Received: 21 February 2020 | Accepted: 18 June 2020

DOI: 10.1111/2041-210X.13442

APPLICATION

Methods in Ecology and Evolution
BRITISH ECOLOGICAL SOCIETY

msocc: Fit and analyse computationally efficient multi-scale occupancy models in R

Christian Stratton[1] | **Adam J. Sepulveda**[2] | **Andrew Hoegh**[1]

[1]Department of Mathematical Sciences, Montana State University, Bozeman, MT, USA

[2]Northern Rocky Mountain Science Center, U.S. Geological Survey, Bozeman, MT, USA

Correspondence
Christian Stratton
Email: christianstratton@montana.edu

Funding information
USGS Ecosystems Mission Area

Handling Editor: Robert B. O'Hara

Abstract

1. Environmental DNA (eDNA) sampling is a promising tool for the detection of rare and cryptic taxa, such as aquatic pathogens, parasites and invasive species. Environmental DNA sampling workflows commonly rely on multi-stage hierarchical sampling designs that induce complicated dependencies within the data. This complex dependence structure can be intuitively modelled with Bayesian multi-scale occupancy models. However, current software for such models are computationally demanding, impeding their use.
2. We present an R package, msocc, that implements a data augmentation strategy to fit fully Bayesian, computationally efficient multi-scale occupancy models. The msocc package allows users to fit multi-scale occupancy models, to estimate and visualize posterior summaries of site, sample and replicate-level occupancy, and to compare different models using Bayesian information criterion. Additionally, we provide a supplemental web application that allows users to investigate study design for multi-scale occupancy models and acts as a graphical user interface to the msocc package.
3. The utility of the msocc package is illustrated on a published dataset and the functions in msocc are compared to the primary Bayesian toolkit for multi-scale occupancy modelling, eDNAoccupancy, using various computational benchmarks. These benchmarks indicate that msocc is capable of fitting models 50 times faster than eDNAoccupancy.
4. We hope that access to software that efficiently fits, analyses and conducts study design investigations for multi-scale occupancy models facilitates their implementation by the research and wildlife management communities.

KEYWORDS

Bayesian estimation, detection, environmental DNA, hierarchical modelling, multi-scale occupancy models, occupancy surveys, study design

1 | INTRODUCTION

Environmental DNA (eDNA) surveys continue to gain popularity for detecting invasive, cryptic and rare species (Klymus, Richter, Chapman, & Paukert, 2015; Lodge et al., 2012; Schmelzle & Kinziger, 2016; Sepulveda et al., 2019), as these techniques are often easier, less expensive and more sensitive than non-molecular detection tools (Eiler, Löfgren, Hjerne, Nordén, & Saetre, 2018; Hunter, Meigs-Friend, Ferrante, Smith, & Hart, 2019; Sengupta et al., 2019; Sepulveda et al., 2019; Williams, Huyvaert, Vercauteren, Davis, & Piaggio, 2018). These surveys rely on hierarchical sampling techniques to accommodate spatial heterogeneity in the occurrence of DNA within a study region. This hierarchical design is necessary, as the presence of a species within a site does not imply that its DNA will be present in every sample taken from that site (Dorazio & Erickson, 2018). Therefore, multiple samples are typically taken from each site. These samples are then assessed for the presence of DNA using polymerase chain reaction (PCR) chemistry on multiple replicates from each sample. However, even if DNA is present in the sample, it may not be present in all PCR replicates (Dorazio & Erickson, 2018).

This sampling design induces a hierarchical dependence structure in the occupancy state of the target species at the site, sample and replicate levels that must be appropriately modelled. Following Nichols et al. (2008), Mordecai, Mattsson, Tzilkowski, and Cooper (2011) and Dorazio and Erickson (2018) proposed a hierarchical model of latent state variables to handle this dependence structure. The resulting multi-scale occupancy model provides a flexible platform for modelling multiple levels of uncertainty in the detection of the target species by accounting for false negatives at the site, sample and replicate levels while simultaneously respecting the hierarchical dependence.

Dorazio and Erickson (2018) provide an R package, eDNAoccupancy, for fitting their model from a Bayesian perspective. Bayesian techniques provide a natural way to quantify the uncertainty in the estimated occupancy parameters, and are valuable tools for hierarchical modelling of ecological data (Dorazio, 2015). While Bayesian hierarchical techniques are valuable for understanding complex data structures, fitting hierarchical models can be computationally demanding. Fitting such models for binary data, with a logistic link, has historically required the use of a Markov chain Monte Carlo (MCMC) technique known as Metropolis–Hastings; this sampling technique is implemented in eDNAoccupancy. This technique can be slow to converge to the target posterior distribution and also requires time-consuming tuning (Robert, 2015). Additionally, the algorithm is not well-suited to estimate parameters on vastly different scales (Roberts & Rosenthal, 2001), which often forces users to centre and scale their continuous predictors, resulting in cumbersome interpretations. Due to these limitations, software packages that rely on this technique can be time-consuming to use, especially when fitting multiple models with the intent of model comparison. The msocc package provides an alternative to these packages, implementing a Gibbs sampler that quickly converges to the posterior distribution and features a web application capable of investigating study design, fitting models and analysing the results.

