Author/s:
Wilkinson, DP;Golding, N;Guillera-Arroita, G;Tingley, R;McCarthy, MA

Title:
A comparison of joint species distribution models for presence-absence data

Date:
2019-02-01

Citation:
Wilkinson, D. P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M. A. (2019). A comparison of joint species distribution models for presence-absence data. METHODS IN ECOLOGY AND EVOLUTION, 10 (2), pp.198-211. https://doi.org/10.1111/2041-210X.13106.

Persistent Link:
https://hdl.handle.net/11343/254331

Methods in Ecology and Evolution

MR DAVID PETER WILKINSON (Orcid ID : 0000-0002-9560-6499)

DR NICK GOLDING (Orcid ID : 0000-0001-8916-5570)

DR GURUTZETA GUILLERA-ARROITA (Orcid ID : 0000-0002-8387-5739)

DR REID TINGLEY (Orcid ID : 0000-0002-7630-7434)

DR MICHAEL ANDREW MCCARTHY (Orcid ID : 0000-0003-1039-7980)

Article type : Research Article

Editor : Dr Pedro Peres-Neto

Running title: Comparing JSDMs for presence-absence data.

Title: A comparison of joint species distribution models for presence-absence data.

David P. Wilkinson[1*], Nick Golding[1], Gurutzeta Guillera-Arroita[1], Reid Tingley[1], Michael A. McCarthy[1]

1. School of BioSciences, University of Melbourne, Parkville, 3010, Victoria, Australia

*Corresponding author: davidpw@student.unimelb.edu.au

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.13106

ABSTRACT

1. Joint species distribution models (JSDMs) account for biotic interactions and missing environmental predictors in correlative species distribution models. Several different JSDMs have been proposed in the literature, but the use of different or conflicting nomenclature and statistical notation potentially obscures similarities and differences among them. Furthermore, new JSDM implementations have been illustrated with different case studies, preventing direct comparisons of computational and statistical performance.
2. We aim to resolve these outstanding issues by (i) highlighting similarities among seven presence-absence JSDMs using a clearly-defined, singular notation; and (ii) evaluating the computational and statistical performance of each JSDM using six datasets that vary widely in numbers of sites, species, and environmental covariates considered.
3. Our singular notation shows that many of the JSDMs are very similar, and in turn parameter estimates of different JSDMs are moderate to strongly, positively-correlated. In contrast, the different JSDMs clearly differ in computational efficiency and memory limitations.
4. Our framework will allow ecologists to make educated decisions about the JSDM that best suits their objective, and enable wider uptake of JSDM methods among the ecological community.

Keywords: biotic interactions, community assembly, hierarchical models, joint species distribution model, latent factors, presence-absence, residual correlation

INTRODUCTION

Understanding the factors underlying species' geographic distributions is a fundamental goal of ecology. Abiotic factors, such as temperature or rainfall, and biotic factors, such as species interactions or population dynamics, ultimately drive species distributions (Hutchinson 1957). Yet correlative species distribution models (SDMs), the most common tool for predicting species distributions, typically quantify species' environmental relationships without explicitly considering effects of species interactions (Dormann *et al.* 2012; Wisz *et al.* 2013). In contrast, community ecology generally studies co-occurrence patterns and community structure without accounting for co-occurrence due to shared environmental responses among species (Hardy 2008).

Several statistical approaches have been proposed to quantify species' environmental relationships and interactions in a single model. Within the occupancy-detection modelling framework (Guillera-Arroita 2017), early examples explicitly modelled effects of biotic interactions via linear models, including pair-wise and higher-order effects (MacKenzie, Bailey & Nichols 2004), or dominate/subordinate species relationships, where the dominant species' occupancy state influences that of the subordinate species, but not vice versa (Richmond, Hines & Beissinger 2010; Waddle *et al.* 2010). Practical and computational limits restrict the application of these methods to communities of only a handful of species (Richmond, Hines & Beissinger 2010).

Other methods incorporate species co-occurrence data into the classical SDM framework. Distribution estimates of additional species can be used alongside abiotic variables as predictors (Leathwick & Austin 2001; Araújo & Luoto 2007; Meier *et al.* 2010; Pellissier *et al.* 2010), or a species' predicted distribution can be restricted to that of another it depends on (Schweiger *et al.* 2012). Like dominate/subordinate models, however, these approaches only incorporate unidirectional interactions (Kissling *et al.* 2012; Pollock *et al.* 2014).

Joint species distribution models (JSDMs) have emerged as extensions of SDMs to capture the effects of biotic interactions in communities (Ovaskainen, Hottola & Siitonen 2010; Kissling *et al.* 2012; Clark *et al.* 2014, 2017; Pollock *et al.* 2014; Warton *et al.* 2015; Golding, Nunn & Purse 2015; Letten *et al.* 2015; Harris 2015; Thorson *et al.* 2016; Ovaskainen *et al.* 2016b,a; Hui 2016; Nieto-Lugilde *et al.* 2017). JSDMs simultaneously model multiple species' distributions, accounting for both environmental relationships and residual associations (that might arise from species interactions) on species co-occurrence (Pollock *et al.* 2014). Thus, JSDMs inform about biotic and abiotic constraints on species' distributions (Pollock *et al.* 2014), and can improve predictions of community composition (Harris 2015). Promising advances in JSDMs have led to the development of several statistical models and computational implementations. This diversity of methods and their presentation might hamper wider use of JSDMs by ecologists, since prospective users have little guidance on which implementation best suits their objectives.

Our study formally compares seven presence-absence JSDMs that consider species interactions by characterizing the residual correlation among species not captured by environmental predictors. These models are statistically complex, and defined in their respective papers using different, potentially conflicting, notation and terminology. In addition, each JSDM was evaluated on a different dataset; any ecological relationships inferred are therefore not directly comparable, nor are reports of an implementation's computational efficiency. To help ecologists better understand the different models, we define them using a consistent notation to elucidate their similarities and differences, and compare their parameter estimates and computational performance against six datasets. The results can help ecologists make informed decisions about appropriate JSDM selection.

