf{z}}$ is approximated using the NNGP approach described in Section 4.2 and Web Appendix B. Conditional on the other parameters in the model, these spatial terms can be updated from the full-conditional distribution

$$[\widetilde{\mathbf{z}} \mid \mathbf{y}, \cdot] \propto [\widetilde{\mathbf{z}}][\mathbf{y} \mid \widetilde{\mathbf{z}}, \cdot], \tag{11}$$

where $[\mathbf{y} \mid \widetilde{\mathbf{z}}, \cdot]$ denotes the observed data likelihood conditional on the spatial terms and all other parameters in the model. The form of (11) can be sampled from using an elliptical slice sampler (Murray et al., 2010) because the prior $[\widetilde{\mathbf{z}}]$ is multivariate normal. Elliptical slice sampling is appealing for models using a latent Gaussian process because there are no restrictions on the form of the likelihood, it is easy to implement, and does not require tuning (Murray et al., 2010). Conditional on $\widetilde{\mathbf{z}}$, we sample $\boldsymbol{\beta}$ from its full-conditional distribution.

When the spatial covariance parameters $\boldsymbol{\theta}$ are not fixed, we update them and the spatial terms $\widetilde{\mathbf{z}}$ jointly using a modified version of the surrogate data slice sampler developed by Murray and Adams (2010). A joint update for these parameters is important because they are highly correlated in the posterior distribution, and conditional updates result in poor mixing of the MCMC chains (Murray and Adams, 2010). The surrogate data slice sampler introduces auxiliary parameters into the model that allow for a series of convenient Gibbs updates for $\widetilde{\mathbf{z}}$ and $\boldsymbol{\theta}$. Directly applying the approach from Murray and Adams (2010) would introduce surrogate data for every spatial location $\mathbf{s}_d$. We found that this was not conducive to using the NNGP approach for the spatial terms and instead introduce surrogate data for only a subset of the spatial locations. This allows us to retain the computationally efficient calculations facilitated by the NNGP when applying the surrogate data slice sampler.

To define our surrogate data slice sampler, we partition the spatial terms such that

$$\widetilde{\mathbf{z}} \equiv \begin{pmatrix} \widetilde{\mathbf{z}}_1 \\ \widetilde{\mathbf{z}}_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right), \tag{12}$$

where we have simplified the notation by assuming $\boldsymbol{\mu}_\beta = \mathbf{0}$, which implies that $\boldsymbol{\mu}_{\widetilde{z}} = \mathbf{0}$, and omitting the $\widetilde{z}$ subscripts from the partitioned covariance matrix. The partitions are chosen such that $\widetilde{\mathbf{z}}_1$ has relatively few locations—these are the locations where we introduce surrogate data. We expand this model with auxiliary variables $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$, and $\mathbf{g}$ that are assumed to be marginally distributed

$$\begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \mathbf{g} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{11} + \boldsymbol{\Sigma}_g \end{pmatrix} \right), \tag{13}$$

where $\boldsymbol{\Sigma}_g$ is a user-specified covariance matrix corresponding to surrogate data $\mathbf{g}$. Conditional on these auxiliary variables, we set

$$\widetilde{\mathbf{z}}_1 = \mathbf{m}_1 + \mathbf{L}_1 \boldsymbol{\nu}_1, \tag{14}$$

$$\widetilde{\mathbf{z}}_2 = \mathbf{m}_2 + \mathbf{L}_2 \boldsymbol{\nu}_2, \tag{15}$$

where

$$\mathbf{m}_1 = \boldsymbol{\Sigma}_{11} \left(\boldsymbol{\Sigma}_{11} + \boldsymbol{\Sigma}_g \right)^{-1} \mathbf{g}, \tag{16}$$

$$\mathbf{L}_1 \mathbf{L}_1' = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{11} \left(\boldsymbol{\Sigma}_{11} + \boldsymbol{\Sigma}_g \right)^{-1} \boldsymbol{\Sigma}_{11}, \tag{17}$$

$$\mathbf{m}_2 = \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \left(\mathbf{m}_1 + \mathbf{L}_1 \boldsymbol{\nu}_1 \right), \tag{18}$$

$$\mathbf{L}_2 \mathbf{L}_2' = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}. \tag{19}$$

Routine calculations show that this construction of the auxiliary variables induces the same marginal distribution for $(\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2)$ as that in (12). This construction also defines a joint normal distribution for $(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathbf{g}, \widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2)$. Thus, we obtain a sample from our target posterior distribution using a Gibbs sampler for this parameter expanded model.