The msocc package implements the data augmentation strategy described by Polson, Scott, and Windle (2013) to fit the hierarchical model described by Dorazio and Erickson (2018) using a Gibbs sampler; this procedure requires no tuning, easily handles covariates on differing scales and quickly converges to the posterior distribution. The remainder of this article is organized as follows: Section 2 contains a description of our package's features, including an example analysis on published data. Section 3 describes the supplemental web application developed to conduct precision analyses and act as a graphical user interface to the msocc package. Section 4 illustrates the computational advantage of the msocc package over eDNAoccupancy by comparing computational benchmarks.

2 | PACKAGE OVERVIEW

The main function in the msocc package is msocc_mod(), which fits the multi-scale occupancy model described by Dorazio and Erickson (2018); for complete model notation, readers are referred to Dorazio and Erickson (2018). The syntax of msocc_mod() is as follows:

```
> msocc_mod(wide_data, num.mcmc = 1000,
+         site = list(model = ~1, cov_tbl),
+         sample = list(model = ~1, cov_tbl),
+         rep = list(model = ~1, cov_tbl),
+         priors = list(site = list(mu0 = 0, Sigma0 = 9),
+                       sample = list(mu0 = 0, Sigma0 = 9),
+                       rep = list(mu0 = 0, Sigma0 = 9),
+                       a0 = 1, b0 = 1),
+         progress = T, print = NULL, seed = NULL, beta_bin = T)
```

The site, sample and rep arguments each take lists containing items named model and cov_tbl respectively. The model item is used to specify the formula which determines the model frame for each level of the hierarchy; the cov_tbl item is used to specify the data frame containing the covariates used in model at each level. This design allows users to specify models using familiar lm and glm syntax, resulting in a function that is widely accessible.

The remaining arguments of msocc_mod() are the number of iterations in the Gibbs sampler (num.mcmc), the detection data (wide_data), whether progress should be printed (progress), how often progress should be printed (print), whether a faster beta-binomial sampler should be used when applicable (beta_bin), specification of priors (priors) and an optional seed to set for reproducibility (seed). The msocc_mod() function returns a list (of class msocc) with the following elements:

- beta = matrix of posterior samples of the regression coefficients at the site level
- psi = vector of posterior samples of the site-level occupancy probability parameter; only returned if beta_bin = TRUE and site$model = ~1
- alpha = matrix of posterior samples of the regression coefficients at the sample level
- theta = matrix of posterior samples of the sample level occupancy probability parameter; only returned if beta_bin = TRUE and sample$model = ~1 or sample$model = ~site
- delta = matrix of posterior samples of the regression coefficients at the replicate level
- p = vector of posterior samples of the replicate level occupancy probability parameter; only returned if beta_bin = TRUE and rep$model = ~1

- model.info = list of model information, including the design matrices for each level of the hierarchy, the number of sites, the number of samples per site and the number of replicates per sample.

2.1 | Posterior summary tools

The msocc package is equipped with multiple functions to summarize the joint posterior distribution of the derived probability parameters. The posterior_summary() function provides a numerical summary table of the derived probability parameters at each level of the hierarchy. The syntax of posterior_summary() is as follows:

```
> posterior_summary(msocc_mod, burnin = 0, level = "overall",
+       quantiles = c(0.025, 0.975), unique = T)
```

The arguments of this function include a fitted model of class msocc (msocc_mod), the number of samples to discard as warm-up (burnin), the level of the model to summarize which may be one of ‘overall’, ‘site’. ‘sample’, or ‘rep’ (level), quantiles defining the credibility intervals to be provided (quantiles) and whether only unique rows of the summary table should be printed (unique).

The cred_plot() function provides a graphical summary of credibility intervals for each of the derived probability parameters at each level of the hierarchy. The syntax of cred_plot is as follows:

```
> cred_plot(msocc, level = "site", truth = NULL,
+       n = "all", quantiles = c(0.025, 0.975), burnin = 0)
```

The arguments of this function include a fitted model of class msocc (msocc), the level of the model to visualize which may be one of ‘site’, ‘sample’, or ‘rep’ (level), the true values of the probability parameters which may be used during simulation (truth), the number of credibility intervals to plot at a time (n), quantiles defining the credibility intervals to be provided (quantiles) and the number of samples to discard as warm-up (burnin).

2.2 | Example analysis

The functions in the msocc package are demonstrated on an eDNA survey of tidewater goby, found along the coast of California, USA (Schmelzle & Kinziger, 2016). In this survey, water samples were collected from 39 sites along the California coast. The number of samples collected at each site ranged from two to 23, and six PCR replicates were tested for the presence of goby DNA from each sample. In addition to detection data, environmental covariates were collected at all 39 sites; the first five rows of the data are provided in Table 1.

In these data, twg represents an index of goby density (catch-per-effort), sal represents the salinity in parts per thousand, turb represents the turbidity in filtration time, fish represents an index of non-goby fish density (catch-per-effort) and veg is a logical indicator for the presence of vegetation. Schmelzle and Kinziger (2016) originally fit a suite of models to these data using the WINBUGS package, and Dorazio and Erickson (2018) recreated those results using their eDNAoccupancy package. Those results are again recreated here for comparative purposes.

```
> # prep data frames
> site.df <- goby %>%
+      distinct(site, .keep_all = TRUE)
> sample.df <- goby
> detect.df <- goby %>%
+       select(-c(twg:veg))
> # fit model
> goby_mod <- msocc_mod(detect.df, num.mcmc = 11000,
+      site = list(model = ~ veg, cov_tbl = site.df),
+      sample = list(model = ~ sal + twg, cov_tbl = sample.df),
+      rep = list(model = ~ sal + fish + turb, cov_tbl = sample.df),
+      priors = list(site = list(mu0 = 0, Sigma0 = 1),
+            sample = list(mu0 = 0, Sigma0 = 1),
+            rep = list(mu0 = 0, Sigma0 = 1),
+            a0 = 1, b0 = 1),
+      progress = T, print = NULL, seed = 02042020, beta_bin = T)
Iteration 11000 of 11000; 100% done. Current runtime of 0.66
  minutes.
|=========================================| 100%
```

An overall summary of occupancy at all three levels is provided using posterior_summary(); the first six sites are presented below. Note that by default, posterior_summary() returns only unique combinations of site, sample and replicate probabilities.