MATERIALS & METHODS

Approach

Each JSDM considered here consists of three components: a Bayesian statistical model; software to fit the model to data; and default or suggested settings for model fitting (e.g., choice of priors and Markov chain Monte Carlo (MCMC) sampling regime). A JSDM's computational efficiency and parameter estimates will undoubtedly be affected by all of these components. Our aim was not to disentangle the effects of each of these components, but to highlight the differences and similarities between the overall methodologies, as applied in their respective papers. The term JSDM hereafter refers to these overall methodologies, unless otherwise stated.

Datasets

Each JSDM was fit to six presence-absence datasets from recent JSDM papers (Ovaskainen, Hottola & Siitonen 2010; Pollock *et al.* 2014; Golding, Nunn & Purse 2015; Harris 2015; Ovaskainen *et al.* 2016b) to compare parameter estimates and computational efficiency across a range of datasets. Datasets not made publicly available were sourced from the authors. These datasets covered a broad range of taxa, geographic locations, and numbers of sites and species (Table 1**)**. Datasets were used as originally published with these exceptions: testing and training splits were merged; the Breeding Bird Survey online repository had been updated resulting in more sites; and we only used one of the eighteen fungi datasets from Ovaskainen, Hottola & Siitonen (2010). Geographic coordinates were rescaled to have a maximum distance between sites of 1 to avoid issues of numerical instability in the spatial model.

The Models

We restrict our evaluation of JSDMs to seven models for species presence-absence data that account for species interactions via residual correlations. These JSDMs are extensions of the generalised linear modelling (GLM) framework, which is widely used for modelling species distribution data (Gelfand *et al.* 2006). The statistical models are defined below using a common notation. The following terms are consistent across all models: $\boldsymbol{y}$, the response variable; $1(.)$, an indicator function that returns 1 when the expression in brackets is true and 0 otherwise; $\boldsymbol{z}$, a normally-distributed latent variable; $\boldsymbol{\mu}$, the linear predictor for the measured covariates; $\mathbf{X}$, the matrix of measured covariates; $\boldsymbol{\beta}$, the matrix of regression coefficients; and $\mathbf{I}$, the identity matrix. Subscript notation for sites is $i = 1, \ldots, n$; for species $j = 1, \ldots, J$; and for predictors $k = 1, \ldots, K$.

Each JSDM is built on the foundation of Chib & Greenberg's (1998) multivariate probit regression model (hereafter, the core model). This model uses a latent variable parameterisation of a probit model rather than the probit link directly. The probability of species presence is modelled as the probability of a latent multivariate normally-distributed variable exceeding a threshold, such that $y_{ij} = 1$ if $z_{ij} > 0$, and $y_{ij} = 0$ otherwise. The probability of species presence at a site is represented by a one-dimensional latent variable, while the community present at a site is represented by a multi-dimensional latent variable (see Pollock *et al* (2014) for a visual representation). This core model is described as follows:

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\boldsymbol{e}_i \sim MVN(\mathbf{0}, \mathbf{R})$$

The probability of species j being present at site i is the probability that latent variable z_{ij} is greater than zero. The latent variable is the sum of the linear predictor μ_{ij} and the correlated residual error e_{ij}. The linear predictor is the product of the measured environmental variables $\boldsymbol{X}_{i,.}$, and their corresponding regression coefficients $\boldsymbol{\beta}_{.,j}$, as in generalised linear models. Correlations in the residual error $\boldsymbol{e}_i$ are captured in $\mathbf{R}$, a symmetric and positive-definite matrix; its diagonal elements are 1 and its off-diagonal elements are restricted between -1 and 1. Standard deviations are constrained to equal to 1 in probit regression, thus covariance and correlation matrices are equivalent. The elements of $\mathbf{R}$ reflect species co-occurrence patterns not described by the environmental predictors (i.e. species interactions, or missing predictors). A limitation of this model is that it does not specify overdispersion in species occurrences like other extensions of the GLM framework (e.g., generalised linear mixed models), but it could be extended to do so.

The following sections define statistical models and describe the software and MCMC regimes of the JSDMs. All models were fit using their default setting for priors and MCMC regimes. Table 2 provides an overview of the statistical models'respective features, and Table 3 provides a glossary of all introduced notation.

(1) Multivariate Probit Regression (MPR)

The multivariate probit regression (MPR) model used by Golding *et al* (2015), *BayesComm*, is identical to the core model. The regression coefficients have a normal prior, $\beta \sim N(0, 10)$, and the correlation coefficients an inverse Wishart prior with $n + 2J$ degrees of freedom and scale matrix $\mathbf{I}$. MPR was fit in R (R Core Team 2016) by MCMC using a Gibbs sampler implemented in R and C++. A single MCMC chain of 11,000 samples, discarding the first 1,000 as burn-in, sampled the posterior

distribution.

(2) Multivariate Logistic Regression (MLR)

Ovaskainen, Hottola & Siitonen (2010) introduced a logistic regression version of the core model based on O'Brien & Dunson (2004). The model (hereafter MLR) is as follows:

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + logit\left(\Phi\left[e_{ij}\right]\right)$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\boldsymbol{e}_i \sim MVN(\mathbf{0}, \mathbf{R})$$

Here, Φ is the cumulative density function of the standard normal distribution. The regression coefficients have a normal prior, $\beta \sim N(0{,}10)$, and the correlation coefficients a uniform prior, $U(-1{,}1)$. MLR was fit in Mathematica (Wolfram Research Inc 2016), and the posterior distributions sampled via MCMC using a Metropolis-within-Gibbs sampler with 100,000 samples (thinned to keep 1 every 10) discarding the first 5,000 as burn in.

(3) Hierarchical Multivariate Probit Regression (HPR)

The hierarchical multivariate probit regression model of Pollock *et al* (2014) (hereby HPR) models the regression coefficients hierarchically such that β_{jk} is drawn from a normal distribution with mean ω_k and standard deviation σ_k.

$$y_{ij} = 1(z_{ij} > 0)$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\beta_{jk} \sim N(\omega_k, \sigma_k)$$

$$e_i \sim \boldsymbol{MVN}(\mathbf{0}, \mathbf{R})$$

Regression coefficients have a normal prior, $N(0, 100)$, on ω and a uniform prior, $U(0,100)$, on σ. Correlation coefficients have an inverse Wishart prior with $J + 1$ degrees of freedom and an $\mathbf{I}$ scale matrix. HPR was fit in R, and the posterior distributions sampled via MCMC using Gibbs sampling in JAGS (Plummer 2003), with three chains of 1,000,000 samples (thinned to keep 1 every 1,000) with the first 15,000 discarded as burn in.

