Received: 22 May 2020 | Accepted: 12 October 2020

DOI: 10.1111/2041-210X.13518

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Defining and evaluating predictions of joint species distribution models

David P. Wilkinson[1] | Nick Golding[1,2,3] | Gurutzeta Guillera-Arroita[1] | Reid Tingley[4] | Michael A. McCarthy[1]

[1]School of BioSciences, University of Melbourne, Parkville, Vic., Australia
[2]Telethon Kids Institute, Perth Children's Hospital, Nedlands, WA, Australia
[3]Curtin University, Bentley, WA, Australia
[4]School of Biological Sciences, Monash University, Clayton, Vic., Australia

Correspondence
David P. Wilkinson
Email: dwilkinson@unimelb.edu.au

Funding information
Australian Research Council, Grant/Award Number: DE160100904, DE170100601 and DE180100635; Australian Government Research Training Program Scholarship

Handling Editor: Robert Freckleton

Abstract

1. Joint species distribution models (JSDMs) simultaneously model the distributions of multiple species, while accounting for residual co-occurrence patterns. Despite increasing adoption of JSDMs in the literature, the question of how to define and evaluate JSDM predictions has only begun to be explored.
2. We define four different JSDM prediction types that correspond to different aspects of species distribution and community assemblage processes. *Marginal* predictions are environment-only predictions akin to predictions from single-species models; *joint* predictions simultaneously predict entire community assemblages; and *conditional marginal* and *conditional joint* predictions are made at the species or assemblage level, conditional on the known occurrence state of one or more species at a site. We define five different classes of metrics that can be used to evaluate these types of predictions: threshold-dependent, threshold-independent, community dissimilarity, species richness and likelihood metrics.
3. We illustrate different prediction types and evaluation metrics using a case study in which we fit a JSDM to a frog occurrence dataset collected in Melbourne, Australia.
4. Joint species distribution models present opportunities to investigate the facets of species distribution and community assemblage processes that are not possible to explore with single-species models. We show that there are a variety of different metrics available to evaluate JSDM predictions, and that choice of prediction type and evaluation metric should closely match the questions being investigated.

KEYWORDS

biotic interactions, community assemblage, evaluation metrics, joint species distribution models, prediction, species richness

1 | INTRODUCTION

Species distribution models (SDMs) are commonly used to predict un-sampled areas or new environmental conditions (Elith & Leathwick, 2009). Joint species distribution models (JSDMs) are an extension of standard correlative SDMs that allow multiple species to be modelled simultaneously while accounting for species correlations not explained by available environmental predictors, for example, due to species interactions or important missing covariates (Clark et al., 2017; Golding et al., 2015; Kissling et al., 2012; Ovaskainen, Roy, et al., 2016; Pollock et al., 2014). Despite increasing adoption of JSDMs in the literature, it remains unclear how predictions of JSDMs differ from those of standard SDMs, and how specifically JSDM predictions can be used to address different questions in ecology and conservation (Ovaskainen, Abrego, et al., 2016; Zhang et al., 2018; but see: Norberg et al., 2019).

Modelling distributions of single species with correlative SDMs ignores the impacts of species interactions, which potentially biases estimated coefficients and resultant predictions (Kissling et al., 2012; Wisz et al., 2013). There have been attempts to account for biotic interactions by using other species' occurrence states as predictor variables alongside abiotic variables (Araújo & Luoto, 2007; Leathwick & Austin, 2001; Meier et al., 2010; Pellissier et al., 2010), or by constraining predicted distributions to observed or predicted distributions of species on which the target depends (Schweiger et al., 2012), but these approaches are restricted to unidirectional interactions (Kissling et al., 2012) and require that distributions of the non-target species are known a priori or estimated using analogous single-species SDMs. Stacked species distribution models (SSDMs) combine, or stack, multiple single-species SDMs to estimate community structure and species richness (Gelfand et al., 2005; Mateo et al., 2012; Parviainen et al., 2009). Yet, it has been suggested that, because they do not account for species interactions, SSDMs tend to overpredict species richness (Calabrese et al., 2014; Guisan & Rahbek, 2011; Pineda & Lobo, 2009; Thuiller et al., 2015). By accounting for interactions between multiple species, the expectation is that JSDMs might allow for more accurate predictions. The captured interactions might be true biotic interactions, or reflect the effect of relevant missing predictors.

An important decision when aiming to obtain predictions from a JSDM is defining what to predict, based on the JSDM's multivariate output. In a single-species SDM, the focus of prediction is clear: some feature of the target species (presence/absence, abundance, etc). Multivariate JSDMs, however, have multiple response variables (one per species), all of which are potentially correlated. Prediction can therefore be approached in several ways. For instance, we may aim to predict community composition at sites where we have no knowledge of distributions, or, we may be interested in exploring how having data on the distributions of some species changes our prediction for a focal species.

Once we have a prediction, we need to evaluate it. Even ill-fitting models can generate predictions, and poor predictions can hamper species management. More broadly, by evaluating predictions we can use predictive performance to assess which combinations of environmental variables best predict species distributions, compare performance between different modelling approaches and assess the reliability of the predictions we generate (Guisan & Zimmermann, 2000; Lawson et al., 2014). How do we approach this for JSDMs? The substantial literature on the evaluation of SDM predictions (Fielding & Bell, 1997; Lawson et al., 2014; Liu et al., 2009) covers a wide variety of metrics, but are they appropriate in a multi-species context? Are the most common metrics used for single-species SDMs (such as AUC) still the most relevant? JSDMs can predict community assemblages, which opens up a suite of potential evaluation metrics in the form of the dissimilarity indices widely used in community ecology (Legendre & De Cáceres, 2013). What insight can be gained from these additional metrics?

Our article outlines different types of prediction from JSDMs. Marginal predictions are environment-only predictions that average over the occurrence and co-occurrence patterns of other species, and would be the most familiar to users of single-species SDMs. Joint predictions simultaneously predict the occurrence of multiple species while accounting for environmental responses and species correlations. Both prediction types (marginal and joint) can be calculated conditional on the known occurrence state(s) of one or more species in the community. We outline which evaluation methods are appropriate for different prediction types. Finally, we fit a standard JSDM to a frog occurrence dataset from Melbourne, Australia. Our case study is not intended to authoritatively compare JSDM performance, which would require multiple datasets and modelling methods, but rather to provide practical examples of the different types of JSDM predictions and evaluation metrics.

