Received: 13 August 2021 | Accepted: 11 November 2021

DOI: 10.1111/2041-210X.13777

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

APPLICATION

ubms: An R package for fitting hierarchical occupancy and N-mixture abundance models in a Bayesian framework

Kenneth F. Kellner[1] | Nicholas L. Fowler[1,2] | Tyler R. Petroelje[1] | Todd M. Kautz[1] | Dean E. Beyer Jr.[3] | Jerrold L. Belant[1]

[1]Global Wildlife Conservation Center, State University of New York College of Environmental Science and Forestry, Syracuse, NY, USA

[2]Alaska Department of Fish and Game, Division of Wildlife Conservation, Soldotna, AK, USA

[3]Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA

Correspondence
Kenneth F. Kellner
Email: kfkellner@esf.edu

Funding information
Federal Aid in Restoration Act under Pittman-Robertson, Grant/Award Number: W-1476-R; Michigan Department of Natural Resources; Camp Fire Conservation Fund; CI Michigan Involvement Committee; Safari Club International Foundation

Handling Editor: Aldo Compagnoni

Abstract

1. Obtaining unbiased estimates of wildlife distribution and abundance is an important objective in research and management. Occupancy and N-mixture abundance models, which correct for imperfect detection, are commonly used for this purpose. Fitting these models in a Bayesian framework has advantages but doing so can be challenging and time-consuming for many researchers.
2. We developed an R package, `ubms`, which provides an easy-to-use, formula-based interface for fitting occupancy, N-mixture abundance and other models in a Bayesian framework using Stan. The package also provides tools for visualizing parameter effects, calculating residuals, assessing goodness-of-fit and comparing models.
3. We demonstrate the use of `ubms` by fitting an N-mixture model to ruffed grouse *Bonasa umbellus* count data from drumming surveys conducted at roadside points sampled on five occasions annually during 2013–2015. To demonstrate the functionality of `ubms`, we used survey site as a random effect, and occasion date and per cent aspen cover at each site as covariates of detection and abundance respectively. The top-ranked model included a positive effect of per cent aspen on grouse abundance.
4. `ubms` has the potential to greatly increase the range of users who will be able to rigorously assess species distribution and abundance while correcting for imperfect detection in a Bayesian framework.

KEYWORDS
Bayesian methods, modelling, population ecology, statistics

1 | INTRODUCTION

Reliable assessments of species distribution and abundance are important goals of a wide range of fields including wildlife research and management. Imperfect detection of individual animals and plants is a ubiquitous source of bias in these assessments (Kellner & Swihart, 2014). To correct the resulting errors, ecologists and statisticians have proposed a suite of hierarchical models which separate the state process (e.g. occurrence or abundance) and the observation, or detection, process (Kéry & Royle, 2016, 2021; Royle & Dorazio, 2008). These models typically require field protocols where a site is surveyed repeatedly, allowing the estimation of probability of detection. Occupancy modelling (MacKenzie et al., 2002; Tyre et al., 2003) is among the most widely used of these hierarchical approaches. In a typical survey protocol for occupancy modelling, M sites are surveyed J times for a species of interest. If the species is detected at C of the M sites, then a simplistic estimate of occurrence probability is obtained as C/M. However, this estimate ignores the possibility that some of the $M - C$ sites without detections were in fact occupied, but the species was simply overlooked. Occupancy

models correct for this error by separately estimating occupancy probability ψ and detection probability p during a given survey. The underlying latent (unobserved) occupancy state at site i, Z_i, is modelled as

$$Z_i \sim \text{Bernoulli}\left(\psi_i\right)$$

and the detections during each of the J surveys at site i, y_{ij}, ..., y_{iJ}, are modelled as conditional on Z_i:

$$y_{ij} \sim \text{Bernoulli}\left(Z_i p_{ij}\right).$$

Usually, the logit-transformed values of ψ and p will be expressed as a function of covariates of interest, allowing researchers to determine, for example, if occupancy probability of a species is affected by habitat type or elevation. Thus, an occupancy model can be seen simply as a species distribution model with an explicit measurement error model attached. Similar hierarchical modelling approaches can be used to estimate abundance from presence–absence data (Royle & Nichols, 2003), or abundance from repeated counts (Royle, 2004a) or time-to-event data (Strebel et al., 2021), to name but a few of many options.