(4) Multivariate Probit Regression with Latent Factors (LPR)

The *boral* JSDM (Hui 2016) is a multivariate probit regression model with latent factors (hereafter LPR).

$$y_{ij} = 1(z_{ij} > 0)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$$

$$\varepsilon_i \sim N(\mathbf{0}, \mathbf{I})$$

This differs from the core model by using latent factors ("hypothetical" unmeasured variables) to explain any residual covariation that is not accounted for by the measured variables. This explained variation is the product of H unmeasured latent factors, $\boldsymbol{\eta}_{i,.}$ ($h = 1, \ldots, H$), and the factor loadings, $\boldsymbol{\lambda}_{.,j}$, of species j to latent factor h. We can interpret $\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$ and $\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$ as linear predictors for measured and unmeasured variables respectively. Now z_{ij} is the sum of the linear predictor of measured covariates, μ_{ij}, the linear predictor of unmeasured covariates, ν_{ij}, and uncorrelated residual error ε_{ij}. This model can be defined equivalently as an extension of MPR where the correlation matrix has low-rank structure, modelled as: $\mathbf{R} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{I}$ (Warton *et al.* 2015). Species interactions can be estimated by converting the factor loadings to correlation coefficients using this expression.

Latent factors bring considerable computational benefits by reducing the number of coefficients to estimate. A full-rank correlation matrix has $J(J-1)/2$ parameters, while a factor loadings matrix only JH. The parameter ratio between latent factor and correlation matrix models is therefore $2H/(J-1)$, so the parameter reduction increases with J. Where $J = 10$ and $H = 2$, the latent factor model has 20 parameters rather than 45 (44%), whilst for $J = 100$ and $H = 5$ the latent factor model has 500 parameters rather than 4,950 (10%). This reduction does not necessarily correspond to an increase in parsimony, as the choice of priors for these matrices also impact model complexity. As latent factors are trying to approximate a fully unstructured correlation matrix, a greater H provides a better approximation, while a smaller H reduces computational requirements. The number of latent factors balances model simplicity and model fit, and 2-8 latent factors has been suggested as suitable in practice (Warton *et al.* 2015; Hui 2017).

The regression coefficients in the LPR implementation have a normal prior, $\beta \sim N(0, 10)$, and the latent factors a normal prior, $\eta \sim N(0,4.47)$, constrained such that upper diagonal elements are 0, and the diagonal elements are positive, $U(0,20)$. LPR was fit in R using the *boral* package. The posterior distribution was sampled via MCMC, with a Gibbs sampler using JAGS, in a single chain of 60,000 samples (thinned to keep 1 every 50) discarding the first 10,000 as burn in.

(5) Dimension Reduction Model (DPR)

The multivariate generalised regression model of Clark *et al* (2017), *gjam* (hereby DPR), fits various types of response data. For presence-absence data it is a multivariate probit regression model that takes on two different forms depending on the size of the dataset. The small dataset form is equivalent to the core model. For datasets above a size threshold ($J > 100$ or $J > \frac{2}{3} * n - 1$) , it is similar to LPR, but with an additional dimension reduction step (with respect to species) using the Dirichlet process (Taylor-Rodríguez *et al.* 2016):

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$$

$$\boldsymbol{\lambda}_{.,j} = \mathbf{A}_{\boldsymbol{l}_j}$$

$$\varepsilon_i \sim N(\mathbf{0}, \mathbf{I})$$

The Dirichlet process assumes that groups of species will have the same response to the unmeasured variables, so *a priori* reduces $\boldsymbol{\lambda}$ along the j-dimension to a matrix, $\mathbf{A}$, for a fixed number, L, of species archetypes (rather than J species). The Dirichlet process can be seen as a distribution of distributions, and is used to cluster species into archetypes, via an index variable $\boldsymbol{l}$. Using $\mathbf{A}$ as a look-up table for the low-rank $\boldsymbol{\lambda}$ allows conversion back to full-rank $\boldsymbol{\lambda}$, and then estimate correlation coefficients via the factor loadings. The parameter ratio of DPR to a full-rank correlation matrix is $2LH/(J(J-1))$, which is typically much smaller than 1. This decouples the number of parameters to estimate from J, though adequate L and H must be used to accurately model interactions.

The regression coefficients in the DPR implementation have an improper uniform prior, $U(-Inf, Inf)$, correlation coefficients use an inverse-Wishart prior with $n-K+J-1$ degrees of freedom and an $\mathbf{I}$ scale matrix, and latent factors are normally distributed as $\eta_{ih} \sim N(0,1)$. DPR was fit in R using the *gjam* package. The posterior distribution was sampled via MCMC using a Gibbs sampler in R of a single chain of 60,000 samples (thinned to keep 1 every 50), discarding the first 10,000 as burn in.

(6) Hierarchical Multivariate Probit Regression with Latent Factors (HLR-S and HLR-NS)

The last two JSDMs are spatially-explicit (HLR-S) and non-spatially-explicit (HLR-NS) versions of the hierarchical multivariate probit regression with latent factors used by Ovaskainen *et al*. (2016b):

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\beta_j \sim MVN\left(\boldsymbol{\omega}_{\boldsymbol{l}_j}, \boldsymbol{\sigma}\right)$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$$

$$\varepsilon_i \sim N(0,1)$$

The HLR models make use of latent factors like LPR and DPR, but also use hierarchical regression coefficients like HPR. However, instead of independent normal distributions unique to the K covariates, $\beta_{.,j}$ is drawn from a multivariate normal distribution, allowing for correlation among coefficients. The mean of this distribution, $\boldsymbol{\omega}_{\boldsymbol{l}_j}$, is unique to species' archetype, $\boldsymbol{l}$. Archetype is specified *a priori* in HLR models using trait data, assuming that species with similar traits exhibit similar responses. Here, $\boldsymbol{l}_j$ is set to 1 as trait data was unavailable for most datasets. Thus, each species' $\beta_{.,j}$ are independently drawn from a shared distribution. The hierarchical regression coefficients have a normal prior, $N(0,1)$ for $\boldsymbol{\omega}$, and an inverse Wishart prior with five degrees of freedom and a $\mathbf{I}$ scale matrix for $\boldsymbol{\sigma}$. The latent factors are normally distributed as $\eta_{ih} \sim N(0,1)$ in HLR-NS, and as independent spatially homogenous Gaussian processes with exponential covariance in HLR-S:

$$Cov(\eta_{ih}, \eta_{i'h}) = exp(-d_{ii'}/\alpha_h)$$

Here, d is the spatial distance between sampling units, and α is a positive parameter controlling decay in correlation with distance. At zero distance the covariance function is 1, so the latent factors have zero mean and unit variance like in the non-spatial model, and this declines towards zero with increasing α.