2 | MATERIALS AND METHODS

2.1 | Joint, marginal and conditional probabilities

Probability theory defines the relationships between the joint, conditional and marginal probabilities. Considering the presence/absence of two species at a single site, let A denote the presence of species A and B denote the presence of species B (and conversely A′ and B′ would denote the absence of species A and B respectively). The joint probability of both species being present is:

$$\begin{aligned} Pr(\mathrm{A,B}) &= Pr(\mathrm{A|B}) \cdot Pr(\mathrm{B}) \\ &= Pr(\mathrm{B|A}) \cdot Pr(\mathrm{A}), \end{aligned} \tag{1}$$

where Pr(A|B) is the (conditional) probability that species A is present given species B is present, and Pr(A) is the marginal probability of species A being present.

The two-species scenario can be extended to more species to consider the relationship between the joint probability of occurrence and the probability of presences (or absences) of a subset of the species conditional on the presence (or absence) of the complement of species. In a four-species scenario:

$$\begin{aligned} Pr(\mathrm{A, B, C, D}) &= Pr(\mathrm{A, B|C, D}) \cdot Pr(\mathrm{C, D}) \\ &= Pr(\mathrm{A|B, C, D}) \cdot Pr(\mathrm{B, C, D}). \end{aligned} \tag{2}$$

Here we define Pr(A, B|C, D) as a joint conditional probability (the probability of the joint occurrence of species A and B at the site, conditional on the presence of both species C and D). Pr(A|B, C, D) is a conditional probability (the probability of the occurrence of species A at the site conditional on the presence of species B, C and D).

2.2 | Joint SDMs

Most occurrence-based JSDMs are built on the foundation of the Chib and Greenberg (1998) multivariate probit regression model. In what follows, we focus on the multivariate probit model formulation, though the same principles can be applied to the multivariate logistic and latent factor models that have also been used as JSDMs. For this model, the occurrence state (present or absent) of species *j*, for $j = 1, \ldots, J$, at a site *i*, for $i = 1, \ldots, n$, is y_{ij} and modelled via a normally distributed latent variable, z_{ij}, with y_{ij} equal to 1 when $z_{ij} > 0$, and 0 otherwise (see Figure 1a for a visual representation). This latent variable is not to be confused with the latent variable/latent factor concept of latent factor models (Warton et al., 2015; Wilkinson et al., 2019). The model is as follows:

$$\begin{gathered} y_{ij} = 1\,(z_{ij} > 0) \\ z_{ij} = \mu_{ij} + e_{ij} \\ \mu_{ij} = \boldsymbol{X}_{i,\cdot}\boldsymbol{\beta}_{\cdot,j} \\ \boldsymbol{e}_i \sim \mathrm{MVN}\,(0,\,\mathbf{R}) \end{gathered} \tag{3}$$

where the latent variable, z_{ij}, 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 $\boldsymbol{X}_{i,\cdot}$, and their corresponding regression coefficients $\boldsymbol{\beta}_{\cdot,j}$, as in standard GLMs. 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—the residual correlations between species—are restricted between −1 and 1. The elements of **R** reflect species co-occurrence patterns not described by the environmental predictors (i.e. species interactions or missing predictors). Standard deviations, and in turn variances, are constrained to equal 1 in probit regression, thus covariance and correlation matrices are equivalent. Because the variance of the latent variable distribution remains constant, the probability of presence for a single species is controlled only by the mean value of the distribution, μ (compare Figure 1a,b).

The correlation in e_i makes the space of the latent variables multivariate. Therefore, the whole species community at a site is represented by a multidimensional latent normal random variable with as many dimensions as species. Integration over the relevant portion of the multivariate latent variable space yields the *joint* probability of observing a given configuration of the presence/absence of all species at a site. For instance, in the two-species scenario (Figure 1c), the joint probability Pr(A, B) that both species A and B are present is:

$$Pr(\mathrm{A},\,\mathrm{B}) = Pr\left(z_{\mathrm{A}} > 0,\, z_{\mathrm{B}} > 0\right) = \int_0^{\infty}\int_0^{\infty} f\left(z_{\mathrm{A}},\, z_{\mathrm{B}}\right)\, \mathrm{d}z_{\mathrm{A}}\mathrm{d}z_{\mathrm{B}}, \tag{4}$$

where $f\left(z_{\mathrm{A}},\, z_{\mathrm{B}}\right)$ is the joint probability distribution for the bivariate normal. The double integral calculates the probability in the region where