Conditional on the detection-level parameters, the posterior distribution of the parameter expanded model is $[\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathbf{g}, \widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2, \boldsymbol{\theta} \mid \mathbf{y}]$. The first step of our Gibbs sampler updates the auxiliary parameters from

$$\begin{aligned} &[\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathbf{g} \mid \mathbf{y}, \widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2, \boldsymbol{\theta}] \\ &\quad = [\mathbf{g} \mid \widetilde{\mathbf{z}}_1, \boldsymbol{\theta}][\boldsymbol{\nu}_1, \boldsymbol{\nu}_2 \mid \mathbf{g}, \widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2, \boldsymbol{\theta}], \end{aligned} \tag{20}$$

where $(\mathbf{g} \mid \widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2, \boldsymbol{\theta}) \overset{d}{=} (\mathbf{g} \mid \widetilde{\mathbf{z}}_1, \boldsymbol{\theta}) \sim \mathrm{N}(\widetilde{\mathbf{z}}_1, \boldsymbol{\Sigma}_g)$ is as shown in Web Appendix A. The distribution of $(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2 \mid \mathbf{g}, \widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2, \boldsymbol{\theta})$ is degenerate as implied by (14) and (15). The second step of the Gibbs sampler is to update the spatial terms and covariance parameters from

$$\begin{aligned} &[\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2, \boldsymbol{\theta} \mid \mathbf{y}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathbf{g}] \\ &\quad = [\boldsymbol{\theta} \mid \mathbf{y}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathbf{g}][\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2 \mid \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathbf{g}, \boldsymbol{\theta}], \end{aligned} \tag{21}$$

where $[\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2 \mid \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathbf{g}, \boldsymbol{\theta}]$ is also degenerate by (14) and (15). Updating the spatial covariance parameters from $[\boldsymbol{\theta} \mid \mathbf{y}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathbf{g}]$ can be performed efficiently using a slice sampler because $\boldsymbol{\theta}$ will generally contain only a few parameters (Neal, 2003). As suggested by Murray and Adams (2010), in our full MCMC algorithm, we perform the surrogate data slice sampling step every 10 iterations (simulated data examples) to reduce computation time. For the remaining iterations, we fix $\boldsymbol{\theta}$ and update the spatial terms using elliptical slice sampling.

Our surrogate data slice sampling algorithm uses the NNGP approximation when updating $\boldsymbol{\nu}_2$ and $\widetilde{\mathbf{z}}_2$. In general, we can reparameterize our model by introducing a vector of independent standard normal random variables $\boldsymbol{\nu}$ and redefining the spatial terms as $\widetilde{\mathbf{z}} = \boldsymbol{L}\boldsymbol{\nu}$, where $\boldsymbol{L}$ is a factorization of $\boldsymbol{\Sigma}_{\widetilde{z}}$ (and therefore depends on $\boldsymbol{\theta}$) such that $\boldsymbol{L}\boldsymbol{L}' = \boldsymbol{\Sigma}_{\widetilde{z}}$. Using standard NNGP results, $\boldsymbol{L}^{-1}$ is readily available, and efficient algorithms exist for calculating $\widetilde{\mathbf{z}}$ by solving the sparse system $\boldsymbol{L}^{-1}\widetilde{\mathbf{z}} = \boldsymbol{\nu}$ (Saha and Datta, 2018; Datta, 2022). Given $\widetilde{\mathbf{z}}_1$, (15) defines $\widetilde{\mathbf{z}}_2$ using mean $\mathbf{m}_2$ and variance $\mathbf{L}_2\mathbf{L}_2'$ which are equivalent to the conditional mean $\mathrm{E}(\widetilde{\mathbf{z}}_2 \mid \widetilde{\mathbf{z}}_1)$ and variance $\mathrm{Var}(\widetilde{\mathbf{z}}_2 \mid \widetilde{\mathbf{z}}_1)$, respectively. Thus, in the second Gibbs step of the surrogate data slice sampler, $\widetilde{\mathbf{z}}_2$ can be found using the same iterative algorithm for solving $\boldsymbol{L}^{-1}\widetilde{\mathbf{z}} = \boldsymbol{\nu}$ when using an NNGP approximation (see Web Appendix B for more details). Similarly, we can reverse this algorithm to find $\boldsymbol{\nu}_2$ conditional on $\boldsymbol{\nu}_1$, $\widetilde{\mathbf{z}}_1$, and $\widetilde{\mathbf{z}}_2$ in the first Gibbs step. Introducing the surrogate data for $\widetilde{\mathbf{z}}_1$ does not allow for the same NNGP calculations to be used. However, by choosing the dimension of the surrogate data to be sufficiently small, we can perform the requisite calculations in (14), (16), and (17) exactly. That is, we do not need to rely on the sparsity of $\widetilde{\boldsymbol{\Sigma}}_{11}^{-1} \approx \boldsymbol{\Sigma}_{11}^{-1}$ when updating $\widetilde{\mathbf{z}}_1$ or $\boldsymbol{\nu}_1$.