These JSDMs were fit in MATLAB (The MathWorks Inc 2016). The posterior distribution was sampled via MCMC using a Gibbs sampler in a single chain of 50,000 samples (thinned to keep 1 every 50) with the first 10,000 discarded as burn in.

Model Fitting

Model fitting was carried out on desktop computers with identical specifications (Windows 7 OS, i7-6700 CPU @ 3.40GHz, 16GB RAM) to compare computational performance. JSDM software that explicitly enabled model fitting in parallel was run so (HPR), while those that did not were run sequentially. An upper time limit for runtime was set at seven days, after which the JSDM was deemed incompatible with that dataset on a desktop computer. JSDMs that reached this limit were then run on high performance computing infrastructure with increased memory availability (Boab: Dual Intel Xeon E5-2699 CPUs @ 2.30GHz, ~40GB RAM per model. Spartan: Intel(R) Xeon(R) CPU E5-2643 v3 @ 3.40GHz, 200GB RAM (University of Melbourne 2017)) with the same time limit. JSDMs were fit using: Mathematica 10.4 (Wolfram Research Inc 2016) (MLR), MATLAB R2016a (The MathWorks Inc 2016) (HLR models), and R 3.2.5 (R Core Team 2016) (all others). Model-specific R packages were *BayesComm* 0.1-2 (Golding & Harris 2015) (MPR), *boral* 1.3.1 (Hui 2017) (LPR), and *gjam* 2.1.1 (Clark *et al.* 2017) (DPR). JAGS 4.3.0 (Plummer 2003) was used for some MCMC sampling.

Statistical and computational comparison

We compared the computational performance and parameter estimation of the different JSDMs. Computational metrics included: ease of use (time required for dataset formatting); dataset compatibility (model completion within seven days on a desktop computer); runtime; and sampler efficiency (effective sample size (ESS) per unit CPU-time). To prevent bias in the *ease of use* metric due to user experience gained during the process, the JSDM/dataset combinations were fit by the

same author (DW) in a random order, with two exceptions: (i) HLR-S and HLR-NS differed only by a single setting, so ease of use was measured for HLR-S only, and (ii) LPR and DPR were not available when the study began and were included later.

ESS estimates are not directly comparable unless each JSDM was fit using identical MCMC regimes, so each JSDM was fit to the mosquito dataset a second time with identical MCMC regimes, consisting of a single chain of 11,000 iterations with no thinning and the first 1,000 discarded as burn-in. ESS was calculated separately for regression and correlation coefficients by dividing the mean effective number of samples of each coefficient in the posterior distribution by the run-time.

Statistical inference was assessed by comparing the estimated regression and correlation coefficients. For latent factor JSDMs, the residual covariance matrixwas recovered from the factor loadings ($\mathbf{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{I}$) and scaled to the correlation matrix $\mathbf{R}$. To evaluate the similarity of JSDM regression coefficient estimates, we compared posterior mean values and their 95% credible intervals, and calculated the Pearson correlation coefficient between mean estimates of each coefficient of each JSDM. Variation in coefficient estimates was evaluated using four uncertainty metrics calculated from post-hoc standardised posterior distributions: coefficient of variation, Gini coefficient, quartile coefficient of dispersion, and 90% quantile coefficient of dispersion. Similarity of correlation coefficient estimates was evaluated by calculating the Pearson correlation coefficient between mean estimates for each JSDM, and the magnitude and uncertainty (95% credible interval width) of estimates was calculated relative to the simplest JSDM (MPR) as the baseline.

Simulation comparison

We also compared a subset of these models (MPR, HPR, LPR, DPR, and HLR-NS) when fit to datasets simulated under different modelling component assumptions. See Appendix S5.

RESULTS

Computational comparison

Only four JSDMs ran to completion within the time limit for all datasets: MPR, LPR, DPR, and HLR-NS (Table 4). MLR did not complete for four datasets, took 142 hours for the eucalypt dataset (nearly three times longer than HLR-S, and eight times longer than the closest non-spatial model, HPR), and 14 hours for the small frog data set; it was therefore deemed too slow for practical use and not considered further. Both HPR and HLR-S failed to complete within the time limit for the bird and butterfly datasets on desktop or high-performance computing resources. HLR-S was incompatible with the non-spatial fungi dataset.

MLR, HLR-S, and HLR-NS were the slowest for converting the datasets into the required format (40 +/- 29 minutes). These were the only JSDMs implemented in programming languages other than R, the language with which the experimenter was most familiar. Datasets for all other JSDMs were prepared within an average of five minutes.

MPR was the fastest to run to completion for all datasets, followed by LPR for the two smallest datasets, and HLR-NS was fastest for the four largest (Table 5**)**. HPR and HLR-S were considerably slower for all datasets. MPR and HLR-NS had substantially greater sampler efficiency than the other JSDMs; MPR was most efficient for the regression coefficients (Figure 1a), whereas HLR-NS was most efficient for correlation coefficients (Figure 1b).

Similarity of Parameter Estimates

All JSDMs yielded similar regression coefficients across all datasets (see Figure 2 for one species; all species in Appendix S1). LPR parameter estimates for binary variables in the fungi and mosquito datasets were markedly more uncertain and frequently stronger than those estimated by other JSDMs for species-environment pairs with complete separation (see Appendix S1, Figure 458). Results presented here are based on default priors, but the impact of complete separation is alleviated with more informative priors (Appendix S2).

Figure 3 shows the correlation of mean regression coefficient estimates between JSDMs, averaged across all datasets (individual datasets in Appendix S4). There is strong correlation (>0.69) between all JSDM estimates, with the strongest correlations between HPR, HLR-S and HLR-NS (>0.9), and between MPR and all models (>0.75).