TABLE 1 First five rows of tidewater goby data

Site	Sample	pcr1	pcr2	pcr3	pcr4	pcr5	pcr6	twg	sal	turb	fish	veg
Big_Lagoon	1	1	1	1	1	1	1	26.63	1.75	132.00	80.00	1
Big_River	1	0	0	0	0	0	0	0.00	26.00	78.80	17.20	0
Caspar_Creek	1	0	0	0	0	0	0	0.00	20.70	413.00	1.20	0
Elk_Creek	1	0	0	0	0	0	0	0.00	30.00	144.67	19.30	0
Garcia_River	1	0	0	0	0	0	0	0.00	23.50	41.00	0.00	0

```
> # numerical summary
> head(posterior_summary(goby_mod, level = "overall", burnin = 1000))
  site        sample  rep  psi        theta      p
1 Big_        1       1    0.7706495  0.8845290  0.8669207
    Lagoon
2 Big_River   1       1    0.2231190  0.6650525  0.5446235
3 Caspar_     1       1    0.2231190  0.7222719  0.8575822
    Creek
4 Davis_      1       1    0.7706495  0.8920972  0.8278296
    Lake
5 Dead_       1       1    0.2231190  0.7242514  0.8292023
    Mouse_
    Marsh
6 Eel_River_  1       1    0.7706495  0.6399665  0.6679964
    Estuary_
    Preserve
```

These results match those provided by the eDNAoccupancy package. Credibility intervals are available for each estimate by specifying a particular level of the model using the level argument. These intervals can also be visualized using the cred_plot() function; credibility intervals for the sample-level occupancy parameter for the first six sites are provided in Figure 1.

2.3 | Additional tools

In addition to the functions described in Sections 2.1 and 2.2, the msocc package contains tools to calculate Bayesian information criterion and simulate data from multi-scale occupancy models. The waic() function can be used to calculate the Watanabe–Akaiki information criterion (WAIC) on a suite of models (Gelman, Hwang, & Vehtari, 2013), while the msocc_sim() function is used to simulate data from a multi-scale occupancy model. For details on each functions' use, please see the GitHub page for msocc or the R help page for either function.

3 | SUPPLEMENTAL WEB APPLICATION

The msocc package is also equipped with an R Shiny web application capable of fitting models, visualizing the results from fitted models and conducting precision analyses. The web application accommodates the first two tasks by providing a point-and-click interface into the msocc_mod(), posterior_summary() and cred_plot() functions. To fit a multi-scale occupancy model with the msocc web application, users must upload the detection and covariate data frames in either .Rdata or .csv format. Once uploaded, the web application allows the user to visualize each data frame and specify the model to be fit (Figure 2). The application also allows the user to download the fitted model in .Rdata format.

To analyse a fitted model, the user must upload the model in .Rdata format. Once uploaded, the application provides a review of the fitted model and prompts the user to view table summaries of the occupancy parameters or visualize credibility intervals for the occupancy parameters (Figure 3). Users are also able to visualize traceplots for each parameter, though more formal assessments of convergence are available through the coda package (Plummer, Best, Cowles, & Vines, 2006).

The web application conducts precision analyses by repeatedly calling msocc_sim() in the background for varying sample sizes and fitting a model to each simulated dataset; the user determines at which level of the model to vary the sample size. Credibility intervals are then calculated for each of the parameters in these models and the widths are stored. These credibility interval widths are then plotted for each level of the model (Figure 4). This process allows users of the web application to assess the value of increasing the sample size at any level of the model in terms of precision, which is an important aspect of eDNA-based work (Erickson, Merkes, & Mize, 2019). Figure 4 suggests little to no additional precision is gained by collecting more than five replicates from each sample for the specified design.