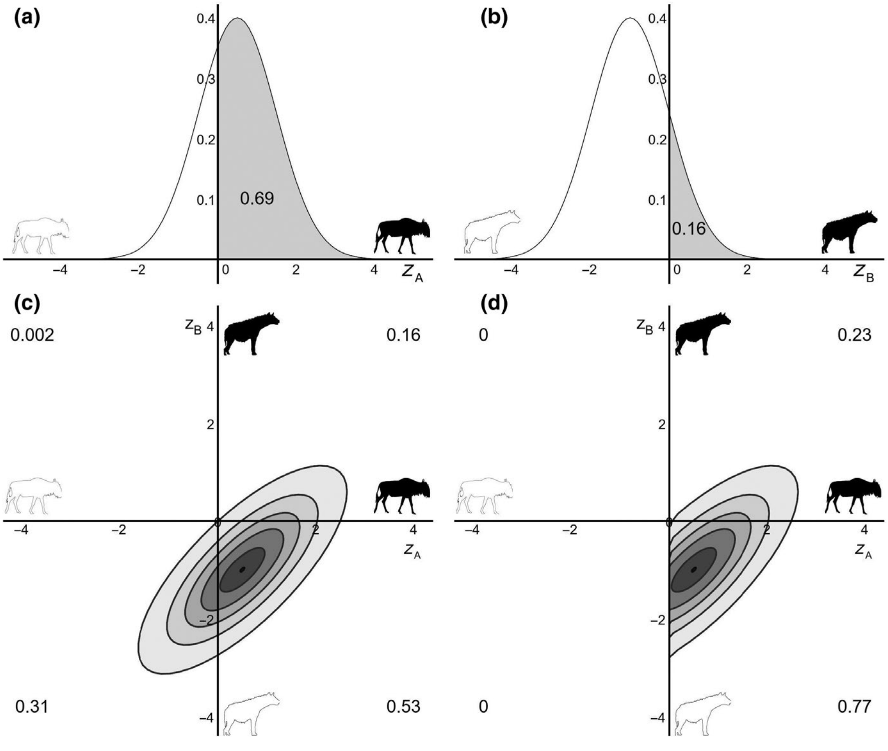

FIGURE 1 Visualisations of different normally distributed latent variables representing species probabilities of presence for two species (A and B). (a) The normally distributed latent variable, z_{A}, with mean, μ_{A}, of 0.5 for species A. Pr(A) is equal to the area under the curve where $z_{\mathrm{A}} > 0$, shown here in grey. (b) The normally distributed latent variable, z_{B}, with mean, μ_{B}, of −1 for species B. Pr(B) is equal to the area under the curve where $z_{\mathrm{B}} > 0$, shown here in grey. (c) The multivariate, normally distributed latent variable, z_{ij}, for a two-species scenario. The mean of the distribution, μ_{ij}, on each species' respective axis is the same as their independent distributions in (a) and (b) and there is positive correlation of 0.75 between them. The contours of the probability distribution, the grey ellipses, indicate probability density values of 0.1, 0.3, 0.5, 0.7 and 0.9. The numbers in the four corners are the probabilities of the multivariate latent variable integrated in that quadrant, for example, in the upper right quadrant; there is a probability of 0.16 that both species will occur at the site. (d) The multivariate, normally distributed latent variable from (c) truncated on the known occurrence state of species A

both z_A and z_B are greater than zero (i.e. both species are present). In this two-species case, this probability can be visualised as the volume under the corresponding region of a three-dimensional surface (as in Figure 1c).

Similar to Equation 1, a joint probability distribution can be written as a function of conditional and marginal density functions; therefore, the joint probability in Equation 4 can be rewritten as follows:

$$\begin{aligned} Pr(\mathrm{A}, \mathrm{B}) &= \int_0^\infty \int_0^\infty f\left(z_\mathrm{A} | z_\mathrm{B}\right) \cdot f\left(z_\mathrm{B}\right) \mathrm{d}z_\mathrm{A} \mathrm{d}z_\mathrm{B} \\ &= \int_0^\infty \int_0^\infty f\left(z_\mathrm{B} | z_\mathrm{A}\right) \cdot f\left(z_\mathrm{A}\right) \mathrm{d}z_\mathrm{B} \mathrm{d}z_\mathrm{A}. \end{aligned} \tag{5}$$

Here the conditional probability distribution $f\left(z_A | z_B\right)$ reflects how likely different values of z_A are (and therefore Pr(A) *given* a fixed value of z_B). The marginal probability distribution $f(z_B)$ reflects how likely different values of z_B are (and therefore Pr(B)) *independent* of the occurrence of species A). The order of items in the joint probability is inconsequential, so species can be marginalised out in any order (lines 1 and 2 in Equation 5 are equivalent).

The marginal probability is obtained by summing the probabilities for all community assemblages in which the species is present, regardless of the presence of the other species; in our two-species scenario Pr(A) = Pr(A|B) + Pr(A|B′). Computing the *conditional* probability of presence of a species given the presence (or absence) of another species involves restricting the probability space to that indicating presence (or absence) of the other species. These probability distributions can be combined in different ways to define a number of prediction types for JSDMs. We next consider four types of JSDM prediction, mapping onto different ecological questions. These are illustrated in Figure 2.

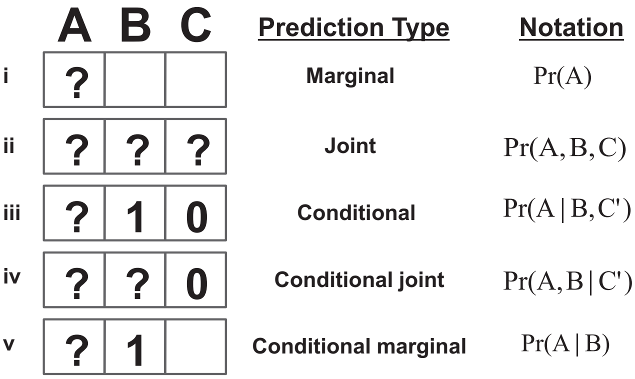

FIGURE 2 Five different prediction types possible with joint species distribution models. The boxes show the occurrence states of species A, B and C. Question marks denote the species being predicted. Empty boxes indicate that the occurrence state for those species is not informing the prediction, while 1/0 denote a known presence/absence state of species which is used to inform prediction. *Prediction Type* is the name of the prediction methodology, and *Notation* shows the corresponding probabilistic notation, following the definitions in Section 2.1

2.3 | Prediction types

2.3.1 | Marginal prediction

Similar to single-species SDM predictions, marginal JSDM predictions are based solely on environmental attributes (they do not consider the presence or absence of the other species, see Figure 2i). This corresponds to predicting, for example, the occurrence of plant species A based solely on its response to environmental variables, such as soil nutrient and water availability, without accounting for co-occurrence patterns with species B–E in the community (e.g. Ovaskainen, Abrego, et al., 2016; Zhang et al., 2018). Marginal JSDM predictions can be calculated using the estimated regression coefficients and the corresponding covariate values at each site, as in a standard GLM. The sole difference between marginal predictions for JSDMs and those of single-species SDMs is that inter-species correlations are accounted for in the estimation of JSDM regression coefficients.

2.3.2 | Joint prediction

We can predict species community composition that accounts for both the environmental covariates *and* species co-occurrence, by using the joint probability distribution. This corresponds to predicting the assemblage of plant species A–E in a community simultaneously, while accounting for their individual responses to environmental conditions (e.g. soil nutrients) and co-occurrence patterns of the species (e.g. Norberg et al., 2019; Ovaskainen, Abrego, et al., 2016).