There was no consistent, substantial difference in the variation of regression coefficient estimates between JSDMs (see Figure 4for the Gini coefficient for all species in the Eucalypt dataset). In a small number of instances, there were differences between JSDMs for particular coefficients (Appendix S3, Figure 21), but there was no apparent pattern to these occurrences, and Appendix S3 Figures 25-28 show only minor differences between JSDMs for each uncertainty metric across all datasets. There was no difference in observed patterns between different uncertainty metrics.

All JSDMs identified similar patterns in estimated correlation coefficients (Figure 5). The direction of species interactions was largely the same among JSDMs, but estimates differed in strength and uncertainty. Observed differences in the direction of estimated correlations between JSDMs involved weak and uncertain coefficients.

Figure 6 shows the correlation of mean correlation estimates between JSDMs for all datasets (individual datasets in Appendix S4). Strong, positive correlations (>0.7) were observed between most JSDMs. However, there was only moderate correlation (>0.49) between DPR and all JSDMs except HPR (0.77), and between LPR and the HLR-S (0.57) and HLR-NS (0.63).

The strength of the JSDM correlation estimates is shown in Figures 7a and 7b. We split this comparison between the two largest (birds and butterflies) and four smallest datasets, as sample size imbalances between datasets masked observable trends. Averaged across the four smallest datasets, DPR, HLR-S, and HLR-NS estimated the weakest correlation strengths relative to MPR, yet for the two largest datasets HLR-NS estimates were slightly stronger than MPR on average. HPR and LPR consistently estimated stronger correlations than MPR.

For the smallest four datasets (Figure 8a**)** DPR, HLR-S, and HLR-NS had more certain estimates of correlation coefficients than MPR, HPR was relatively more uncertain, while MPR and LPR were generally equal. However, for the largest two datasets (Figure 8b), HLR-NS was relatively more uncertain than MPR; LPR and MPR were similarly uncertain, and DPR tended to be more certain than MPR.

For results of the simulation study see Appendix S5.

DISCUSSION

Only four JSDMs could be successfully fit to all six datasets, illustrating that different JSDMs scale differently with dataset size. That MLR could only be successfully fit to the two smallest datasets, and was significantly slower than the other JSDMs, highlights the computational benefits of the probit link over the logit link. Since probit models allow a latent variable parameterisation instead of

using the link function directly, they can be fitted using a highly efficient Gibbs sampler (Chib & Greenberg 1998; O'Brien & Dunson 2004). Of the probit models, HPR and HLR-S scaled the worst, and LPR exhibited signs of slowing down for the two largest datasets. The effect of scaling will only become more apparent as datasets become larger.

Estimates of runtime and sampler efficiency show that MPR and HLR-NS are markedly faster than the other JSDMs. HPR took longer to run than DPR, yet this difference is due to different MCMC regimes, since the HPR sampler was more efficient. HPR also exhibited the most consistent ESS across all coefficients. The MCMC sampling of HPR is implemented in JAGS, which prioritises ease of use for those with limited programming experience, rather than more efficient, customised samplers utilised in other JSDMs. HLR-S had the least efficient sampler as it was the only one that included spatial Gaussian process components. This requires inversion of an $n \times n$ covariance matrix per latent factor, which is computationally expensive and scales poorly with dataset size (Rasmussen & Williams 2006).

Datasets were coerced into the required formats for most JSDMs within five minutes, which suggests data formatting is not a barrier to JSDM usage. A vignette can be found in the code repository for this analysis (Wilkinson 2018) that explains data formatting. The HLR models took longer for dataset formatting but formatting time reduced markedly with experience. HLR models are fit in MATLAB not R, the language most ecologists are familiar with, so many ecologists might experience a slight learning curve. MATLAB also requires a software license.

All JSDMs estimated similar regression coefficients across all datasets (>0.7 correlation). The three JSDMs with hierarchically-drawn regression coefficients (HPR, HLR-S, HLR-NS) were strongly correlated (>0.9); however, this does not mean the additional hierarchical structure is necessarily beneficial. Minor differences between the uncertainty in JSDM coefficient estimates were observed but with no clear pattern. We suggest that the JSDMs were as uncertain as each other when considered across all coefficients.

For this comparison we used the default priors for each JSDM, though these will not always be the ideal choice. The fungi and mosquito datasets exhibit complete separation for multiple species in the case of binary environmental variables. For maximum-likelihood estimation methods, a finite maximum likelihood would not exist and standard calculation of standard errors would fail. This is less of an issue for models using a Bayesian framework, as using prior information switches the focus of parameter estimation to posterior distribution summaries (Rainey 2016). LPR had convergence issues for these datasets with default priors, but the use of more informative priors led to performance similar to the other JSDMs (see Appendix S2), highlighting the importance of careful consideration of priors.

Different JSDMs estimated residual correlations between species in the same direction (with some exceptions for near-zero strength interactions whose 95% credible intervals included zero), suggesting that all JSDMs lead to similar ecological inferences. Correlations between each JSDM's estimated correlation coefficients were moderate to very strong when averaged across all datasets. For the smaller datasets, DPR estimated weaker residual correlations than other JSDMs, and showed only weak correlation (~0.3) with the other JSDMs in the two largest datasets. The reason for this is unclear; the Dirichlet process was only implemented for the larger of the two, so it cannot be affecting both datasets.

Whilst the sign of residual correlations was the same across all JSDMs, the estimates differed in their relative strength and uncertainty. For the four smallest datasets, HLR-S and HLR-NS estimated relatively weaker correlations between species than the non-latent factor JSDMs (e.g. MPR and HPR). This is because latent factor models account for correlations through shared responses to a small set of latent factors, which provides a form of regularisation on residual correlations. For the two largest datasets, however, HLR-NS estimated stronger correlations between species than MPR (HLR-S failed to complete within the time limit). LPR, despite being a latent factor model like HLR-S and HLR-NS, estimated stronger correlations between species. This is likely a result of the number of latent factors used to fit the model, as the use of more latent factors provides a better approximation of the full-rank correlation matrix. By default, LPR uses two latent factors, while HLR-S and HLR-NS estimate the number of latent factors during the model fitting process (here between 3 and 12). This further highlights the trade-off between model fit and model simplicity when using the latent factor approach. DPR estimated relatively weaker correlations than the other JSDMs in most cases (similar to HLR-S/HLR-NS in the small datasets); the reason for this is unclear.