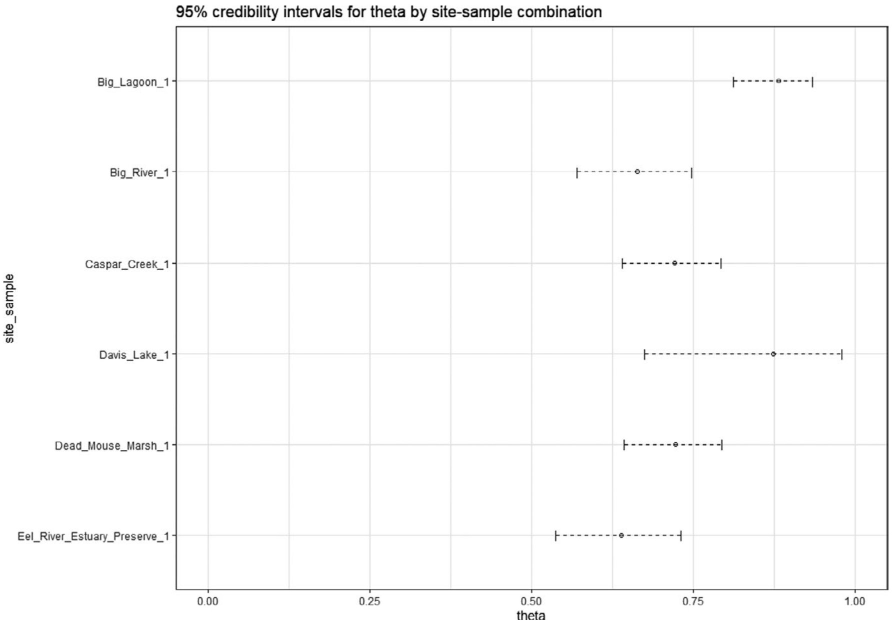
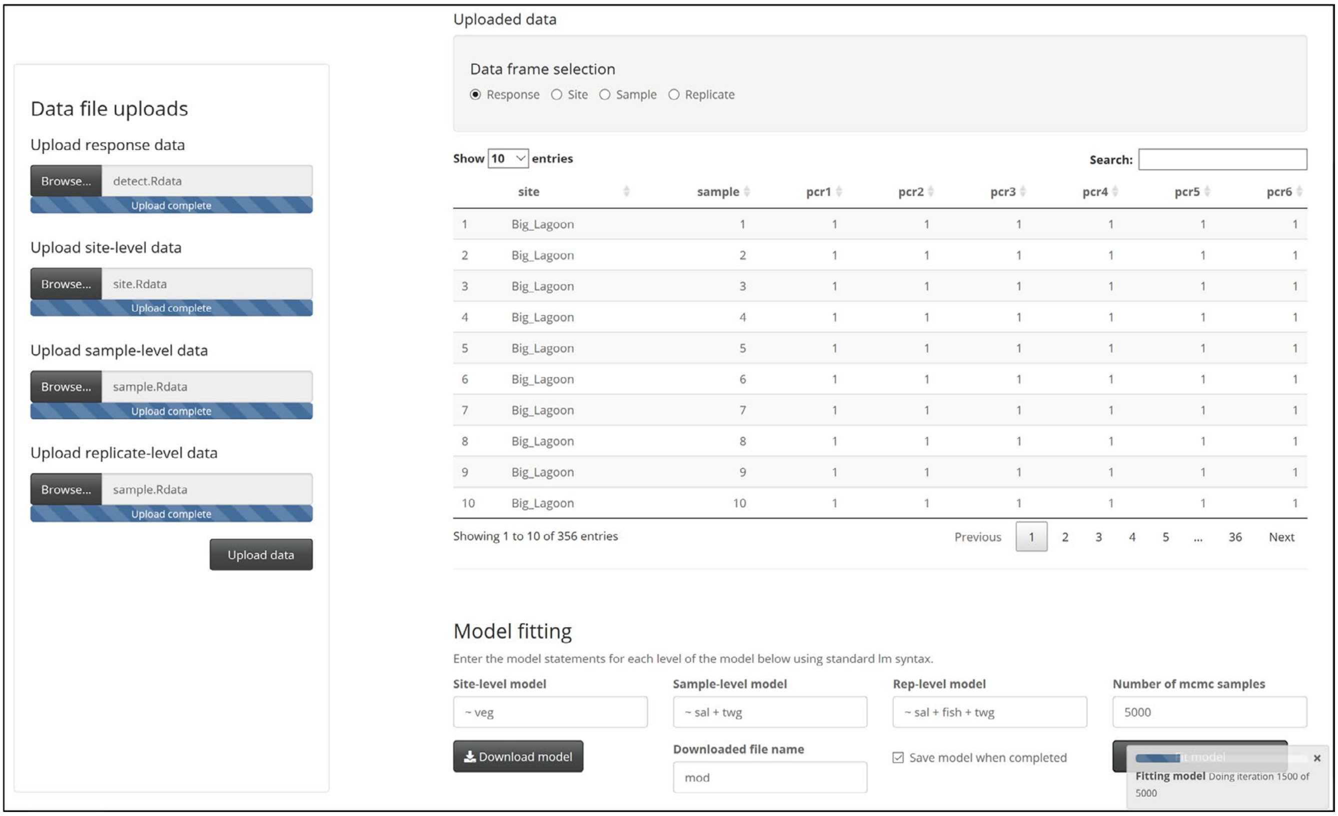


FIGURE 1 Plot of 95% credibility intervals for the sample level occupancy probability parameters of the first six sites of the tidewater goby data



Data file uploads

Upload response data: detect.Rdata

Upload site-level data: site.Rdata

Upload sample-level data: sample.Rdata

Upload replicate-level data: sample.Rdata

Uploaded data

Data frame selection

● Response ○ Site ○ Sample ○ Replicate

Show 10 entries

Search:

	site	sample	pcr1	pcr2	pcr3	pcr4	pcr5	pcr6
1	Big_Lagoon	1	1	1	1	1	1	1
2	Big_Lagoon	2	1	1	1	1	1	1
3	Big_Lagoon	3	1	1	1	1	1	1
4	Big_Lagoon	4	1	1	1	1	1	1
5	Big_Lagoon	5	1	1	1	1	1	1
6	Big_Lagoon	6	1	1	1	1	1	1
7	Big_Lagoon	7	1	1	1	1	1	1
8	Big_Lagoon	8	1	1	1	1	1	1
9	Big_Lagoon	9	1	1	1	1	1	1
10	Big_Lagoon	10	1	1	1	1	1	1

Showing 1 to 10 of 356 entries

Model fitting

Enter the model statements for each level of the model below using standard lm syntax.