Tuning is required for the surrogate data slice sampler. First, the number of points to introduce surrogate data must be chosen. In general, we choose $\widetilde{\mathbf{z}}_1$ to include one point per site or grid cell used to order the points as described in Section 4.2. The second choice is the specification of the surrogate data covariance $\boldsymbol{\Sigma}_g$. We specify this covariance to be a diagonal matrix with elements $\sigma_{g,d}^2$ tuned to be approximately twice the posterior variance of the corresponding spatial term $\widetilde{z}_{1,d}$.

5 SIMULATIONS

5.1 Simulated example

We start by demonstrating our model using a simulated data example. We simulated occupancy in continuous space as a clipped

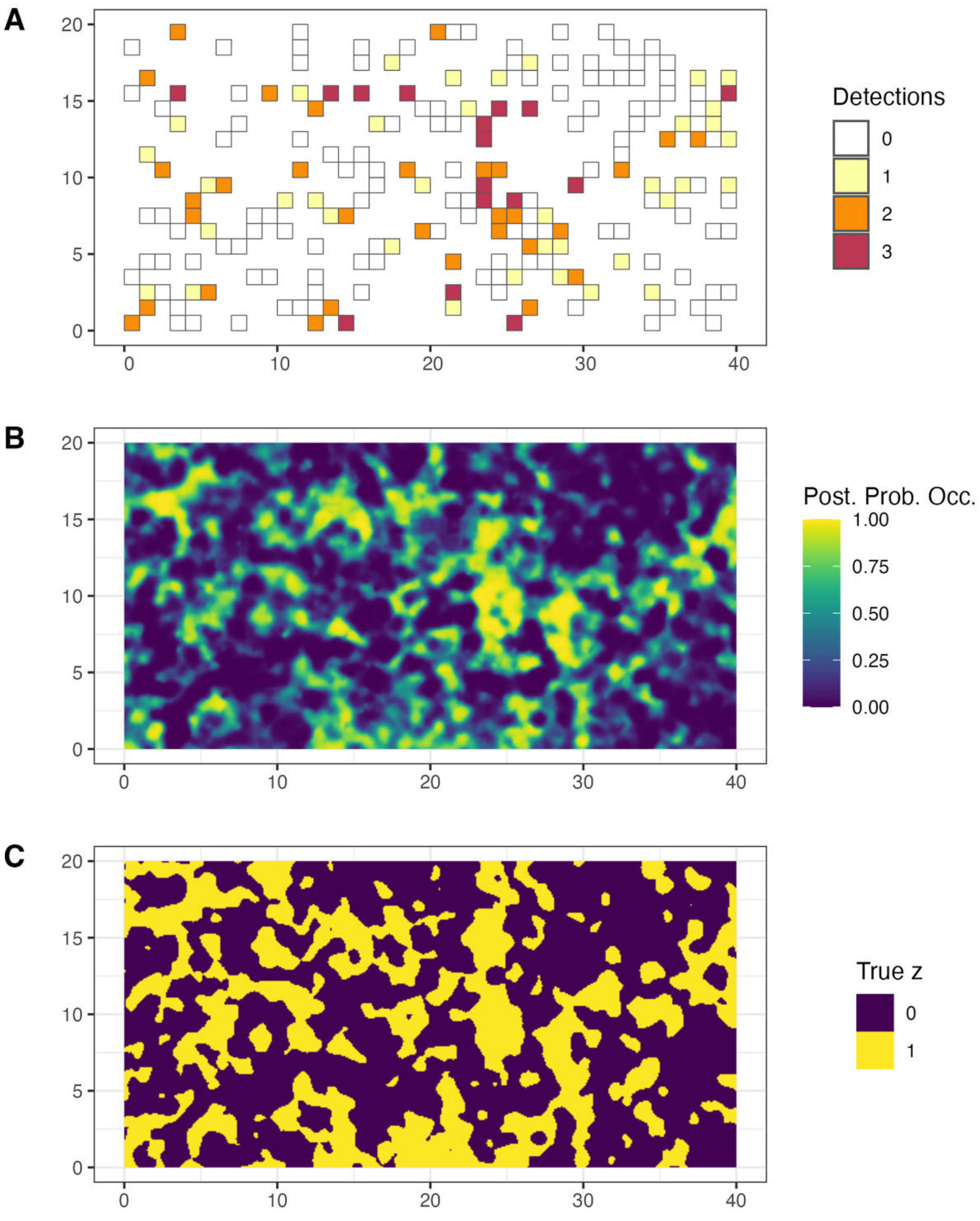

FIGURE 2 Simulated data example with occupancy related to one spatial covariate and an additional covariate related to detection probabilities. The number of observed detections out of three visits is shown in (A) for the sampled sites. After fitting our model, the posterior probability of occupancy (B) is well-aligned with the true underlying occupancy (C).