Whereas the marginal prediction of a community at a given site can be represented by a single vector of probabilities (one for each species), the joint prediction instead yields a probability value for each possible realisation of the community composition (each realisation is a vector of 1s and 0s indicating the presence/absence of each species; Figure 2ii). With J species, the number of possible community assemblages is 2^J. This number increases very quickly with J. While with three species there are eight possible assemblages, with 10 species there are 1,024, and with 20 species over a million. It is therefore generally infeasible to compute and store the probabilities of all possible assemblages when evaluating the predictions. One alternative is to use the model to simulate community assemblages by taking random draws from the joint probability distribution. The frequencies of the simulated assemblages reflect the probability of plausible assemblages. Also, in some cases we are only interested in a subset of the probabilities. For instance, for model evaluation purposes, we only need to compute joint predictions for the species assemblages observed in held-out data.

2.3.3 | Conditional joint prediction

In some cases, we might be interested in estimating a species' occurrence probability given the known occurrence state of other species. If we know the correlation between species (estimated by the JSDM) and the occurrence state of some of those species, we can make a more informed prediction of the unknown occurrence state of the remaining species. This corresponds to predicting plant species A–D in the community simultaneously when we know the occurrence state of species E, and leveraging that knowledge to constrain the possible predicted assemblages (e.g. Taylor-Rodríguez et al., 2017).

For each species with a known occurrence state, we can truncate the multivariate normal distribution over the latent variable in the dimension that represents that species, either to be positive if the species is present, or negative if it is absent (see Figure 1d for a visual representation). Because probability distributions integrate to one, truncating this distribution alters the probabilities of the remaining community assemblage possibilities. For instance, in our example, the marginal probability of species B, Pr(B), is 0.16 (Figure 1b), but if we know that species A is present, then the conditional probability of species B is Pr(B|A) = 0.23 (Figure 1d).

2.3.4 | Conditional marginal prediction

Conditional marginal predictions, which are simultaneously conditional on and marginal to the occurrence states of the other species (Figure 2v), are another type of prediction that can be made with JSDMs. As per conditional joint predictions, we can make more informed predictions by using the known occurrence state of other species; however, this method makes use of the updated marginal distribution of the remaining species. This corresponds to predicting plant species Abased on its response to environmental variables (e.g. soil nutrients), leveraging information on the known occurrence state of species E, and still being independent of the remaining species in the community B–D. This prediction type is also simple to represent; like marginal predictions, it can be stored as a vector of probabilities for each species.

2.4 | Evaluation metrics

To date, the choice of evaluation metric in the multi-species context of JSDMs has not been thoroughly explored. In a single-species context, choice of metric depends largely on factors such as data type (e.g. presence–absence vs. presence-only) and prediction format (binary or probabilistic; Lawson et al., 2014). Here we consider traditional single-species metrics for JSDMs and new metrics that may be suitable. We broadly classify metrics for evaluating JSDM predictions into five groups, in terms of the aspects of performance on which they focus (see Table 1).

Threshold-independent metrics evaluate continuous predicted probabilities against observed presence–absence data. A widely used threshold-independent metric used in single-species SDMs is the area under the receiver operating characteristic curve (AUC), but other metrics include root mean square error (RMSE), the coefficient of determination (R^2) and the Pearson's or point-biserial correlation coefficient.

Threshold-dependent metrics compare binary predictions against observed presence–absence data. This requires simplifying predicted values (which may be probabilities or some other metric) to binary outcomes by considering them presences if they exceed a set threshold value, or absences otherwise. Then, a confusion matrix contrasts observed and predicted occurrence states, and metrics derived from it. Examples include precision, sensitivity and true/false positive/negative rates. The value of thresholding continuous predictions has been debated in the SDM literature (Freeman & Moisen, 2008; Guillera-Arroita et al., 2015; Liu et al., 2005), for prediction and evaluation (Lawson et al., 2014). How to determine the threshold value when binary conversion is wanted is also debated. It is common to set the threshold at an arbitrary value of 0.5 (Freeman & Moisen, 2008), which provides an obvious decision threshold (i.e. the species is more likely to be present than absent) when predictions are calibrated estimates of probability of presence. Another frequent suggestion is to set the threshold to the observed prevalence of the species in question (Hanberry & He, 2013). In the multi-species context of JSDMs, a logical extension of this debate is whether to define community-wide or species-specific thresholds. However, Lawson et al. (2014) showed that by using a probabilistic confusion matrix we can calculate threshold-dependent types of metrics without the need to threshold probabilistic predictions. To avoid any issues with the choice of threshold impacting our analysis, we followed this idea and used the probabilistic confusion matrix approach when calculating our threshold-dependent metrics.

Community dissimilarity indices are widely used in community ecology to quantify the dissimilarity between two realisations of species assemblages. Examples of common metrics are Bray–Curtis dissimilarity and Jaccard distance (Chao et al., 2004). These metrics compare predicted assemblages with observed assemblages and thus probabilistic predictions require thresholding or samples from a binomial distribution before these metrics can be evaluated on them.

Species richness metrics consider the ability of models to predict a single (but widely studied) aspect of community composition—the number of species present. We have evaluated species richness difference, defined as predicted minus observed richness, as it directly compares predictions with observed richness values.

Likelihood metrics assess model fit by computing the probability of observing a given community assemblage, assuming the model is 'correct' in its structure and parameter estimates. For reasons of numerical stability, it is common to work with the log of the likelihood. We use the term independent log-likelihood to represent the typical log-likelihood metric used in SSDMs. This metric assesses each species individually across all sites—computing the probability of observing that species' presence/absence observations—and then combines these into a single metric, assuming the species' distributions to be independent (i.e. the log-likelihoods are summed up). We can also define a joint log-likelihood that assesses all species simultaneously as an assemblage at each site, accounting for the correlation structure encoded in the JSDM formulation.