Differences in the sizes of residual correlations between the smaller and larger datasets could indicate the effect of spatial scale on the implementation and interpretation of latent factors (Ovaskainen *et al.* 2016b,a). For datasets collected over small spatial scales (coincidently our smaller datasets) HLR-S estimated small α values: 0 (frogs/eucalypts) or 0.04 (mosquitoes), suggesting no benefit of a spatial implementation. Datasets collected at larger spatial scales, like the birds and butterflies (computationally incompatible with HLR-S in this study), would be more likely to exhibit spatial correlation (i.e., a moderate to large α).Indeed, Ovaskainen *et al*. (2016a,b) observed an α between 0.25-0.26 for the butterfly dataset (in which a training/testing split of the data reduced the computational requirements). We suggest the additional effort required for spatially-explicit JSDMs might only be necessary for large spatial scale datasets where spatial correlation could be

reasonably expected, and it becomes important to account for any decay in correlation between latent factors due to distance. Correlations between species at finer spatial scales might reflect species interactions, but at larger scales correlations are likely influenced by other factors like shared environmental responses and dispersal ability (Götzenberger *et al.* 2012; Tikhonov *et al.* 2017). This can be explored by defining latent factors at different spatial scales (Ovaskainen *et al.* 2016a).

In Appendix S5 we compared model performance for a subset of these models on simulated datasets. As for the real datasets included in the main analysis, we found that there were no significant differences in parameter estimation between JSDMs. We observed that models that accounted for hierarchical regression coefficients performed better than the other models for datasets generated hierarchically but only if they also matched the latent factor assumptions. Models also performed worse when fit to datasets that were simulated using unmeasured variables as well as the measured ones used to fit the models, but models that could account for that using latent factors (like HLR-NS) had smaller performance decreases. This shows that the added complexity of these models can provide a benefit to parameter estimation, where the ecological system that generates the data has these features.

Methodological differences between the different JSDMs beyond their core statistical framework might have biased our results. The program in which JSDMs are developed (e.g., R vs MATLAB) could influence computational efficiency, as could their respective MCMC samplers and regimes. The choice of priors and other default settings, such as the number of latent factors, could also impact statistical inference. However, since the goal of this study was to compare the default implementations of these different JSDMs and their underlying statistical frameworks, these differences are unavoidable. Clearly, particular case studies should explore the appropriateness of default settings prior to implementation.

While we only considered presence-absence JSDMs in which species interactions are modelled via multivariate regression methods, other JSDMs have been proposed. *mistnet* (Harris 2015), a neural network JSDM, focuses on predicting community composition but does not provide coefficient estimates comparable to the JSDMs considered here. The JSDM of Johnson & Sinclair (2017) models species interactions by grouping them into guilds by environmental responses rather than via residual correlation. JSDMs for abundance data also exist (e.g.Dorazio, Connor & Askins 2015; Letten *et al.* 2015; Thorson *et al.* 2015; Warton *et al.* 2015).

We have compared the ability of JSDMs to make inferences about species' responses to biotic and abiotic factors. We have used both real and simulated datasets, but a more comprehensive simulation study accounting for all nuances of the model fitting process presents an interesting avenue for further research. The predictive ability of different JSDMs also warrants further study. Single species SDMs are commonly used to predict into un-sampled areas or under new environmental conditions, such as climate change scenarios (Elith & Leathwick 2009), and these are clearly areas where JSDMs are applicable. There is an open question, however, of how best to define prediction in multivariate space. JSDMs can predict community composition (joint prediction), distributions of individual species (marginal prediction), or distributions of individual species conditional on the presence or absence of other species (conditional prediction). Ecological interpretations, and the statistical and computational aspects of these different types of prediction should be considered.

Another important consideration is that all methods described here assume perfect detection (i.e. that absence records reflect true species absence). However, imperfect detection is often an issue in survey data. Extending these models to account for imperfect detection is an important avenue for

future JSDM research (Beissinger *et al.* 2016; Warton *et al.* 2016), as is the evaluation of impacts of imperfect detection in the estimation of residual co-occurrence. Rota *et al* (2016) introduced a multispecies occupancy model accounting for imperfect detection, but in principle, any of the methods above could be extended to include a description of the detection process.

CONCLUSION

Akin to comparisons of single species SDMs (Elith *et al.* 2006), this study serves as a guide for prospective JSDM users to make informed decisions about the JSDM, or JSDMs, that might be relevant to their objective. Expressing the different statistical JSDM frameworks with a common notation has helped to identify their differences and similarities. The core statistical differences between different JSDMs can be broadly defined by two methodological choices: the inclusion of a hierarchical structure on the regression coefficients, and the use of latent factors to account for shared species responses to "hypothetical" unmeasured environmental variables impacting species co-occurrence patterns. There are statistical and computational implications related to these methodological differences, but they also have ecological interpretations. As for all ecological modelling, the user should be aware of all ecological assumptions inherent in these models when making an informed selection of JSDM.

ACKNOWLEDGEMENTS

We thank Peter Vesk, Brendan Wintle, and Jian Yen for insightful discussions, and Els Van Burm and Erica Marshall for early manuscript reviews. We thank Robert Dorazio, Devin Johnson, and one anonymous reviewer for their valuable input. DW is funded by an Australian Government Research Training Program Scholarship. NG is funded by a McKenzie fellowship from the University of Melbourne. GGA and RT are supported by Australian Research Council (ARC) Discovery Early Career

Researcher Awards (DE160100904 and DE170100601).

CONFLICTS OF INTEREST

Some of the authors of this manuscript are authors on the JSDM papers being compared. NG, RT, and MM on Pollock *et al* (2014), and NG on Golding *et al* (2015). No other conflicts of interest are declared.

AUTHORS' CONTRIBUTIONS

All authors conceived the ideas and methodology. DW implemented the analysis. DW led writing the manuscript but all authors contributed significantly throughout and gave final approval before submission.