Multiple software tools have been developed to facilitate fitting occupancy and abundance models and estimating covariate effects. Among the pioneers were program `PRESENCE` (Hines, 2006) and program `MARK` (White & Cooch, 2001), stand-alone software that can be used to fit various model types using maximum likelihood. As adoption of the R programming language has greatly increased among ecologists and wildlife biologists, specialized R packages including `RMark` (an interface to `MARK`; Laake, 2013) and `unmarked` (which also uses maximum likelihood; Fiske & Chandler, 2011) have become available. Finally, as Bayesian approaches have become increasingly accessible and popular, researchers have begun to fit Bayesian occupancy and abundance models with Markov chain Monte Carlo (MCMC) tools like JAGS (Plummer, 2003), NIMBLE (de Valpine et al., 2017) and Stan (Carpenter et al., 2017). Bayesian approaches are particularly well-suited to this application, since they allow researchers to obtain posterior distributions of latent occupancy and abundance parameters, which can then be used to calculate point estimates and uncertainty intervals for these parameters (Kéry & Royle, 2016, 2021; Royle & Dorazio, 2008).

Maximum likelihood and Bayesian approaches to fitting occupancy and abundance models each provide advantages and disadvantages. The available tools based on maximum likelihood are generally fast and user-friendly, either via use of a GUI (e.g. `MARK`) or a formula-based interface which is familiar to `R` users (e.g. `unmarked`). Maximum likelihood-based tools also facilitate model comparison with AIC. Conversely, Bayesian approaches require users to code custom models using special programming languages and typically are slower to run, but have advantages such as easier error propagation, ability to incorporate prior information and more intuitive and flexible handling of latent variables such as random effects (Ellison, 2004; Kéry & Royle, 2016). Model selection can also be more challenging in a Bayesian framework (Hooten & Hobbs, 2015). We developed an R package, `ubms`, which combines advantages of both approaches. `ubms` uses a formula-based interface familiar to R and `unmarked` users, while fitting occupancy and abundance models in a Bayesian framework using Stan (Carpenter et al., 2017) as the computational engine. `ubms` allows users to easily include random effects in occupancy, abundance and detection models. The estimation of random effects may be useful in a variety of scenarios; for example, to estimate region-level effects when sites are nested in regions, or to allow detection probability to vary among different observers. Finally, `ubms` facilitates model comparison via calculation of WAIC or leave-one-out cross-validation (LOO cross-validation; Vehtari et al., 2017), and model checking via goodness-of-fit tests. We believe this package will be a valuable tool for researchers interested in rigorous assessments of the distribution and abundance of animal and plant populations.

2 | DESCRIPTION OF `ubms`

2.1 | Fitting a model

`ubms` includes several core functions (hereafter called 'fitting functions'), each of which is designed to fit a particular type of occupancy or abundance model in Stan. These functions were designed to be drop-in replacements for corresponding functions in the popular package `unmarked`, and thus have similar names. For example, in `unmarked`, single-season occupancy models can be fit with function `occu`, while the equivalent in `ubms` is `stan_occu`. Use of the `stan_` prefix follows the naming conventions of package `rstanarm`, which can be used as an easy alternative to other R modelling packages and functions (Goodrich et al., 2020). At present, seven different fitting functions are available in `ubms` (Table 1). In addition to a similar naming convention, required arguments and inputs provided to fitting functions in `ubms` also match those in `unmarked` wherever possible. This allows users to quickly switch between and compare fitting models with the two packages, and should facilitate adoption by existing `unmarked` users.

After selecting an appropriate fitting function, `ubms` users must provide two key inputs: a set of formulas and a dataset. There is a right-sided formula for each model parameter (e.g. occupancy probability), and the formulas are specified using the typical syntax used in many common R functions. For example, to specify that occupancy probability should be a function of elevation, the corresponding right-sided formula would be `~elevation`. Random slopes and intercepts may be specified using the syntax popularized by the `lme4` R package (Bates et al., 2015). For example, to add random intercepts by region, the corresponding formula would be `~elevation + (1|region)`. It is possible to include multiple random effects (e.g. region and year) to a given model, but some more complicated random effect structures such as correlated random slopes and intercepts and nested random effects are not yet supported.