TABLE 1 Summary of evaluation metrics for joint species distribution model predictions

Name	Applicable to binary predictions	Applicable to probabilistic predictions	Threshold-dependent metric	Threshold-independent metric	Community dissimilarity metric	Species richness metric	Likelihood metric
Accuracy/true skill statistic	X	X	X				
Area under the receiver operating characteristic curve (AUC)		X		X			
Bray–curtis dissimilarity	X	X			X		
Canberra index	X	X			X		
Cohen's Kappa	X		X				
Diagnostic odds ratio	X	X	X				
F_1 score	X	X	X				
False discovery ratio	X	X	X				
False negative rate	X	X	X				
False omission rate	X	X	X				
False positive rate	X	X	X				
Gower index	X	X			X		
Gower index (alternative)	X	X			X		
Jaccard distance	X	X			X		
Kendall rank correlation coefficient		X		X			
Kulczynski index	X	X			X		
Log likelihood—independent		X					X
Log likelihood—joint		X					X
Mountford index	X	X			X		
Mean error (bias)		X		X			
Mean square error		X		X			
Negative likelihood ratio	X	X	X				
Negative predictive performance	X	X	X				
Pearson correlation coefficient		X		X			
Positive likelihood ratio	X	X	X				
Positive predictive performance/ precision	X	X	X				
R^2/coefficient of determination	X	X		X			
Raup-crick dissimilarity	X	X	X				
Root mean square error		X		X			
Spearman rank correlation coefficient		X		X			
Species richness difference	X	X				X	
Sum of squared errors		X		X			
True negative rate/specificity	X	X	X				
True positive rate/sensitivity	X	X	X				
Youden's J statistic	X	X	X				

2.5 | Case study

We illustrate the application of different JSDM prediction types and the appropriate evaluation metric classes with a case study in which we fit a JSDM to a presence–absence dataset of frog species in the Greater Melbourne area of Victoria, Australia (Parris, 2006). The dataset contains nine species, 104 waterbodies (sites) and three measured covariates (area, road density and presence of vertical wall), and was previously analysed in Pollock et al. (2014) and Wilkinson et al. (2019). Species prevalence ranged from 0.02 to 0.52, and were mainly positively correlated with each other except for a single species negatively correlated with the rest. In our analysis, we standardised the two continuous variables and used fivefold random cross validation for model evaluation.

We chose to use a standard multivariate probit regression JSDM (Chib & Greenberg, 1998) implemented with *BayesComm* v0.1–2 (Golding & Harris, 2015) in R v3.5.2 (R Core Team, 2018). The JSDM applies independent normal priors, $\beta_k \sim N(0, 100)$, on the regression coefficients and an inverse Wishart prior, with $n + 2J$ degrees of freedom and scale matrix $\mathbf{I}$, on the correlation coefficients. The model was fit by MCMC using a Gibbs sampler implemented in R and C++. We used a single MCMC chain of 11,000 samples, discarding the first 1,000 as burn-in, to sample the posterior distribution. Model convergence was examined via visual assessment of trace plots for the posterior distributions. Model fitting and predictions were undertaken on The University of Melbourne's Spartan HPC infrastructure (Meade et al., 2017).

Once fit to the training data, the JSDM was used to predict to the held-out test data from fivefold cross validation. We calculated four prediction types: marginal, conditional marginal, joint and conditional joint. The two conditional prediction types were undertaken under the assumption that we knew the occurrence state of one randomly selected 'high prevalence' species in the community (*Litoria ewingii*). Conditional marginal prediction returned probabilistic predictions for the remaining eight species and conditional joint prediction generated plausible assemblages of the remaining eight species.

While we have defined 35 metrics that can be calculated to assess different aspects of JSDM predictions, we present only a subset in this analysis for illustrative purposes. More detail on the metrics, including how they are calculated, which prediction types they are appropriate for, and how to interpret them can be found in Appendix S1. Threshold-dependent and threshold-independent metrics are calculated on a per-species basis, whereas community dissimilarity and species richness metrics are calculated per-site. Most metrics can theoretically be calculated for either species or sites (as they are just comparing two binary vectors: observations and predictions) so this split is based on how they are historically used in ecological literature. Metrics were evaluated once for each of the 1,000 posterior samples drawn.

3 | RESULTS

The patterns of predicted probabilities of presence for each species obtained from the marginal and conditional marginal predictions both broadly matched the community assemblage observed in the held-out testing data (Figure 3). The AUC for most species using the marginal predictions was typically high (>0.88) with a standard deviation of <0.05 across all posterior samples. Exceptions were *L. ewingii* and *Litoria peronii*, which were predicted approximately randomly with an AUC = ~0.5, and *Litoria raniformis*, which had no recorded presences in the held-out data and was thus incompatible with AUC calculations. Marginal predictions conditioned on the known occurrence state of *L. ewingii* returned higher AUC values by a mean of 0.03 (Figure 3). The largest AUC gain was 0.08 for *Litoria verreauxii*. For most other species-level evaluation metrics, we found the conditional marginal prediction to outperform the marginal prediction, but in most cases the differences were relatively minor. We