Gaussian process with a single spatial covariate and a spatial term that had a Gaussian covariance function. We considered a 40 × 20 unit rectangular study area and defined sites using a 1 × 1 unit regular grid over the region. Occupancy data were simulated at 200 randomly selected sites out of the 800 total sites over the region with three visits per site (Figure 2A). We also simulated a visit-level covariate from a Uniform$(-1, 1)$ distribution. Data were generated based on parameters $\boldsymbol{\beta} = (-0.5, 2)'$, $\boldsymbol{\alpha} = (-2, 1)'$, and $\gamma = 3$. Because γ is positive, detection probabilities within a site increase as the proportion of the site that was occupied by the species increases.

We fit our continuous-space occupancy model to these simulated data using 200 000 iterations to tune the surrogate data slice sampler, followed by 200 000 iterations for posterior inference. The 95% marginal credible intervals captured values of the parameters used to generate these data. Our focus is on

TABLE 1 For the proportion of area occupied and proportion of sites occupied, the empirical bias of posterior means and the coverage of 95% CIs for different spatial occupancy models based on analyses of 100 simulated datasets.

	Proportion of area occupied		Proportion of sites occupied	
Model	**Bias**	**Coverage**	**Bias**	**Coverage**
Areal	0.143	0%	−0.268	0%
Centroid	0.156	0%	−0.255	3%
Continuous-space	−0.002	94%	−0.010	92%

Data were simulated under our continuous-space occupancy model. We compared our model to two other approaches that both ignore the change of spatial support. The first approach ("Areal") models the occupancy process at the areal sites used to collect data. The second approach ("Centroid") ignores the defined sites and treats all data as point-level observations corresponding to the site centroids.

inferences for mapping occupancy and summarize the posterior probability of occupancy for spatial locations throughout the region considered. We calculate the marginal posterior probability of occupancy for spatial location $\mathbf{s}$ as $\mathrm{P}(z(\mathbf{s}) = 1 \mid \mathbf{y}) \approx T^{-1} \sum_{t=1}^{T} z(\mathbf{s})^{(t)}$ where $z(\mathbf{s})^{(t)}$ denotes the sampled value for the binary spatial process z on MCMC iteration t and T is the total number of iterations. Overall, the map of the posterior probabilities of occupancy recovers the underlying occupancy process well for this simulated example (Figures 2B and C).

In general, this example illustrates how we can recover an occupancy process in continuous space using data observed at areal sites. The downscaling of inferences is possible because (i) the spatial structure in the occupancy process, (ii) the spatial predictor variable, and (iii) the relationship between detection probabilities and the proportion of a site that is occupied. All of these provide information at a finer resolution than the areal sites.

We also used this example to compare the performance of our surrogate data slice sampler to a Metropolis-Hastings (MH) update for the spatial range parameter ρ. This MH update conditions on $\widetilde{\mathbf{z}}$ when updating ρ, which can lead to slow MCMC mixing because these parameters are highly correlated (Murray and Adams, 2010). This issue is exacerbated as the dimension of $\widetilde{\mathbf{z}}$ increases. In our simulated example, the surrogate data slice sampler resulted in an effective sample size for ρ that was 150 times higher compared to using the MH update.

5.2 Comparisons to other models

We also performed simulations to compare our continuous-space occupancy model to other spatial occupancy models. Data were generated using the same parameter values as specified in Section 5.1. We defined sites using a 1×1 unit regular grid and considered a study area that was 20×15 units in size. Data were generated for three visits at each of the 300 sites within this region. We considered a smaller region and surveys at all sites to simplify the computation in this simulation study.

We generated 100 different datasets and fit three different spatial occupancy models to each. We first fit our continuous-space occupancy model that matches the data-generating process. The first alternative approach models the occupancy process at the areal sites and includes spatial structure among sites using an intrinsic conditional autoregressive model (Johnson et al., 2013). The second alternative approach ignores the areal support of the survey data and treats each site as a point location at the site centroid. Spatial dependence in occupancy is included in the second alternative model using a clipped Gaussian process to make it comparable to our continuous-space model. Neither of these alternative approaches accounts for the change of spatial support between the occupancy process and the observed data. Consequently, these alternative approaches are unable to model how the within-site occurrence proportions lead to spatial heterogeneity in detection probabilities. Additional details for these alternative models and their implementation are in Web Appendix C.

We compared the different models by considering the bias of the posterior means and coverage of the 95% credible intervals (CIs) for both the proportion of area occupied and the proportion of sites occupied (see Web Appendix C for details on calculating these quantities). Both of the alternative models were biased and had minimal coverage for the proportion of area occupied and the proportion of sites occupied (Table 1). Our continuous-space occupancy model was unbiased and had high coverage for the proportion of area occupied. Our approach also provided unbiased inferences for the proportion of sites occupied and had high coverage (Table 1). Both of the alternative approaches were positively biased for the proportion of area occupied because they fail to account for the change of spatial support. However, even considering the proportion of sites occupied resulted in biased inferences from these models—due to unaccounted for heterogeneity in detection probabilities resulting from the species only occurring in a portion of a site.

