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

OXFORD

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

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

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

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

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