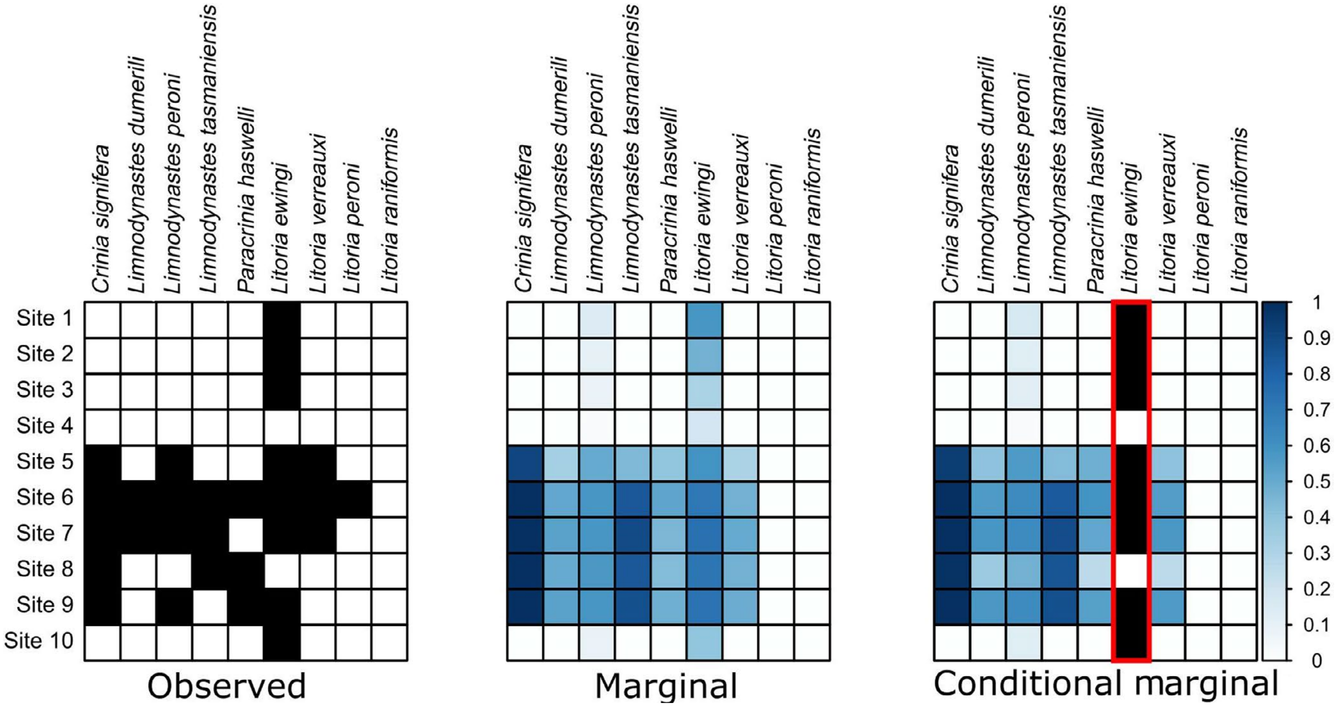

FIGURE 3 Species-level predictions for the frog community at a subset of 10 sites in the testing dataset. The left-hand plot depicts the known occurrence state of the species in the testing dataset (black = present and white = absent). The middle plot depicts the median probability of presence from the posterior distributions of the marginal predictions. The right-hand plot depicts the median probability of presence from the posterior distributions of the conditional marginal predictions. These predictions are conditional on the known occurrence state of *L. ewingii* which are highlighted with the red border

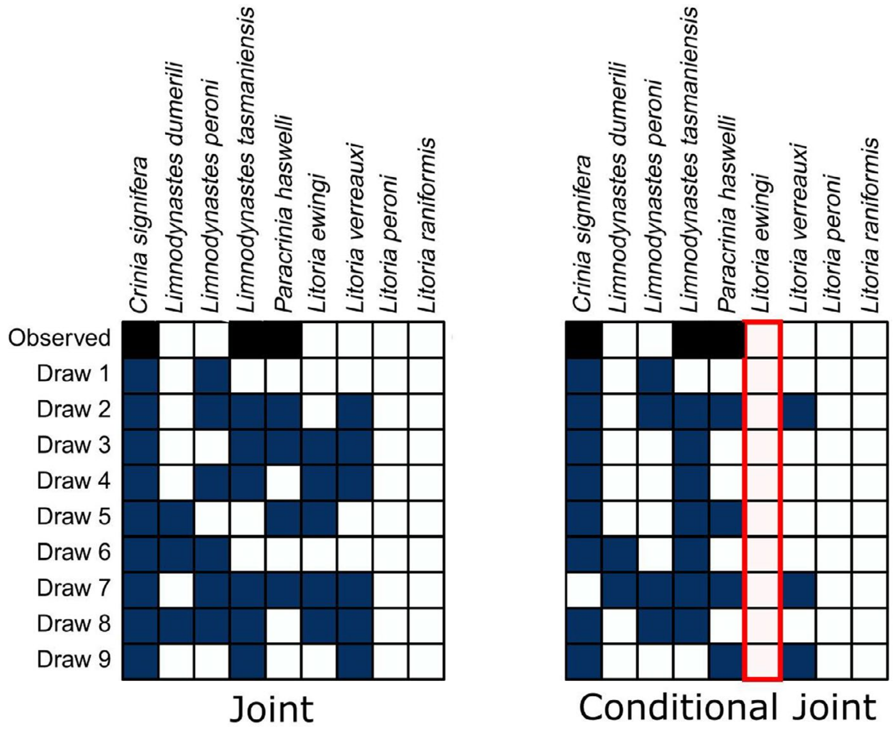

FIGURE 4 Binary community-level predictions of the frog community at a single site in the testing dataset. Community-level predictions here are random draws from the multivariate normal distribution representing plausible community assemblages under the environmental conditions at the site (blue = present and white = absent). The left-hand plot depicts draws of plausible community assemblages from joint predictions. The right-hand plot depicts draws of plausible community assemblages from joint predictions, conditional on the known occurrence state of *Litoria ewingii* at the site (highlighted with the red border). The top row of both plots shows the known community assemblage in the testing dataset. The site considered in this figure corresponds to site 8 in Figure 3

did observe some exceptions to this, such as an increased R^2 value for *L. verreauxii* of 0.26.

The conditional marginal predictions outperformed the marginal predictions for the majority of community-level metrics. For most community dissimilarity metrics, the conditional marginal predictions returned dissimilarity values between 0.05 and 0.1 lower than the marginal predictions which indicates a better estimate of community assemblages. Jaccard distance and Gower Index improved by 0.09, and Bray–Curtis dissimilarity improved by 0.06.

The community-level approach of joint predictions yields probabilities of specific assemblages occurring at a site or, as we present here, one can take random draws from the posterior distribution to generate plausible community assemblages at a site. In Figure 4, we show random draws from the joint and conditional joint prediction types for a single site in our held-out data. While joint predictions generally overpredicted the number of species found in the assemblage overall, conditioning on the known occurrence state (absence) of *L. ewingii* led to draws that were closer to the observed assemblage.

Community dissimilarity metrics were generally between 0.02 and 0.05 higher for conditional joint predictions relative to standard joint predictions. Raup-Crick dissimilarity was the exception. Conditional joint predictions had values that were, on average, 0.25 lower; however, with a standard deviation of 0.37 across all predictions, this metric is possible struggling to calculate correctly across all draws. For the site considered in Figure 4, we observed improvements of between 0.08 and 0.1 for the Bray–Curtis dissimilarity, Jaccard distance, Gower index and Canberra index.