DATA ACCESSSIBILITY

- Script files to replicate the model running, data extraction, analysis, and plot generation are available in an online repository (Wilkinson 2018).
- The bird dataset is from the USGS Breeding Bird Survey and is available to download from https://www.mbr-pwrc.usgs.gov/bbs/bbs2011.html (Sauer *et al.* 2012). A version used in this analysis can also be found in our online repository (Wilkinson 2018).
- The butterfly dataset is available to download in the supplementary material of Ovaskainen *et al* (2016b). A version used in this analysis can also be found in our online repository (Wilkinson 2018).
- The eucalypt dataset is available to download from Pollock (2014) A version used in this analysis can also be found in our online repository (Wilkinson 2018).
- An anonymised version of the frog dataset is available in our online repository (Wilkinson

2018). This dataset will allow you to replicate our results without the ability to identify the particulars of the dataset.

- The fungi dataset is available in our online repository (Wilkinson 2018)
- The mosquito dataset is available to download from Golding (2015). A version used in this analysis can also be found in our online repository (Wilkinson 2018).

REFERENCE LIST

Araújo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

Beissinger, S.R., Iknayan, K.J., Guillera-Arroita, G., Zipkin, E.F., Dorazio, R.M., Royle, J.A. & Kéry, M. (2016). Incorporating Imperfect Detection into Joint Models of Communities: A response to Warton et al. *Trends in Ecology and Evolution*, **31**, 736–737.

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Clark, J.S., Gelfand, A.E., Woodall, C.W. & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, **24**, 990–999.

Clark, J.S., Nemergut, D., Seyednasrollah, B., Turner, P.J. & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs*, **87**, 34–56.

Dorazio, R.M., Connor, E.F. & Askins, R.A. (2015). Estimating the effects of habitat and biological interactions in an avian community. *PLoS ONE*, **10**, 1–16.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F.,

Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J. & Leathwick, J.J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, …*, **40**, 677–697.

Gelfand, A.E., Silander, J.A., Wu, S., Latimer, A., Lewis, P.O., Rebelo, A.G. & Holder, M. (2006). Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis*, **1**, 41–92.

Golding, N. (2015). Mosquito community data for Golding et al. 2015 (Parasites & Vectors) (Version 1). figshare. doi:10.6084/m9.figshare.1420528.v1. *figshare*.

Golding, N. & Harris, D.J. (2015). BayesComm: Bayesian Community Ecology Analysis.

Golding, N., Nunn, M.A. & Purse, B. V. (2015). Identifying biotic interactions which drive the spatial distribution of a mosquito community. *Parasites & Vectors*, **8**, 367.

Götzenberger, L., de Bello, F., Bråthen, K.A., Davison, J., Dubuis, A., Guisan, A., Lepš, J., Lindborg, R., Moora, M., Pärtel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel, K. & Zobel, M. (2012). Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews*, **87**, 111–127.

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, **40**, 281–295.

Hardy, O.J. (2008). Testing the spatial phylogenetic structure of local communities: Statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology*, **96**, 914–926.

Harris, D.J. (2015). Generating realistic assemblages with a joint species distribution model (D. Warton, Ed.). *Methods in Ecology and Evolution*, **6**, 465–473.

Hui, F.K.C. (2016). <scp>boral</scp> - Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in <scp>r</scp> (T. Poisot, Ed.). *Methods in Ecology and Evolution*, **7**, 744–750.

Hui, F.K.C. (2017). boral: Bayesian Ordination and Regression AnaLysis.

Hutchinson, G.E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

Johnson, D.S. & Sinclair, E.H. (2017). Modeling joint abundance of multiple species using Dirichlet process mixtures. *Environmetrics*, **28**, 1–13.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., Mcinerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.C., Zimmermann, N.E. & O'Hara, R.B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Leathwick, J.R. & Austin, M.P. (2001). Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology*, **82**, 2560–2573.

Letten, A.D., Keith, D.A., Tozer, M.G. & Hui, F.K.C. (2015). Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology*, **103**, 1264–1275.

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

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.C., Thuiller, W., Araújo, M.B., Guisan, A. & Zimmermann, N.E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33**, 1038–1048.

Nieto-Lugilde, D., Maguire, K.C., Blois, J.L., Williams, J.W. & Fitzpatrick, M.C. (2017). Multiresponse algorithms for community-level modeling: review of theory, applications, and comparison to species distribution models. *Methods in Ecology and Evolution*, **12**, 3218–3221.

O'Brien, S.M. & Dunson, D.B. (2004). Bayesian multivariate logistic regression. *Biometrics*, **60**, 739–746.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales (D. Warton, Ed.). *Methods in*

Ecology and Evolution, **7**, 549–555.

Ovaskainen, O., Hottola, J. & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, **91**, 2514–2521.

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

Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E. & Guisan, A. (2010). Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004–1014.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Pollock, L.J. (2014). Data from: Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Dryad Digital Repository*.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & Mccarthy, M.A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, **5**, 397–406.

R Core Team. (2016). R: A Language and Environment for Statistical Computing.

Rainey, C. (2016). Dealing with separation in logistic regression models. *Political Analysis*, **24**, 339–355.

Rasmussen, C.E. & Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*, Secondn. The MIT Press, Cambridge, Massachusetts.

Richmond, O.M.W., Hines, J.E. & Beissinger, S.R. (2010). Two-species occupancy models: a new paramaterization applied to co-occurence of secretive rails. *Ecological Applications*, **20**, 2036–2046.

Rota, C.T., Ferreira, M.A.R., Kays, R.W., Forrester, T.D., Kalies, E.L., McShea, W.J., Parsons, A.W. & Millspaugh, J.J. (2016). A multispecies occupancy model for two or more interacting species. *Methods in Ecology and Evolution*, **7**, 1164–1173.

Sauer, J.R., Hines, J.E., Fallon, J.E., Pardieck, K.L., Ziolkowski, Jr., D.J. & Link, W.A. (2012). The North American Breeding Bird Survey, Results and Analysis 1966 - 2011. Version 12.13.2011.

Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J. & Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, **21**, 88–99.

Taylor-Rodríguez, D., Kaufeld, K., Schliep, E.M., Clark, J.S. & Gelfand, A.E. (2016). Joint Species Distribution Modeling: Dimension Reduction Using Dirichlet Processes. *Bayesian Analysis*, 1–29.

The MathWorks Inc. (2016). MATLAB.

Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C. & Zipkin, E.F. (2016). Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, **25**, 1144–1158.

Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J. & Kristensen, K. (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, **6**, 627–637.

Tikhonov, G., Abrego, N., Dunson, D. & Ovaskainen, O. (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. *Methods in Ecology and Evolution*, **8**, 443–452.

University of Melbourne. (2017). Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility.