Site-level model: ~ veg

Sample-level model: ~ sal + twg

Rep-level model: ~ sal + fish + twg

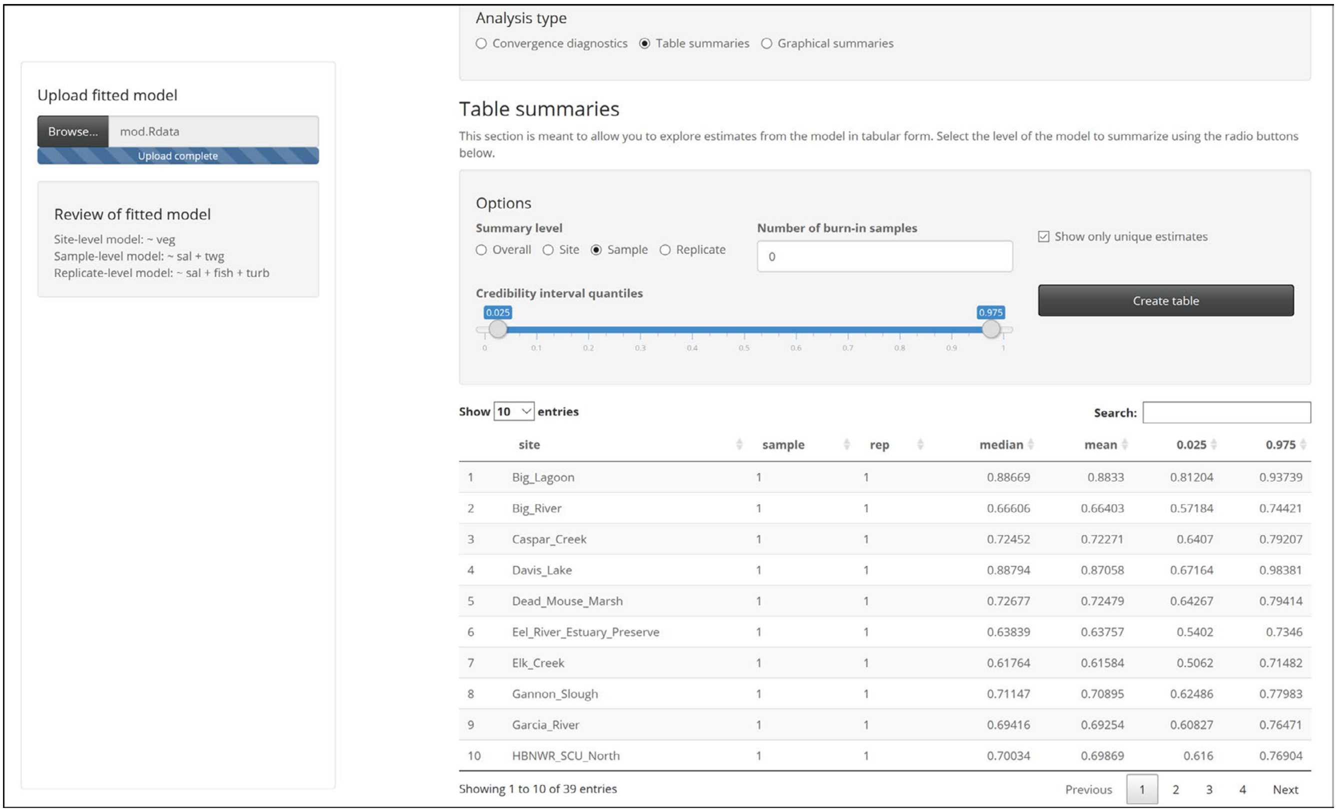
Number of mcmc samples: 5000

Downloaded file name: mod

☑ Save model when completed

Fitting model Doing iteration 1500 of 5000

FIGURE 2 Model fitting tab of the msocc web application



Upload fitted model: mod.Rdata

Review of fitted model

Site-level model: ~ veg
Sample-level model: ~ sal + twg
Replicate-level model: ~ sal + fish + turb

Analysis type

○ Convergence diagnostics ● Table summaries ○ Graphical summaries

Table summaries

This section is meant to allow you to explore estimates from the model in tabular form. Select the level of the model to summarize using the radio buttons below.