As shown in Figure 4, in our dataset, both the joint and conditional joint prediction types overestimated species richness. The joint prediction type had a mean species richness difference estimate of 0.95 species in the held-out data, while the conditional joint prediction type had a mean estimate of 0.7 species per site. By conditioning the community-level prediction on the known occurrence state of a species, we are able to obtain a prediction closer to the observed data.

4 | DISCUSSION

We have reviewed and clarified the ways in which predictions of species distributions can be approached with JSDMs. Our review highlights the additional functionality that JSDMs enable relative to simple stacking of single-species models. JSDMs have the following two main advantages: an ability to partition the effect of measured variables and residual correlations between species, and predictions of community assemblages that account for these correlations. By partitioning the effect of measured variables from residual correlations, which may include species interactions, JSDMs potentially enable more accurate estimates of environmental drivers of species' distributions. In addition, by leveraging the information in the residual correlations, JSDMs can potentially better predict community assemblages.

The different prediction types defined here for JSDMs correspond to different ecological aims. Marginal predictions correspond to the traditional single-species predictions; joint predictions predict entire assemblages; while the conditional and conditional marginal predictions let us inform these predictions with additional information such as easy-to-detect indicator species. Different aims also imply different evaluation metrics, so each practitioner should determine an appropriate prediction type and evaluation metric, or metrics, for their ecological question.

We have shown that conditioning on the known occurrence status of species can improve both species- and community-level predictions by exploiting the correlations estimated in JSDMs

(Harris, 2015). The extent of this improvement is likely dependant on the strength of the correlation between the known species and other species in the community. We can expect that a species that is highly correlated with the known species (positively or negatively) would benefit more than a species that occurs more or less independent of the known species. Therefore, benefits of JSDM predictions may be observed for only some species in the community, or to differing extents between species.

The JSDM tended to overpredict species richness for all prediction types that account for correlations between species. For our case study of nine frog species, the JSDM predicted approximately one extra species per site. Zurell et al. (2019) similarly found that a JSDM overpredicted species richness compared to a SSDM. As the JSDM only overpredicted species richness for prediction types that account for residual correlations, the estimated correlations could potentially explain why the JSDM tended to overpredict. A largely positive correlation matrix between most species combined with high marginal occurrence probabilities, as seen here, could cause the JSDM to predict likely assemblages in excess of restrictions, such as site carrying capacities and/or dispersal limitations.

We have presented JSDM prediction methods using only a single dataset for illustrative purposes. A more in-depth study using a wider array of datasets is required, as dataset properties likely influence the relative merits of the different prediction types. For datasets in which all species are reasonably prevalent, the expected performance difference between prediction types would be smaller than in scenarios of relatively sparse data, in which the additional information in joint or conditional prediction types could outperform the other prediction types. The value of additional information in the form of species co-occurrence will also be greater when correlations between species are stronger; the known occurrence of a species will provide minimal benefit if it exhibits no correlation with the other species in a dataset. Future studies incorporating multiple case studies could also usefully evaluate different prediction types when making versus extrapolative predictions, as our case study only focused on the former type of evaluation and the latter has been shown to be more difficult (Norberg et al., 2019).

A small body of literature focuses on large-scale comparisons of JSDMs and single-species models. These studies have included many species and model types (Norberg et al., 2019; Zurell et al., 2019) but have not addressed all of the prediction types available to JSDMs. Norberg et al. (2019) used joint predictions but differently than presented here. First, they generated binary assemblage predictions but took the average of a large number of random draws to generate species-specific probabilities, which is an approximation of the marginal distribution. Second, they used joint predictions to generate species richness metrics equivalent to ours. Third, they used joint predictions and community dissimilarity metrics to test predictions of species turnover between sites rather than to test community composition at a site. To our knowledge, no JSDM studies have yet included conditional or conditional marginal predictions, although they have been identified as desirable avenues of research (Blanchet et al., 2020; Norberg et al., 2019; Zurell et al., 2019). Evaluations with a broader array of JSDM implementations, prediction types and datasets are warranted. Both Norberg et al. (2019) and Zurell et al. (2019) included latent factor JSDMs in their comparisons, whereas we use a multivariate probit model. Norberg et al. (2019) found that the HMSC JSDM (Ovaskainen, Roy, et al., 2016) outperformed both other JSDMs and SSDMs, while Zurell et al. (2019) found that the *boral* JSDM (Hui, 2016) had similar results to those described here. Zurell et al. (2019) suggested that the poor performance of *boral* was a result of how the latent factor model is used when extrapolating predictions. Prediction using latent factor JSDMs can be performed in two ways: (a) marginalising over the latent factors by assigning the mean value of modelled sites to prediction sites when performing regression-style predictions and (b) defining the latent factor models in the same way as multivariate probit models (see Wilkinson et al., 2019 for notation), which lets one use the prediction types defined in this paper. This second method may be a better alternative to latent factor model prediction than marginalising over the latent factors.

5 | CONCLUSIONS

Joint species distribution models enable a variety of different ways to predict species distributions and community assemblages. Here we have defined environment-only marginal predictions, joint predictions for whole community assemblages, and conditional marginal and conditional joint predictions that can also leverage additional information on known species' occurrences. Previous studies have either not considered prediction with JSDMs, focused on marginal predictions, or considered only limited aspects of joint prediction. We have also shown that there are several classes of evaluation metrics that can be applied to subsets of these predictions. Ecologists seeking to use these prediction methods and evaluation metrics should consider which method and metric are most closely linked to the ecological question they are investigating. A larger comparison of prediction types that considers different JSDM implementations and multiple datasets is required to evaluate general performance of these models and prediction types.

ACKNOWLEDGEMENTS

We thank Peter Vesk, Brendan Wintle and Jian Yen for insightful discussions, and Els Van Burm and Kristina Chyn for manuscript reviews. D.P.W. is funded by an Australian Government Research Training Program Scholarship. N.G., G.G.-A. and R.T. are supported by Australian Research Council (ARC) Discovery Early Career Researcher Awards (DE180100635, DE160100904 and DE170100601).