TABLE 1 Models currently available in `ubms` and their corresponding fitting functions. A complete table of functions available in `ubms` can be found at https://kenkellner.com/ubms/reference/index.html

Model	Fitting function	Citation
Single-season occupancy	`stan_occu`	MacKenzie et al. (2002) and Tyre et al. (2003)
Dynamic occupancy	`stan_colext`	MacKenzie et al. (2003)
Time-to-detection occupancy	`stan_occuTTD`	Garrard et al. (2008)
Royle–Nichols abundance	`stan_occuRN`	Royle and Nichols (2003)
N-mixture abundance	`stan_pcount`	Royle (2004a)
Hierarchical distance sampling	`stan_distsamp`	Royle et al. (2004)
Multinomial-Poisson abundance	`stan_multinomPois`	Royle (2004b)

TABLE 2 Example output summary from an N-mixture model of ruffed grouse *Bonasa umbellus* abundance fit in `ubms` with `stan_pcount` (model ASP). The summary contains point estimates (posterior means) and interval estimates (posterior *SD* and 95% credible intervals) for parameters from the abundance and detection submodels, along with associated MCMC diagnostic information. Note the familiar format of the summary shared with plenty of R model fitting functions including `lm`, `glm` and `lmer` as well as those in package `unmarked`

`stan_pcount(formula = ~scale(date) ~ scale(pctaspen) + (1 | site), data =umf, K = 20, chains =3, iter =2000, seed =123)`

	Estimate	*SD*	2.5%	97.5%	n_eff	Rhat
Abundance (log-scale):						
(Intercept)	0.969	0.191	0.6303	1.381	1890	1
scale(pctaspen)	0.289	0.103	0.0896	0.496	1678	1
sigma [1|site]	0.652	0.105	0.4672	0.870	900	1
Detection (logit-scale):						
(Intercept)	−1.560	0.2154	−2.011	−1.167	1922	1
scale(date)	−0.474	0.0763	−0.626	−0.336	3057	1
LOOIC: 1,613.632						
Runtime: 4.622 min						

Input data must be stored in a special format called an `unmarkedFrame`, as described by the `unmarked` package (Fiske & Chandler, 2011). The `unmarkedFrame` combines the observed occurrence or count data as well as corresponding covariates. Using exactly the same data structure type as `unmarked` further facilitates the ease of `ubms` adoption by `unmarked` users.

Selection of priors is an important part of Bayesian analysis (Banner et al., 2020). `ubms` provides reasonable weakly informative default priors on intercepts, regression coefficients and random effect standard deviations. For probability-type parameters, such as occupancy or detection, intercepts and regression coefficients are given a Logistic(0, 1) prior as suggested by Northrup and Gerber (2018). Intercepts and regression coefficients for abundance parameters receive Normal(0, *SD* = 5) and Normal(0, *SD* = 2.5) priors by default, and random effect standard deviations are assigned a Gamma(1, 1) prior. All prior distributions can be customized by users individually, with additional distributions (including Student's *t*, Cauchy and uniform) available. This facilitates prior sensitivity analysis, an important step in a Bayesian workflow that is often overlooked (Banner et al., 2020).

Using these inputs (along with other potential options, such as the number of MCMC chains to use), the fitting function bundles the data into a format suitable for Stan and passes the data to the appropriate pre-compiled Stan model. Stan's MCMC sampler is then invoked via the `rstan` package (Stan Development Team, 2017a), and the resulting Stan output is wrapped in an object of class `ubmsFit`. Calling this object from the R console returns a summary table of statistics (Table 2).

2.2 | Diagnostics and model fit

The first step after fitting a model is to identify any problems with MCMC sampling. Summary output from `ubms` model fits includes estimates of effective sample size and split-chain $\hat{R}$, which can be used to diagnose the convergence failure of MCMC chains (Vehtari et al., 2021). The `traceplot` function allows users to visually assess chain convergence. The next step should ideally be the assessment of model fit. `ubms` can automatically calculate separate residuals for the state and observation processes following Wright et al. (2019) using the `residuals` function, and these can be plotted against fitted values or covariates with `plot_residuals` (Figure 1a,b). Posterior predictive checks of model fit statistics, such as the MacKenzie–Bailey χ^2 statistic for occupancy models (MacKenzie &