Options

Summary level: ○ Overall ○ Site ● Sample ○ Replicate

Number of burn-in samples: 0

☑ Show only unique estimates

Credibility interval quantiles: 0.025 – 0.975

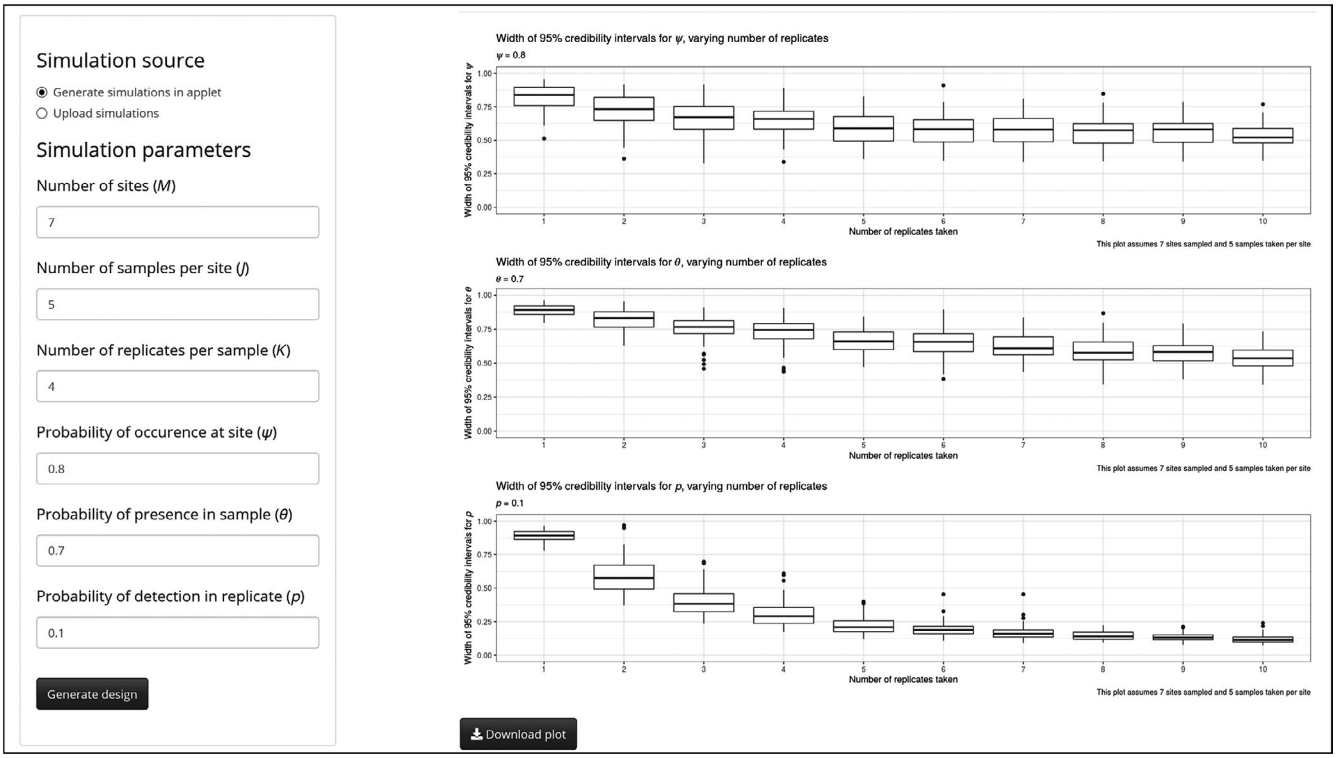
Show 10 entries

Search:

	site	sample	rep	median	mean	0.025	0.975
1	Big_Lagoon	1	1	0.88669	0.8833	0.81204	0.93739
2	Big_River	1	1	0.66606	0.66403	0.57184	0.74421
3	Caspar_Creek	1	1	0.72452	0.72271	0.6407	0.79207
4	Davis_Lake	1	1	0.88794	0.87058	0.67164	0.98381
5	Dead_Mouse_Marsh	1	1	0.72677	0.72479	0.64267	0.79414
6	Eel_River_Estuary_Preserve	1	1	0.63839	0.63757	0.5402	0.7346
7	Elk_Creek	1	1	0.61764	0.61584	0.5062	0.71482
8	Gannon_Slough	1	1	0.71147	0.70895	0.62486	0.77983
9	Garcia_River	1	1	0.69416	0.69254	0.60827	0.76471
10	HBNWR_SCU_North	1	1	0.70034	0.69869	0.616	0.76904

Showing 1 to 10 of 39 entries

FIGURE 3 Table summaries for the fitted goby model from the msocc web application. Convergence diagnostics and graphical summaries of the posterior distribution are also available



Simulation source

● Generate simulations in applet

○ Upload simulations

Simulation parameters

Number of sites (*M*): 7

Number of samples per site (*J*): 5

Number of replicates per sample (*K*): 4

Probability of occurence at site (*ψ*): 0.8

Probability of presence in sample (*θ*): 0.7

Probability of detection in replicate (*p*): 0.1

FIGURE 4 Credibility width analysis from the msocc web application. This analysis assumes seven sites are surveyed and five samples are taken from each site. The number of replicates taken varies from 1 to 10

TABLE 2 Summary of the effective sample size comparison. The values in this table describe the results from drawing 11,000 samples from the the joint posterior distribution of the goby model 10 times

Package	Data	Average time	Average ESS	Average ESS/min
msocc	Unscaled	36.41 s	4,029.61	6,640.39
msocc	Scaled	36.07 s	4,161.01	6,921.56
eDNAoccupancy	Unscaled	2,198.72 s	2,958.19	80.71
eDNAoccupancy	Scaled	1,839.45 s	2,908.97	94.87

4 | COMPARISON TO eDNAoccupancy

The msocc package is designed to be more expedient than eDNAoccupancy. The msocc_mod() function relies on a Gibbs sampler, and therefore converges to the posterior distribution much quicker than occModel(), the model fitting function from eDNAoccupancy. Additionally, it is not necessary to centre and scale covariates before fitting a model with msocc_mod(), thereby allowing users to interpret results on the original scale of the data. Gradient-based sampling frameworks, such as Stan or Greta, also generally provide quicker convergence to the posterior distribution (Carpenter et al., 2017). However, many of these frameworks, including Stan and Greta, do not allow sampling of discrete parameters or latent variables (Golding, 2019; Stan Development Team, 2018). Consequently, to fit multi-scale occupancy models within these frameworks, practitioners must integrate out the latent variables prior to defining the likelihood (e.g. see Mize et al., 2019). In a hierarchical modelling framework with three levels, this task is non-trivial and not likely to be widely accessible to wildlife managers and other eDNA practitioners. Therefore, we have not included these techniques for comparison, as they lack the convenience and ease-of-implementation associated with canned R packages.