6 AVIAN DATA APPLICATION

We analyzed detection/non-detection data for ovenbirds (*Seiurus aurocapilla*) collected during the summer of 2015 in the Hubbard Brook Experimental Forest, New Hampshire, USA. These data are part of ongoing bird surveys within the experimental forest (Rodenhouse and Sillet, 2019) and are available in the `spOccupancy` R package (Doser et al., 2022). These data include two visit-level predictor variables—time of day and survey date—to model detection probabilities. We obtained elevation data for the study region using the `elevatr` R package (Hollister et al., 2023) to use as a spatial predictor variable for occupancy. All predictor variables were standardized to have mean 0 and standard deviation of 1. The code we used to fit our model to these data is available in the online Supplementary Material.

The Hubbard Brook Experimental Forest is located in a valley in the White Mountains (Figure 3A). The detection/non-detection data were obtained from 10 min point count surveys of circular sites with 100 meter radii (Figure 3 B). Most sites were

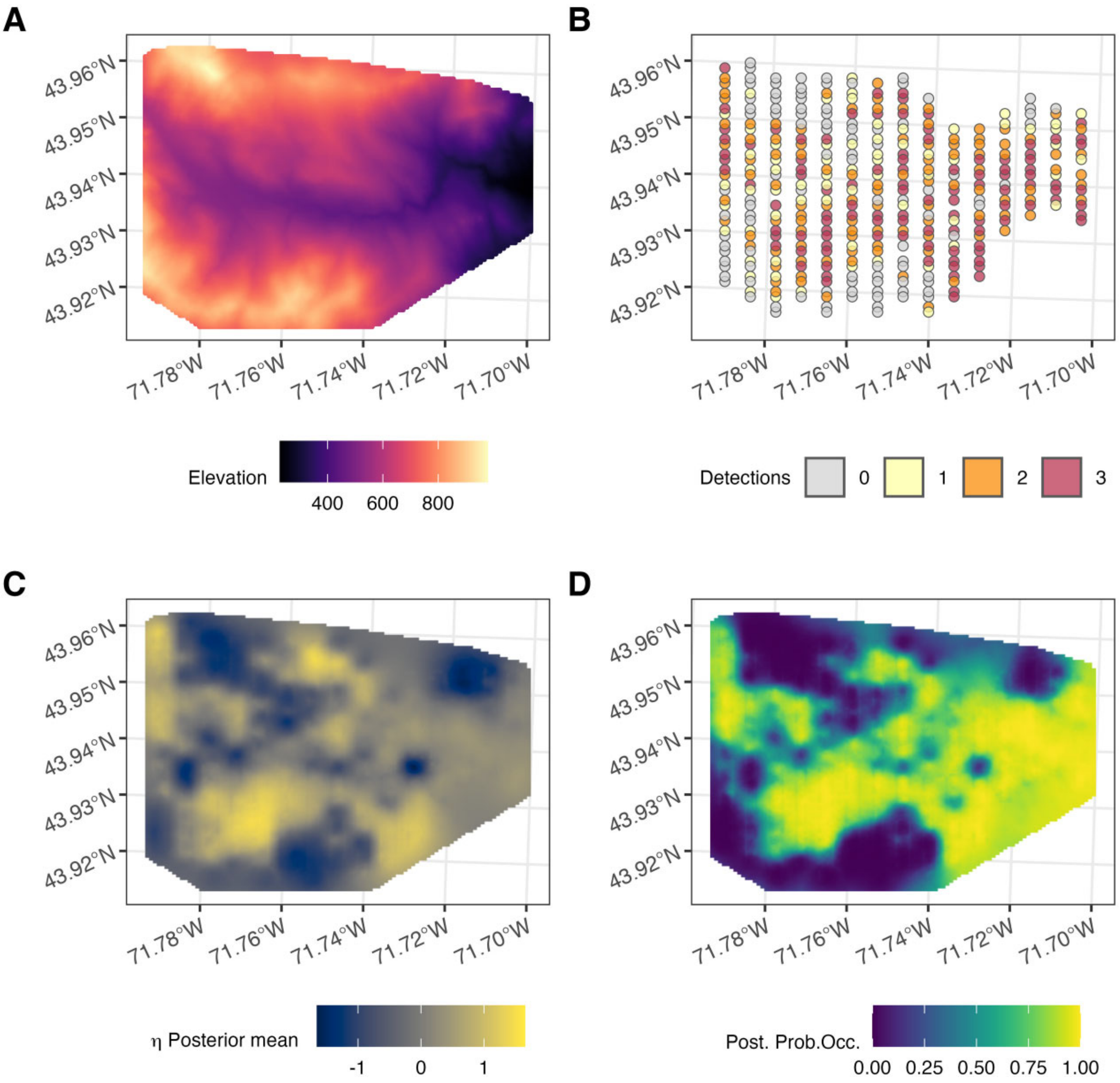

FIGURE 3 Elevation (m) in the Hubbard Brook Experimental Forest (A) and the observed detection/non-detection data for ovenbirds (*Seiurus aurocapilla*) at 100 m radius plots (B). Most sites had 3 visits, but some had only 1 or 2 total visits. After fitting our continuous-space occupancy model, the posterior mean of the spatial effects (C) and the posterior probability of ovenbird occupancy (D) within the study region.