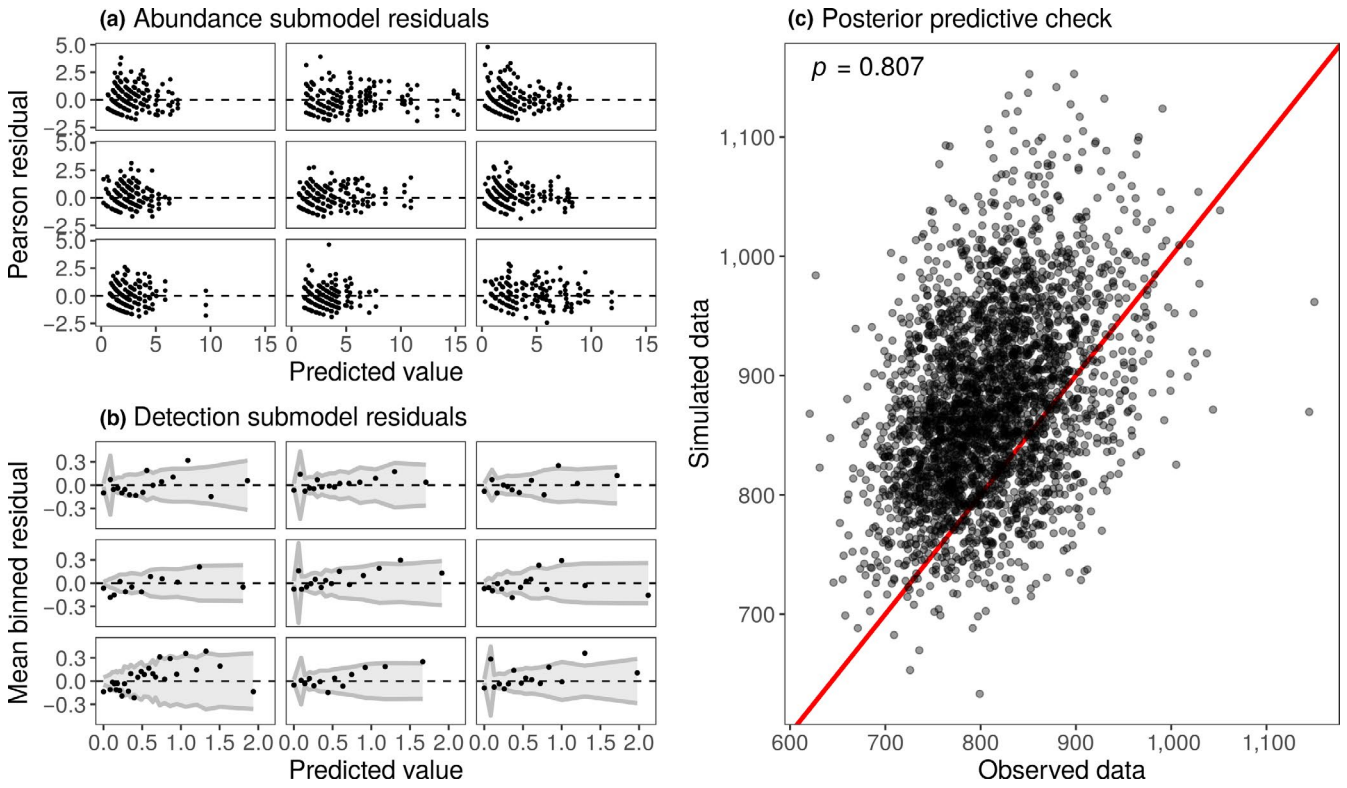

FIGURE 1 Examples of (a and b) residual plots generated with function `plot_residual` and (c) a posterior predictive check plot generated by `gof` from an N-mixture model of ruffed grouse *Bonasa umbellus* abundance fit with `ubms`. Residuals were generated following Wright et al. (2019). The posterior predictive check plot also includes a calculated Bayesian *p*-value ($p = 0.807$) indicating adequate model fit

Bailey, 2004), can be obtained with the `gof` function for most available models (Figure 1c). The draws from the posterior distributions of all model parameters produced by the Stan run are available via the `extract` function, allowing users to calculate their own custom fit statistics as well.

2.3 | Model inference

Multiple models can be compared with WAIC (using the `waic` function), or with LOO cross-validation using the `modSel` function. For an individual model, parameter estimates and 95% credible intervals are found in the summary table for a model as described earlier (Table 2). For a model that includes covariates, the `plot_marginal` function will plot the marginal effects of each covariate (with all other covariates held at their mean observed values) as a quick visual summary of results (Figure 2a). The `predict` function can be used to calculate parameter estimates at various custom combinations of covariate values, for example to generate a predictive map across a study area (Figure 2b). Finally, the `posterior_predict` function can be used to generate posterior distributions of the outcome and of latent variables such as the realized values of abundance and occupancy at the actual sample of studied sites. Additional details on available functions and the `ubms` workflow are available in the package vignettes (available from https://cran.r-project.org/package=ubms).