To compare these functions, we provide the amount of time taken to draw 11,000 samples from the joint posterior of the goby model defined above, and the effective sample size for the msocc_mod() and occModel() functions per minute using both scaled and unscaled covariates. The effective sample size represents the number of uncorrelated posterior samples to which an MCMC chain is equivalent, and therefore describes the degree of autocorrelation present in that chain; the effectiveSize() function in the coda package (Plummer et al., 2006) was used to calculate this value. This function provides effective sample sizes for each parameter in the model; the minimum effective sample size across all parameters for each model was chosen to summarize the fit. This process was repeated 10 times for each model and the results are summarized in Table 2. These models were fit on a Surface Book 2 laptop running Windows 10 with an i7-8650U CPU and 16GB of RAM.

Table 2 suggests that msocc is far more efficient than eDNAoccupancy, allowing models to be fit in seconds as opposed to half-hours. Additionally, models fit using msocc tend to have larger effective sample sizes than those fit by eDNAoccupancy. Consequently, users are required to take fewer samples from the posterior distribution when using msocc, and can do so much quicker than when using eDNAoccupancy.

5 | DISCUSSION

As eDNA surveys continue to gain popularity as a sensitive monitoring strategy for rare and cryptic species, the need for efficient modelling techniques of multi-scale occupancy data increases; the msocc package provides an efficient alternative to existing methods of fitting Bayesian multi-scale occupancy models. This computational advantage allows the research and wildlife management communities the flexibility to fit multiple models when investigating scientific hypotheses, an otherwise time-consuming task. Additionally, msocc requires no tuning when fitting models, easily handles covariates on non-standardized scales and is equipped with a web application capable of conducting precision analyses, fitting models and exploring model results. The culmination of these factors eases the burden of ecologists working with eDNA data, thereby improving their ability to assess their research questions and disseminate that information. Moreover, in situ eDNA workflows continue to gain traction. Such analyses require computationally expedient techniques to understand these data structures in real-time; the ability to do so results in up-to-date information that can be used to minimize negative outcomes and improve management decision-making.

The msocc package continues to be developed. In the future, we hope to add a coherent dynamic modelling framework that accommodates longitudinal eDNA surveys. Continued research into dynamic multi-scale occupancy models is essential as longitudinal eDNA monitoring programs gain popularity (Bálint et al., 2018; Hutchins, Sepulveda, Hopper, & Staigmiller, 2019; Pilliod, Laramie, MacCoy, & Maclean, 2019; Uchii, Doi, Yamanaka, & Minamoto, 2017). Once these methods have been developed, we intend to add them to msocc.

ACKNOWLEDGEMENTS

This work was supported by the USGS Ecosystems Mission Area. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the US Government. The time and effort of Christian Stratton and Dr Andy Hoegh was made possible with a CESU agreement (G17AC00147) between USGS-NOROCK and MSU.

AUTHORS' CONTRIBUTIONS

C.S. wrote the msocc package, wrote the supplemental web application and led the writing of the manuscript; A.H. and A.S. provided intellectual guidance on the development of msocc and the web application. All the co-authors assisted with edits and approved submission.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13442.

DATA AVAILABILITY STATEMENT

Data: The tidewater goby data were originally archived by Schmelzle and Kinziger (2016) at Dryad Digital Repository http://dx.doi.org/10.5061/dryad.6rs23. These data files are also accessible in the msocc package.

Package, web application and simulation scripts: The msocc package and documentation are hosted at https://github.com/StrattonCh/msocc; the package is also archived at the Zenodo online repository (Stratton, 2020a). The supplemental web application is hosted at https://christianstratton.shinyapps.io/eDNAapp/. The script file used to compare msocc and eDNAoccupancy is hosted at https://github.com/StrattonCh/msocc_supplemental_code; this file is also archived at the Zenodo online repository (Stratton, 2020b).

ORCID

Christian Stratton https://orcid.org/0000-0001-8051-2185
Adam J. Sepulveda https://orcid.org/0000-0001-7621-7028
Andrew Hoegh https://orcid.org/0000-0003-1176-4965

REFERENCES

Bálint, M., Pfenninger, M., Grossart, H.-P., Taberlet, P., Vellend, M., Leibold, M. A., … Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology & Evolution*, *33*(12), 945–957. https://doi.org/10.1016/j.tree.2018.09.003

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

Dorazio, R. (2015). Bayesian data analysis in population ecology: Motivations, methods, and benefits. *Population Ecology*, *58*, 31–44. https://doi.org/10.1007/s10144-015-0503-4

Dorazio, R. M., & Erickson, R. A. (2018). eDNAoccupancy: An R package for multiscale occupancy modelling of environmental DNA data. *Molecular Ecology Resources*, *18*(2), 368–380. https://doi.org/10.1111/1755-0998.12735

Eiler, A., Löfgren, A., Hjerne, O., Nordén, S., & Saetre, P. (2018). Environmental DNA (eDNA) detects the pool frog (*Pelophylax lessonae*) at times when traditional monitoring methods are insensitive. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-23740-5

Erickson, R., Merkes, C., & Mize, E. (2019). Sampling designs for landscape-level eDNA monitoring programs. *Integrated Environmental Assessment and Management*, *15*(5), 760–771. https://doi.org/10.1002/ieam.4155

Gelman, A., Hwang, J., & Vehtari, A. (2013). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*, 1–30.