visited 3 times, but a few sites only had 1 or 2 visits. Note that some of the site areas overlap (Figure 3B) and that the sites do not form a regular grid over the study region. Standard occupancy models cannot account for the overlapping sites in these surveys. Typically, this feature of the data would need to be ignored (by treating sites as point locations) or by excluding data from some of the sites that overlap. Our continuous space occupancy model, on the other hand, is able to include all the observed data and naturally accommodate the overlapping sites when making inferences about species occurrence.

We fit our model to these data using three chains with 400 000 MCMC iterations each. The chains were fit in parallel, and on a high performance desktop computer our model took approximately 6 h to run. The iterations from the first half of each chain were discarded as burn-in after using them to tune the surrogate data variances, and the final 200 000 iterations from each chain were saved for inferences. We thinned the iterations by 10 to reduce the amount of memory required to save the results. In this example, we assumed an exponential spatial covariance function so that $K(\mathbf{s}_i, \mathbf{s}_{i'}) = \exp\{-(||\mathbf{s}_i - \mathbf{s}_{i'}||)/\rho\}$ where ρ is the spatial range parameter. Note that we fixed the variance of this covariance function to be 1 for identifiability of the clipped Gaussian process (De Oliveira, 2000, 2020). We found that using a Gaussian spatial covariance function resulted in similar inferences. This is expected because the smoothness parameter is not identifiable for clipped Gaussian processes (De Oliveira, 2000). We ultimately used the exponential spatial covariance function because it is more convenient computationally. We checked model convergence using traceplots for all parameters and summarized posterior inferences using posterior means and 95% CIs.

There was no evidence that detection probabilities varied with the predictor variables date (α_1: posterior mean $= -0.03$ and 95% CI $= (-0.12, 0.06)$) or time of day (α_2: posterior mean $= -0.03$ and 95% CI $= (-0.13, 0.06)$). There was strong evidence that detection probabilities increased as the proportion of the site that was occupied increased (γ: posterior mean $= 1.74$, 95% CI $= (1.19, 2.43)$). Consequently, the posterior mean detection probability at sites with 15% occurrence was 0.29 (95% CI from 0.11 to 0.45), while that for sites with 85% occurrence was 0.73 (95% CI from 0.66 to 0.81). This suggests there is substantial heterogeneity in the detection probabilities due to the variability of within-site species occurrence.

Our analysis also provided strong evidence that ovenbird occurrence was negatively related to elevation (β_1: posterior mean $= -0.68$, 95% CI $= (-1.00, -0.39)$). We obtained posterior inferences on the probability of ovenbird occurrence across the study region, including for areas that were not surveyed. There was evidence of additional spatial variability that was not due to elevation (Figure 3C). We also were able to map the distribution of the species in this region based on the posterior probability of ovenbird occupancy (Figure 3D).

7 DISCUSSION

We developed a new spatial occupancy model for wildlife species. While standard surveys for wildlife species collect detection/non-detection data over areal survey sites, species occupancy is a process that can be defined in continuous space (Efford and Dawson, 2012). Our approach is the first to treat species occurrence in continuous space while accounting for the change of support required for analyzing the areal survey data. Additionally, our model accounts for imperfect detection and allows for heterogeneity in detection probabilities related to the proportion of the site that is occupied. This detection process is similar to that in other occupancy models that incorporate a detection-abundance relationship (Royle and Nichols, 2003) or heterogeneity in detection probabilities using a mixture model (Royle, 2006). Our real data analysis provided strong evidence that detection probabilities of ovenbirds increased as the proportion of the site area that is occupied increased. Failing to account for this variability in detection probabilities can lead to biased inferences even when considering the proportion of sites occupied, as shown in our simulation study. Other sources of variability in detection probabilities, such as the radius used for point counts, can be incorporated in our model as well.

Standard approaches rely on a discrete approximation of species occupancy in continuous space. This approximation can be improved as the size of the sites decreases. In other words, standard occupancy models would have small bias for the proportion of area occupied when the sites are small relative to the scale of spatial variation in species occupancy. However, surveying sufficiently small sites to adequately reduce bias may be challenging because detection probabilities decrease as site area decreases (MacKenzie et al., 2018). Studies commonly define sites so that each has an area that is similar to that of the home range of an individual animal. This leads to substantial variability in within-site species occurrence and, consequently, inferences could be improved by considering our continuous-space model.