3 | COMPARISON WITH EXISTING SOFTWARE

3.1 | Advantages of `ubms`

The majority of occupancy and abundance analyses that account for imperfect detection are currently conducted with maximum likelihood-based tools like `unmarked`, `PRESENCE` and `MARK`, or custom model fit using Bayesian tools like JAGS. Relative to existing maximum likelihood approaches, which provide estimates of uncertainty in the form of standard errors or confidence intervals, an advantage of `ubms` is that it provides full posterior distributions for model parameters. This allows for probabilistic statements about parameter values, inference about latent abundance and occupancy parameters (e.g. using function `posterior_predict`), and calculation of posterior distributions for fit statistics (e.g. using the `residuals` or `gof` functions described earlier). Furthermore, `ubms` allows for the estimation of random slopes and intercepts for all model parameters, which is only partially available in tools such as `unmarked`, `MARK` and `PRESENCE`. Inclusion of such random effects is very often required in abundance and occupancy models. For example, if survey sites are nested within regions, a random effect of region may have to be included in the model. Alternatively, if many different observers collected data, an observer random effect on detection may be required to accommodate the resulting dependency in the data. `ubms` allows specification of random

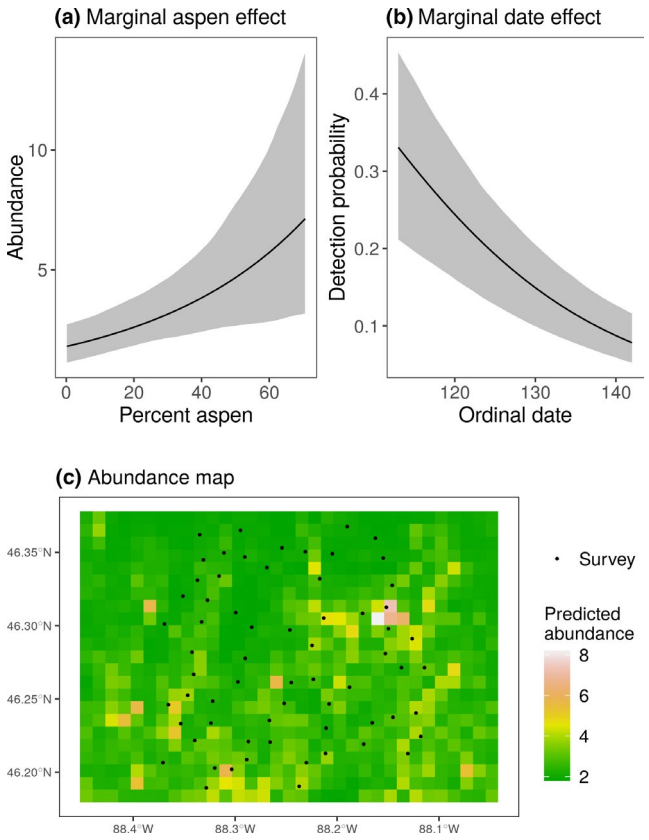

FIGURE 2 Marginal effects plots (posterior means and 95% credible intervals) generated by `plot_marginal` from an N-mixture model of ruffed grouse *Bonasa umbellus* abundance fit with `ubms`, showing (a) the effects of per cent aspen on grouse abundance and (b) ordinal date on grouse detection probability, and (c) a map of model-predicted grouse abundance across the study area generated using the `predict` function

effects by simply changing the formula provided to the fitting function.

The second alternative to `ubms` is fitting custom-written models with one of the several model fitting engines available for Bayesian analysis such as JAGS (Plummer, 2003), NIMBLE (de Valpine et al., 2017) or Stan (Carpenter et al., 2017). An important advantage of `ubms` relative to these tools is the ease of use. Writing custom models requires knowledge of special programming languages and a deep understanding of model structure. While these skills are valuable to learn, limited time and resources mean learning to write custom code for JAGS or Stan may not be feasible for all researchers or managers. `ubms` allows anyone that can use R to fit models in a Bayesian framework and gain the corresponding advantages described above. Even for experienced users of JAGS or Stan, the interface of `ubms` allows for quick iteration and adjustment of models without rewriting complex model code. A second, considerable advantage of `ubms` over software like JAGS and Stan is the availability in `ubms` of a suite of utility functions allowing easy manipulation and visualization of results. Custom models mean that, generally, researchers must also write custom code for making predictions from the fitted model, visualizing output, choosing between competing models and checking goodness-of-fit. In contrast, `ubms` includes a whole series of utility functions like `predict`, `plot_marginal`, `plot_residuals` and `gof`, which can be used on fitted models immediately. Finally, because of Stan's efficient sampler (Hoffman & Gelman, 2014), and because the underlying likelihoods in `ubms` are marginalized, `ubms` may fit models faster than equivalent models in JAGS in some cases (Yackulic et al., 2020).