CONFLICT OF INTEREST

No conflict of interest is declared.

AUTHORS' CONTRIBUTIONS

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

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13518.

DATA AVAILABILITY STATEMENT

Script files to replicate the model running, data extraction, analysis and plot generation are available in an online repository (Wilkinson, 2019). An anonymised version of the frog dataset is available in our online repository (Wilkinson, 2019). This dataset will allow you to replicate our results without the ability to identify the particulars of the dataset.

ORCID

David P. Wilkinson https://orcid.org/0000-0002-9560-6499
Nick Golding https://orcid.org/0000-0001-8916-5570
Gurutzeta Guillera-Arroita https://orcid.org/0000-0002-8387-5739
Reid Tingley https://orcid.org/0000-0002-7630-7434
Michael A. McCarthy https://orcid.org/0000-0003-1039-7980

REFERENCES

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. https://doi.org/10.1111/j.1466-8238.2007.00359.x

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, *23*, 1050–1063. https://doi.org/10.1111/ele.13525

Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, *23*, 99–112. https://doi.org/10.1111/geb.12102

Chao, A., Chazdon, R. L., Colwell, R. K., & Shen, T.-J. (2004). A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters*, *8*, 148–159. https://doi.org/10.1111/j.1461-0248.2004.00707.x

Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*, 347–361. https://doi.org/10.1093/biomet/85.2.347

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. https://doi.org/10.1002/ecm.1241

Elith, J., & Leathwick, J. J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*, 38–49. https://doi.org/10.1017/S0376892997000088

Freeman, E. A., & Moisen, G. G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, *217*, 48–58. https://doi.org/10.1016/j.ecolmodel.2008.05.015

Gelfand, A. E., Schmidt, A. M., Wu, S., Silander, J. A., Latimer, A., & Rebelo, A. G. (2005). Modelling species diversity through species level hierarchical modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*, 1–20.

Golding, N., & Harris, D. J. (2015). *BayesComm: Bayesian community ecology analysis*. R package version 0.1-2. Retrieved from https://CRAN.Rproject.org/package=BayesComm

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. https://doi.org/10.1186/s13071-015-0915-1

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., Mccarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, *24*, 276–292. https://doi.org/10.1111/geb.12268

Guisan, A., & Rahbek, C. (2011). SESAM – A new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, *38*, 1433–1444. https://doi.org/10.1111/j.1365-2699.2011.02550.x

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Hanberry, B. B., & He, H. S. (2013). Prevalence, statistical thresholds, and accuracy assessment for species distribution models. *Web Ecology*, *13*, 13–19. https://doi.org/10.5194/we-13-13-2013

Harris, D. J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution*, *6*, 465–473. https://doi.org/10.1111/2041-210X.12332

Hui, F. K. C. (2016). boral – Bayesian ordination and regression analysis of multivariate abundance data in R. *Methods in Ecology and Evolution*, *7*, 744–750.

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. https://doi.org/10.1111/j.1365-2699.2011.02663.x

Lawson, C. R., Hodgson, J. A., Wilson, R. J., & Richards, S. A. (2014). Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, *5*, 54–64. https://doi.org/10.1111/2041-210X.12123

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

Legendre, P., & De Cáceres, M. (2013). Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, *16*, 951–963.

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, *28*, 385–393.

Liu, C., White, M., & Newell, G. (2009). Measuring the accuracy of species distribution models: A review. In *Proceedings 18th World IMACs/MODSIM Congress*, Cairns, Australia, Vol. 4241, 4247 pp.

Mateo, R. G., Felicísimo, Á. M., Pottier, J., Guisan, A., & Muñoz, J. (2012). Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS ONE*, *7*, e32586.

Meade, B., Lafayette, L., Sauter, G., & Tosello, D. (2017). Spartan HPC-cloud hybrid: Delivering performance and flexibility. *University of Melbourne*, *10*, 49.

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.

Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., … Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, *89*, 1–24.

Ovaskainen, O., Abrego, N., Halme, P., & Dunson, D. (2016). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, *7*, 549–555.

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

Parris, K. M. (2006). Urban amphibian assemblages as metacommunities. *Journal of Animal Ecology*, *75*, 757–764. https://doi.org/10.1111/j.1365-2656.2006.01096.x

Parviainen, M., Marmion, M., Luoto, M., Thuiller, W., & Heikkinen, R. K. (2009). Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biological Conservation*, *142*, 2501–2509. https://doi.org/10.1016/j.biocon.2009.05.030

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. https://doi.org/10.1111/j.1600-0587.2010.06386.x

Pineda, E., & Lobo, J. M. (2009). Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, *78*, 182–190. https://doi.org/10.1111/j.1365-2656.2008.01471.x

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) (J. McPherson, Ed.). *Methods in Ecology and Evolution*, *5*, 397–406.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

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. https://doi.org/10.1111/j.1466-8238.2010.00607.x

Taylor-Rodríguez, D., Kaufeld, K., Schliep, E. M., Clark, J. S., & Gelfand, A. E. (2017). Joint species distribution modeling: Dimension reduction using dirichlet processes. *Bayesian Analysis*, *12*, 939–967. https://doi.org/10.1214/16-BA1031

Thuiller, W., Pollock, L. J., Gueguen, M., & Münkemüller, T. (2015). From species distributions to meta-communities. *Ecology Letters*, *18*, 1321–1328. https://doi.org/10.1111/ele.12526

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 & Evolution*, *30*, 766–779. https://doi.org/10.1016/j.tree.2015.09.007

Wilkinson, D. P. (2019). JSDM_Prediction v0.1.0. *Zenodo*, https://doi.org/10.5281/zenodo.3514766

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*, 198–211. https://doi.org/10.1111/2041-210X.13106

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., … 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. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Zhang, C., Chen, Y., Xu, B., Xue, Y., & Ren, Y. (2018). Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography*, *41*, 1876–1887. https://doi.org/10.1111/ecog.03571

Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R. O. (2019). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, 1–13.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wilkinson DP, Golding N, Guillera-Arroita G, Tingley R, McCarthy MA. Defining and evaluating predictions of joint species distribution models. *Methods Ecol Evol*. 2021;12:394–404. https://doi.org/10.1111/2041-210X.13518