The biggest limitation of our model is the increased computational burden compared to standard spatial occupancy models. In our simulation study we used $D = 36300$ spatial grid points to implement our model, and run times were approximately 3 h. The alternative spatial occupancy models are implemented at 300 spatial sites and can be fit in only a few minutes for these simulated data. Using a nearest neighbor approximation increased the computational efficiency of our spatial model and made Bayesian inferences using MCMC feasible. Additionally, we modified the surrogate data slice sampler developed by Murray and Adams (2010) to better accommodate NNGPs. Our version of the sampler introduces surrogate data for only a portion of the spatial locations and is able harness the computational advantages provided by NNGPs for the remaining spatial locations. While our approach was motivated by our continuous-space occupancy model, the surrogate data slice sampler is applicable to any model that includes a latent spatial Gaussian process. This is because all steps update parameters from multivariate normal distributions, resulting from the Gaussian process prior, except when updating the covariance parameters using slice sampling. The slice sampling step is general and can be applied to any assumed data likelihood. Overall, the surrogate data slice sampler provides an efficient approach for jointly updating spatial terms and spatial covariance parameters when using a latent NNGP.

We defined the spatial terms in (3) using a generic covariance function. In some applications, it will be useful to assume occupancy is a multiscale process and model the spatial terms as

$$\eta(\mathbf{s}) = \sum_{m=1}^{M} \eta_m(\mathbf{s}), \tag{22}$$

$$\eta_m(\mathbf{s}) \sim \text{GP}(0, K_{\eta_m}), \tag{23}$$

where the η_m for $m = 1, \ldots, M$ are assumed to be independent of one another and their corresponding spatial covariance functions K_{η_m} have different parameters, including different range parameters. The independence assumption implies that the overall covariance function K_η is equal to $\sum_{m=1}^{M} K_{\eta_m}$. This allows the spatial dependence in the species occupancy process to vary across different spatial scales. Such a multiscale model in continuous space differs from current multiscale occupancy models (eg, Nichols et al., 2008), and future research could compare inferences from these different approaches.

Other variations to standard occupancy models exist that account for multiple species, multiple seasons, and false positive detections (Bailey et al., 2014). These ideas could also be incorporated into our continuous-space occupancy model. For instance, it would be straightforward to construct a model for multiple seasons by assuming each season is a discrete-time snapshot of species occupancy in continuous space. An alternative would be to consider using a clipped Gaussian process for modeling species occupancy in continuous space-time. Care would be needed to ensure detection probabilities are identifiable given the available survey data. This could require concurrent visits to a site or visits close together in time relative to the effective range of temporal covariance function (analogous to the closure assumptions of standard models). Modeling occupancy in continuous space-time could allow for improved insights into how species occurrence changes over time.

ACKNOWLEDGMENTS

We thank the associate editor and an anonymous reviewer for their helpful comments.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices and code referenced in Sections 4-6 are available at the *Biometrics* website on Oxford Academic.

FUNDING

This research was supported by U.S. National Park Service grants G20AC00322 and P20AC00670 and U.S. National Science Foundation grants 1927177 and 2222525.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The ovenbird data analyzed in this paper are available in the `spOccupancy` R package (Doser et al., 2022).

REFERENCES

Bailey, L. L., MacKenzie, D. I. and Nichols, J. D. (2014). Advances and applications of occupancy models. *Methods in Ecology and Evolution*, 5, 1269–1279.

Bradley, J. R., Wikle, C. K. and Holan, S. H. (2016). Bayesian spatial change of support for count-valued survey data with application to the American Community Survey. *Journal of the American Statistical Association*, 111, 472–487.

Broms, K. M., Johnson, D. S., Altwegg, R. and Conquest, L. L. (2014). Spatial occupancy models applied to atlas data show Southern Ground Hornbills strongly depend on protected areas. *Ecological Applications*, 24, 363–374.

Chambert, T., Miller, D. A. and Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. *Ecology*, 96, 332–339.

Cressie, N. A. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, 3, 159–180.

Datta, A. (2022). Nearest-neighbor sparse Cholesky matrices in spatial statistics. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14, e1574.

Datta, A., Banerjee, S., Finley, A. O. and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111, 800–812.

De Oliveira, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Computational Statistics & Data Analysis*, 34, 299–314.

De Oliveira, V. (2020). Models for geostatistical binary data: Properties and connections. *The American Statistician*, 74, 72–79.

Doser, J. W., Finley, A. O. and Banerjee, S. (2023). Joint species distribution models with imperfect detection for high-dimensional spatial data. *Ecology*, 104, e4137.

Doser, J. W., Finley, A. O., Kéry, M. and Zipkin, E. F. (2022). spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. *Methods in Ecology and Evolution*, 13, 1670–1678.

Eddelbuettel, D. and Balamuta, J. J. (2018). Extending R with C++: A brief introduction to Rcpp. *The American Statistician*, 72, 28–36.

Eddelbuettel, D. and Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics & Data Analysis*, 71, 1054–1063.

Efford, M. G. and Dawson, D. K. (2012). Occupancy in continuous habitat. *Ecosphere*, 3, 1–15.