3.2 | Drawbacks of `ubms`

`ubms` combines advantages of both maximum likelihood model fitting and Bayesian model fitting with JAGS, NIMBLE or Stan which necessitates compromises relative to both. Compared to `unmarked`, `MARK` and `PRESENCE`, fitting models in `ubms` will typically be slower and perhaps markedly so for complex models with large datasets, due to the use of MCMC sampling. Furthermore, the users of `ubms` must have some understanding of how to diagnose and address MCMC problems such as lack of convergence (Kéry & Royle, 2016). To help with this, Stan, and `ubms`, provide tools such as automatic calculation of the $\hat{R}$ statistic (Vehtari et al., 2021), generation of MCMC traceplots and other diagnostics such as those available in the `shinystan` package (Stan Development Team, 2017b). Compared to fitting custom models with JAGS or Stan, `ubms` users are limited to using the supported model types, and cannot further customize model structure.

4 | APPLICATION

We now demonstrate the workflow of a typical N-mixture model analysis using `ubms`, for ruffed grouse *Bonasa umbellus* count data from drumming surveys in Michigan, USA (46.3°N, −88.2°E) conducted during 2013–2015. Code and data used in this analysis, along with an additional example occupancy analysis, are available from Zenodo (Kellner et al., 2021). We established 65 roadside sampling points >1.6 km apart to ensure site independence and assumed grouse had a maximum detection radius of 550 m from each point (Hansen et al., 2011). Each year we visited sampling points on five occasions from late April to early May, a period during which we assumed population closure. We visited sites 0.5 hr before sunrise to 5 hr after sunrise and counted the number of independent detections of drumming grouse during a 5-min period. We extracted aspen (*Populus tremuloides* and *P. grandidentata*) cover types from the National Individual Tree Species Atlas (Ellenwood et al., 2015) within a 550 m buffer of each site to estimate per cent cover given the importance of aspen to grouse (Jakubas & Gullion, 1991; Svoboda & Gullion, 1972).

We fit three candidate N-mixture models (Royle, 2004a) for abundance to the grouse data. The first was a null model with intercepts only for both abundance and detection (NULL). For the second, since we had 3 years of grouse data collected at the same sampling points, we included random intercepts by site for abundance to account for

TABLE 3 Model selection by comparison of predictive accuracy of candidate models for ruffed grouse *Bonasa umbellus* abundance using leave-one-out cross-validation, generated using `ubms` with function `modSel`. Accuracy is measured using the expected log pointwise predictive density (ELPD) relative to the top-ranked model (ΔELPD; Vehtari et al., 2017)

Model	ELPD	ΔELPD	*SE* of ΔELPD
ASP	−806.82	0.00	0.00
RAND	−807.99	−1.17	1.78
NULL	−846.75	−39.94	7.20

site-level heterogeneity (RAND). For the third, we included random site intercepts and a fixed effect of per cent aspen on abundance (ASP). In all three models, we included an effect of ordinal date on detection probability. Covariates were standardized to have a mean of 0 and a standard deviation of 1 before analysis. All models were fit with `stan_pcount` in `ubms` using three MCMC chains of 2,000 iterations each with a burn-in of 1,000 iterations. Candidate models were compared using leave-one-out cross-validation (LOO cross-validation; Vehtari et al., 2017).

The two models which included the random site effect (RAND and ASP) had higher predictive accuracy than the model without the random effect (NULL; Table 3). The top-ranked model ASP included an effect of per cent aspen, but was similar in predictive accuracy to the model without aspen (Table 3). Model fit for ASP was adequate based on residual plots and a posterior predictive check (Bayesian *p*-value = 0.792; Figure 1a). Based on model ASP, aspen had a strong, positive effect on expected grouse abundance (Table 2; Figure 2a). A 1 standard deviation increase in per cent aspen (about 15%) increased expected grouse abundance by 34%. Detection probability strongly declined over the course of the survey season (Table 2; Figure 2a).

Drumming surveys as an index have been criticized as they do not account for heterogeneity introduced by seasonal changes in temperature or breeding activity (Zimmerman & Gutiérrez, 2007). We corroborate the need to account for heterogeneity using ordinal day during the breeding season as a surrogate for decline in breeding activity by males (Figure 2a). In addition, accounting for the unmodelled heterogeneity within sites across years increased predictive accuracy. The importance of aspen to grouse has been long established (Jakubas & Gullion, 1991; Svoboda & Gullion, 1972) and is supported by the inclusion of aspen in the best supported model. The presence of the random site effect in the top model suggests that inclusion of additional land cover classes or fragmentation metrics may further inform the role of landscape complexity or configuration on grouse abundance.