Golding, N. (2019). greta: Simple and scalable statistical modelling in R. *Journal of Open Source Software*, *4*, 1601.

Hunter, M., Meigs-Friend, G., Ferrante, J., Smith, B., & Hart, K. (2019). Efficacy of eDNA as an early detection indicator for Burmese pythons in the ARM Loxahatchee National Wildlife Refuge in the greater Everglades ecosystem. *Ecological Indicators*, *102*, 617–622. https://doi.org/10.1016/j.ecolind.2019.02.058

Hutchins, P., Sepulveda, A. J., Hopper, L., & Staigmiller, K. (2019). The Yellowstone river fish-kill: Fish health informs and is informed by vital signs monitoring. *Yellowstone Science*, *27*(1), 55–57.

Klymus, K. E., Richter, C. A., Chapman, D. C., & Paukert, C. (2015). Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. *Biological Conservation*, *183*, 77–84. https://doi.org/10.1016/j.biocon.2014.11.020

Lodge, D., Turner, C., Jerde, C., Barnes, M., Chadderton, W., Egan, S., … Pfrender, M. (2012). Conservation in a cup of water: Estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, *21*, 2555–2558. https://doi.org/10.1111/j.1365-294X.2012.05600.x

Mize, E. L., Erickson, R. A., Merkes, C. M., Berndt, N., Bockrath, K., Credico, J., … Von Ruden, K. (2019). Refinement of eDNA as an early monitoring tool at the landscape-level: Study design considerations. *Ecological Applications*, *29*(6), e01951. https://doi.org/10.1002/eap.1951

Mordecai, R., Mattsson, B., Tzilkowski, C., & Cooper, R. (2011). Addressing challenges when studying mobile or episodic species: Hierarchical Bayes estimation of occupancy and use. *Journal of Applied Ecology*, *48*, 56–66. https://doi.org/10.1111/j.1365-2664.2010.01921.x

Nichols, J., Bailey, L., O'Connell, A., Talancy, N., Grant, E., Gilbert, A., … Hines, J. (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, *45*, 1321–1329. https://doi.org/10.1111/j.1365-2664.2008.01509.x

Pilliod, D., Laramie, M., MacCoy, D., & Maclean, S. (2019). Integration of eDNA-based biological monitoring within the U.S. Geological Survey's national streamgage network. *JAWRA Journal of the American Water Resources Association*, *55*(6), 1505–1518. https://doi.org/10.1111/1752-1688.12800

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11.

Polson, N., Scott, J., & Windle, J. (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. *Journal of the American Statistical Association*, *108*, 1339–1349. https://doi.org/10.1080/01621459.2013.829001

Robert, C. P. (2015). The Metropolis–Hastings algorithm. *arXiv preprint*.

Roberts, G., & Rosenthal, J. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statistical Science*, *16*, 351–367.

Schmelzle, M., & Kinziger, A. (2016). Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources*, *16*, 1–14. https://doi.org/10.1111/1755-0998.12501

Sengupta, M., Hellström, M., Kariuki, H., Olsen, A., Thomsen, P., Mejer, H., … Vennervald, B. (2019). Environmental DNA for improved detection and environmental surveillance of schistosomiasis. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(18), 8931–8940. https://doi.org/10.1073/pnas.1815046116

Sepulveda, A. J., Schmidt, C., Amberg, J., Hutchins, P., Stratton, C., Mebane, C., … Pilliod, D. S. (2019). Adding invasive species biosurveillance to the U.S. geological survey streamgage network. *Ecosphere*, *10*(8), 1–17. https://doi.org/10.1002/ecs2.2843

Stan Development Team. (2018). *Stan modelling language users guide and reference manual, version 2.18.0*. Retrieved from http://mc-stan.org

Stratton, C. (2020a). StrattonCh/msocc: msocc release (version 1.0.0). *Zenodo*, https://doi.org/10.5281/zenodo.3911239

Stratton, C. (2020b). StrattonCh/msocc_supplemental_code: msocc supplemental code release. *Zenodo*, https://doi.org/10.5281/zenodo.3908568

Uchii, K., Doi, H., Yamanaka, H., & Minamoto, T. (2017). Distinct seasonal migration patterns of Japanese native and non-native genotypes of common carp estimated by environmental DNA. *Ecology and Evolution*, *7*, 8515–8522. https://doi.org/10.1002/ece3.3346

Williams, K., Huyvaert, K. K., Vercauteren, K., Davis, A., & Piaggio, A. (2018). Detection and persistence of environmental DNA from an invasive, terrestrial mammal. *Ecology and Evolution*, *8*, 688–695. https://doi.org/10.1002/ece3.3698

How to cite this article: Stratton C, Sepulveda AJ, Hoegh A. msocc: Fit and analyse computationally efficient multi-scale occupancy models in R. *Methods Ecol Evol*. 2020;11:1113–1120. https://doi.org/10.1111/2041-210X.13442