Waddle, J.H., Dorazio, R.M., Walls, S.C., Rice, K.G., Beauchamp, J., Schuman, M.J. & Mazzotti, F.J. (2010). A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications*, **20**, 1467–1475.

Warton, D.I., Blanchet, F.G., O'Hara, R., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2016). Extending Joint Models in Community Ecology: A Response to Beissinger et al. *Trends in Ecology and Evolution*, **31**, 737–738.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution*, **30**, 766–779.

Wilkinson, D. (2018). JSDM_Inference v0.1.6. Zenodo, https://dx.doi.org/10.5281/zenodo.1452066.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Wolfram Research Inc. (2016). Mathematica.

TABLES AND FIGURES

Dataset	Source	Geographic Location	Species	Sites	Covariates
Birds	Harris (2015)	North America	370	2752	8
Butterflies	Ovaskainen *et al*. (2016a)	Great Britain	55	2609	4
Eucalypts	Pollock *et al*. (2014)	Grampians National Park, Australia	12	458	7
Frogs	Pollock *et al*. (2014)	Melbourne, Australia	9	104	3
Fungi	Ovaskainen *et al*. (2010)	Southern Finland	11	800	15
Mosquitoes	Golding *et al*. (2015)	South-East England	16	167	13

Table 1: Dataset summary

	Source Paper	R Package	Hierarchical Regression Coefficients	Dirichlet Process	Latent Factors	Multivariate	Probit	Logistic	Regression	Spatially-Explicit	Other Compatible Data Types
MPR	Golding *et al*. (2015)	*BayesComm*				X	X		X		
MLR	Ovaskainen *et al*. (2010)	NA				X		X	X		
HPR	Pollock *et al*. (2014)	NA	X			X	X		X		
LPR	Hui (2016)	*boral*			X	X	X		X		Abundance
DPR	Clark *et al*. (2017)	*gjam*		o	o	X	X		X		Multiple including continuous and discrete abundance, and ordinal counts
HLR-S	Ovaskainen *et al*. (2016a)	NA	X		X	X	X		X	X	
HLR-NS	Ovaskainen *et al*. (2016a)	NA	X		X	X	X		X		

Table 2: Modelling method components. X denotes permanent features, o denotes features dependent on dataset size. Models without an accompanying R package have code scripts available in the supplementary material of their source paper.

Symbol	Definition
Subscripts	
i	Site. $i = 1, \dots, n$
j	Species. $j = 1, \dots, J$
k	Measured covariate. $k = 1, \dots, K$
h	Latent factor / Unmeasured covariate. $h = 1, .., H$
l	Species archetype. $l = 1, \dots, L$
Main Terms	
y	Binary response variable (species presence/absence)
$1(.)$	Indicator function
z	Normally-distributed latent variable
μ	Linear predictor for the measured covariates
ν	Linear predictor for the unmeasured covariates
$\mathbf{X}$	Matrix of measured covariates $(n \cdot K)$
$\boldsymbol{\beta}$	Matrix of regression coefficients $(K \cdot J)$
$\boldsymbol{\eta}$	Matrix of latent factors $(n \cdot H)$
$\boldsymbol{\lambda}$	Matrix of factor loadings $(H \cdot J)$
$\mathbf{I}$	Identity matrix $(J \cdot J)$
$\mathbf{A}$	Archetype-reduced factor loadings matrix $(H \cdot J)$
$\boldsymbol{\Sigma}$	Covariance matrix $(J \cdot J)$
$\mathbf{R}$	Symmetric, positive-definite correlation matrix $(J \cdot J)$
e	Correlated residual error
ε	Uncorrelated residual error
Φ	Cumulative density function of $N(0,1)$
ω	Mean of the normal distribution for hierarchical β coefficients
σ	Standard deviation of the normal distribution for hierarchical β coefficients
d	Spatial distance
α	Positive parameter controlling decay in correlation with distance for latent factors

Table 3:Symbology. Matrix dimensions supplied in brackets

	MPR	MLR	HPR	LPR	DPR	HLR-S	HLR-NS
Birds	X			X	X		X
Butterflies	X			X	X		X
Eucalypts	X	X	X	X	X	X	X
Frogs	X	X	X	X	X	X	X
Fungi	X		X	X	X		X
Mosquitoes	X		X	X	X	X	X

Table 4: Model compatibility with datasets. Pairs marked with X were compatible.

	MPR	MLR	HPR	LPR	DPR	HLR-S	HLR-NS
Birds	3.8	>168	NA	120.4	27.3	>168	15.2
Butterflies	0.23	>168	>168	13.9	6.5	>168	2.1
Eucalypts	<0.02	142.1	7.6	0.33	0.25	50.0	0.21
Frogs	<0.02	14.1	0.94	0.04	0.06	1.4	0.13
Fungi	<0.02	>168	15.8	0.62	0.67	NA	0.26
Mosquitoes	<0.02	>168	6.4	0.14	0.73	2.0	0.2

Table 5: Model runtimes (in hours). >168 corresponds to the cut-off time of seven days.

Figure1: Effective sample size of regression (a) correlation (b)estimates for each JSDM's respective MCMC samplers on the mosquito dataset. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively, and the black dots represent outliers.

Figure 2: Mean regression coefficients (dots) and 95% credible intervals (bars) for damselflies in the mosquito dataset, estimated by different JSDMs.

Figure 3: Correlation of mean regression coefficient estimates between JSDMs, averaged across all datasets.

Figure 4: Absolute value of the Gini coefficient of the variation in the regression coefficient estimates in the eucalypt dataset for all JSDMs.

Figure 5: Estimated correlation coefficients (top row) and the uncertainty of those estimates defined as the width of their 95% credible interval (bottom row) of all JSDMs on the frog dataset.

Figure 6: Average correlation of mean correlation coefficients between JSDMs across all datasets.

Figure 7: Relative correlation strength of all JSDMs compared to a baseline of the MPR model. Plot a)

is for the four smallest datasets, and plot b) is for the two largest. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, and the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively.

Figure 8: Relative correlation uncertainty (95% credible interval width) of all JSDMs compared to a baseline of the MPR model. Plot a) is for the four smallest datasets, and plot b) is for the two largest. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, and the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively. HPR and HLR-S are missing from 8b as they did not complete within the time limit.