Finley, A. O., Datta, A., Cook, B. D., Morton, D. C., Andersen, H. E. and Banerjee, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. *Journal of Computational and Graphical Statistics*, 28, 401–414.

Gelfand, A. E. and Shirota, S. (2019). Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. *Ecological Monographs*, 89, e01372.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

Gelfand, A. E., Zhu, L. and Carlin, B. P. (2001). On the change of support problem for spatio-temporal data. *Biostatistics*, 2, 31–45.

Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97, 632–648.

Guinness, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics*, 60, 415–429.

Hefley, T. J. and Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports*, 1, 87–97.

Hoeting, J. A., Leecaster, M. and Bowden, D. (2000). An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics*, 5, 102–114.

Hollister, J., Shah, T., Nowosad, J., Robitaille, A. L., Beck, M. W. and Johnson, M. (2023). *Elevatr: Access Elevation Data from Various APIs*. R package version 0.99.0. [Accessed 15 October 2024]

Hooten, M. B., Larsen, D. R. and Wikle, C. K. (2003). Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, 18, 487–502.

Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C. and Pond, B. A. (2013). Spatial occupancy models for large data sets. *Ecology*, 94, 801–808.

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M. and Stone, L. (2017). Integrated species distribution models: Combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution*, 8, 420–430.

Latimer, A. M., Wu, S., Gelfand, A. E. and Silander, J. A. (2006). Building statistical models to analyze species distributions. *Ecological Applications*, 16, 33–50.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A. and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. and Hines, J. E. (2018). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. San Diego, CA: Elsevier.

Murray, I., Adams, R. and MacKay, D. (2010). Elliptical slice sampling. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Vol. 9, 541–548, PMLR.

Murray, I. and Adams, R. P. (2010). Slice sampling covariance hyperparameters of latent Gaussian models. In: *Advances in Neural Information Processing Systems*(eds. Lafferty, J, Williams, C, Shawe-Taylor, J, Zemel, R and Culotta, A), Vol. 23. 1732–1740. San Diego, CA: NeurIPS.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31, 705–767.

Nichols, J. D., Bailey, L. L., O'Connell, A. F., Talancy, N. W., Campbell Grant, E. H., Gilbert, A. T. et al. (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, 45, 1321–1329.

Noon, B. R., Bailey, L. L., Sisk, T. D. and McKelvey, K. S. (2012). Efficient species-level monitoring at the landscape scale. *Conservation Biology*, 26, 432–441.

Ovaskainen, O., Roy, D. B., Fox, R. and Anderson, B. J. (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, 7, 428–436.

Pacifici, K., Reich, B. J., Miller, D. A. W. and Pease, B. S. (2019). Resolving misaligned spatial data with integrated species distribution models. *Ecology*, 100, e02709.

R Core Team, (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rodenhouse, N. L. and Sillet, S. (2019). Valleywide Bird Survey, Hubbard Brook Experimental Forest, 1999-2016 (ongoing) ver 3. Environmental Data Initiative. https://doi.org/10.6073/pasta/faca2b2cf2db9d415c39b695cc7fc217.

Royle, J. A. (2006). Site occupancy models with heterogeneous detection probabilities. *Biometrics*, 62, 97–102.

Royle, J. A. and Nichols, J. D. (2003). Estimating abundance from repeated presence–absence data or point counts. *Ecology*, 84, 777–790.

Ruiz-Gutierrez, V., Hooten, M. B. and Campbell Grant, E. H. (2016). Uncertainty in biological monitoring: a framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, 7, 900–909.

Saha, A. and Datta, A. (2018). BRISC: Bootstrap for rapid inference on spatial covariances. *Stat*, 7, e184.

Shi, H., Kang, E. L., Konomi, B. A., Vemaganti, K. and Madireddy, S. (2017). Uncertainty Quantification Using the Nearest Neighbor Gaussian Process. In Chen, D.-G., Jin, Z., Li, G., Li, Y., Liu, A. and Zhao, Y., editors, *New Advances in Statistics and Data Science*, pp. 89–107. Cham: Springer.

Stein, M. L., Chi, Z. and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66, 275–296.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 50, 297–312.

Walker, N. B., Hefley, T. J., Ballmann, A. E., Russell, R. E. and Walsh, D. P. (2021). Recovering individual-level spatial inference from aggregated binary data. *Spatial Statistics*, 44, 100514.

Walker, N. B., Hefley, T. J. and Walsh, D. P. (2020). Bias correction of bounded location error in binary data. *Biometrics*, 76, 530–539.

Wright, W. J., Irvine, K. M., Rodhouse, T. J. and Litt, A. R. (2021). Spatial Gaussian processes improve multi-species occupancy models when range boundaries are uncertain and nonoverlapping. *Ecology and Evolution*, 11, 8516–8527.

Received: April 15, 2024; Revised: November 5, 2024; Accepted: April 24, 2025