5 | SUMMARY AND OUTLOOK

The `ubms` package allows R users, and especially those familiar with the popular `unmarked` package, to quickly fit occupancy and abundance models in a Bayesian framework, check model fit and visualize results. We see `ubms` as a valuable intermediate option between maximum likelihood tools like `unmarked`, `MARK` and `PRESENCE`, and custom Bayesian model fitting engines including JAGS, Nimble or Stan. The package was not developed to replace either approach for fitting hierarchical models but rather to supplement them, for situations when users need a Bayesian framework and standard model structures are appropriate. Current developmental work in `ubms` focuses on the continued addition of new models (especially from the larger catalogue present in `unmarked`) and features to the package. One priority is the inclusion of methods for accommodating spatial autocorrelation, inclusion of which can greatly increase the quality of species distribution maps (Guélat & Kéry, 2018). At present, `ubms` includes support for spatially autocorrelated random effects using restricted spatial regression (Johnson et al., 2013).

ACKNOWLEDGEMENTS

The authors thank Marc Kéry, two anonymous reviewers and the associate editor for their helpful comments on a previous version of this manuscript. Funding was provided by the Michigan Department of Natural Resources, Federal Aid in Restoration Act under Pittman-Robertson project W-1476-R, Safari Club International (SCI) Foundation, SCI Michigan Involvement Committee and Camp Fire Conservation Fund.

CONFLICT OF INTEREST

The authors declare that there are no conflict of interest.

AUTHORS' CONTRIBUTIONS

K.F.K., N.L.F. and T.R.P. conceived the ideas and methodology; K.F.K. developed the package; K.F.K., N.L.F., T.R.P. and J.L.B. wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13777.

DATA AVAILABILITY STATEMENT

The `ubms` package is available on CRAN (https://cran.r-project.org/package=ubms), Github (https://github.com/kenkellner/ubms) and Zenodo (Kellner, 2021). Data and code used in the example are also available from Zenodo (Kellner et al., 2021).

ORCID

Kenneth F. Kellner https://orcid.org/0000-0002-6755-0555
Nicholas L. Fowler https://orcid.org/0000-0003-3977-0945
Tyler R. Petroelje https://orcid.org/0000-0003-4957-6732

REFERENCES

Banner, K. M., Irvine, K. M., & Rodhouse, T. J. (2020). The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in Ecology and Evolution*, *11*, 882–889. https://doi.org/10.1111/2041-210X.13407

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, *26*, 403–413. https://doi.org/10.1080/10618600.2016.1172487

Ellenwood, J. R., Krist Jr, F. J., & Romero, S. A. (2015). *National individual tree species atlas*. U.S. Department of Agriculture Forest Service Forest Health Technology Enterprise Team, Fort Collins, Colorado, Publication No. FHTET-15-01.

Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, *7*, 509–520. https://doi.org/10.1111/j.1461-0248.2004.00603.x

Fiske, I., & Chandler, R. (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, *43*, 1–23.

Garrard, G. E., Bekessy, S. A., McCarthy, M. A., & Wintle, B. A. (2008). When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology*, *33*, 986–998. https://doi.org/10.1111/j.1442-9993.2008.01869.x

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2020). *rstanarm: Bayesian applied regression modeling via Stan*. R package version 2.21.1. Retrieved from https://cran.r-project.org/web/packages/rstanarm/index.html

Guélat, J., & Kéry, M. (2018). Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, *9*, 1614–1625. https://doi.org/10.1111/2041-210X.12983

Hansen, C. P., Millspaugh, J. J., & Rumble, M. A. (2011). Occupancy modeling of ruffed grouse in the Black Hills National Forest. *Journal of Wildlife Management*, *75*, 71–77. https://doi.org/10.1002/jwmg.21

Hines, J. E. (2006). *PRESENCE 3.1 Software to estimate patch occupancy and related parameters*. Retrieved from http://www.mbr-pwrc.usgs.gov/software/presence.html

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*, 3–28. https://doi.org/10.1890/14-0661.1

Jakubas, W. J., & Gullion, G. W. (1991). Use of quaking aspen flower buds by ruffed grouse: Its relationship to grouse densities and bud chemical composition. *The Condor*, *93*, 473–485. https://doi.org/10.2307/1368179

Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C., & Pond, B. A. (2013). Spatial occupancy models for large data sets. *Ecology*, *94*, 801–808. https://doi.org/10.1890/12-0564.1

Kellner, K. F. (2021). kenkellner/ubms: ubms v1.1.0 (v1.1.0). *Zenodo*, https://doi.org/10.5281/zenodo.5711279

Kellner, K. F., Fowler, N. L., Petroelje, T. R., Kautz, T. M., Beyer Jr., D. E., & Belant, J. L. (2021). Example code and data for ubms: An R package for fitting hierarchical occupancy and N-mixture abundance models in a Bayesian framework (1.0.0). *Zenodo*, https://doi.org/10.5281/zenodo.5711345

Kellner, K. F., & Swihart, R. K. (2014). Accounting for imperfect detection in ecology: A quantitative review. *PLoS ONE*, *9*, e111436. https://doi.org/10.1371/journal.pone.0111436

Kéry, M., & Royle, J. A. (2016). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and static models*. Academic Press.

Kéry, M., & Royle, J. A. (2021). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and Advanced Models*. Academic Press.

Laake, J. L. (2013). *RMark: An R interface for analysis of capture-recapture data with MARK*. R package version 2.2.7. Retrieved from https://cran.r-project.org/web/packages/Rmark/index.html

MacKenzie, D. I., & Bailey, L. L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, *9*, 300–318. https://doi.org/10.1198/108571104X3361

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., & Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, *84*, 2200–2207. https://doi.org/10.1890/02-3090

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*, 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Northrup, J. M., & Gerber, B. D. (2018). A comment on priors for Bayesian occupancy models. *PLoS ONE*, *13*, e0192819. https://doi.org/10.1371/journal.pone.0192819

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, *124*, 1–10.

Royle, J. A. (2004a). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, *60*, 108–115.

Royle, J. A. (2004b). Generalized estimators of avian abundance from count survey data. *Animal Biodiversity and Conservation*, *27*, 375–386.

Royle, J. A., Dawson, D. K., & Bates, S. (2004). Modeling abundance effects in distance sampling. *Ecology*, *85*, 1591–1597. https://doi.org/10.1890/03-3127

Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. Elsevier.

Royle, J. A., & Nichols, J. D. (2003). Estimating abundance from repeated presence–absence data or point counts. *Ecology*, *84*, 777–790. https://doi.org/10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2

Stan Development Team. (2017a). *RStan: The R interface to Stan*, *version 2.21.2*. Stan Development Team. Retrieved from http://mcstan.org

Stan Development Team. (2017b). *shinystan: Interactive visual and numerical diagnostics and posterior analysis for Bayesian models*, *version 2.3.0*. Stan Development Team. Retrieved from http://mc-stan.org

Strebel, N., Fiss, C. J., Kellner, K. F., Larkin, J. L., Kéry, M., & Cohen, J. (2021). Estimating abundance based on time-to-detection data. *Methods in Ecology and Evolution*, *12*, 909–920. https://doi.org/10.1111/2041-210X.13570

Svoboda, F. J., & Gullion, G. W. (1972). Preferential use of aspen by ruffed grouse in northern Minnesota. *The Journal of Wildlife Management*, *36*, 1166–1180. https://doi.org/10.2307/3799245

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, *13*, 1790–1801. https://doi.org/10.1890/02-5078

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2021). Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC. *Bayesian Analysis*, *16*, 667–718.

White, G. C., & Cooch, E. (2001). *Program Mark. Mark and recapture survival rate estimation*. Version 9. Retrieved from http://www.phidot.org/software/mark/downloads/

Wright, W. J., Irvine, K. M., & Higgs, M. D. (2019). Identifying occupancy model inadequacies: Can residuals separately assess detection and presence? *Ecology*, *100*, e02703. https://doi.org/10.1002/ecy.2703

Yackulic, C. B., Dodrill, M., Dzul, M., Sanderlin, J. S., & Reid, J. A. (2020). A need for speed in Bayesian population models: A practical guide to marginalizing and recovering discrete latent states. *Ecological Applications*, *30*, e02112. https://doi.org/10.1002/eap.2112

Zimmerman, G. S., & Gutiérrez, R. J. (2007). The influence of ecological factors on detecting drumming ruffed grouse. *The Journal of Wildlife Management*, *71*, 1765–1772. https://doi.org/10.2193/2006-184

How to cite this article: Kellner, K. F., Fowler, N. L., Petroelje, T. R., Kautz, T. M., Beyer, D. E. Jr., & Belant, J. L. (2022). `ubms`: An R package for fitting hierarchical occupancy and N-mixture abundance models in a Bayesian framework. *Methods in Ecology and Evolution*, 13, 577–584. https://doi.org/10.1111/2041-210X.13777