1	
2	MR DAVID PETER WILKINSON (Orcid ID: 0000-0002-9560-6499)
3	DR NICK GOLDING (Orcid ID : 0000-0001-8916-5570)
4	DR GURUTZETA GUILLERA-ARROITA (Orcid ID : 0000-0002-8387-5739)
5	
6	
7	Article type : Research Article
8	S
9	
10	Handling editor: Professor Robert Freckleton
11	
12	Running title: Predictions for joint species distribution models
13	Title: Defining and evaluating predictions of joint species distribution models
14	David P. Wilkinson <sup>1*</sup> , Nick Golding <sup>1,2,3</sup> , Gurutzeta Guillera-Arroita <sup>1</sup> , Reid Tingley <sup>4</sup> , Michael
15	A. McCarthy <sup>1</sup>
16	
17	1. School of BioSciences, University of Melbourne, Parkville, 3010, Victoria, Australia
18	2. Telethon Kids Institute, Perth Children's Hospital, 15 Hospital Ave, Nedlands, 6009,
19	Western Australia, Australia
20	3. Curtin University, Kent St, Bentley, 6102, Western Australia, Australia
21	4. School of Biological Sciences, Monash University, Clayton, 3800, Victoria, Australia
22	*Corresponding author: dwilkinson@unimelb.edu.au

This is the author manuscript accepted for publication and has 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 <u>Version of Record</u>. Please cite this article as <u>doi: 10.1111/2041-210X.13518</u>

23

- 24 Number of words: 6744
- 25 Number of tables: 1
- 26 Number of figures: 4
- 27 Number of references: 45

- 28
- 29
- 30 Abstract:
- 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 35 species distribution and community assemblage processes. Marginal predictions are 36 37 environment-only predictions akin to predictions from single-species models; joint predictions simultaneously predict entire community assemblages; and *conditional* 38 39 marginal and conditional joint predictions are made at the species- or assemblage-40 level, conditional on the known occurrence state of one or more species at a site. We 41 define five different classes of metrics that can be used to evaluate these types of 42 predictions: threshold-dependent, threshold-independent, community dissimilarity, 43 species richness, and likelihood metrics.
- 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. JSDMs present opportunities to investigate facets of species distribution and
   47 community assemblage processes that are not possible to explore with single-species
   48 models. We show that there are a variety of different metrics available to evaluate
   49 JSDM predictions, and that choice of prediction type and evaluation metric should
   50 closely match the questions being investigated.

51 Keywords: joint species distribution models, prediction, evaluation metrics, biotic
52 interactions, community assemblage, species richness

#### 53

#### 54

#### 1. Introduction

Species distribution models (SDMs) are commonly used to predict into un-sampled areas or 55 56 to new environmental conditions (Elith & Leathwick 2009). Joint species distribution models 57 (JSDMs) are an extension of standard correlative SDMs that allow multiple species to be 58 modelled simultaneously while accounting for species correlations not explained by available 59 environmental predictors, e.g. due to species interactions or important missing covariates 60 (Kissling et al. 2012; Pollock et al. 2014; Golding, Nunn & Purse 2015; Ovaskainen et al. 2016b; Clark et al. 2017). Despite increasing adoption of JSDMs in the literature, it remains 61 unclear how predictions of JSDMs differ from those of standard SDMs, and how specifically 62 63 JSDM predictions can be used to address different questions in ecology and conservation 64 (Ovaskainen et al. 2016a; Zhang et al. 2018; but see: Norberg et al. 2019).

65 Modelling distributions of single species with correlative SDMs ignores the impacts of 66 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 67 interactions by using other species' occurrence states as predictor variables alongside abiotic 68 69 variables (Leathwick & Austin 2001; Araújo & Luoto 2007; Meier et al. 2010; Pellissier et 70 al. 2010), or by constraining predicted distributions to observed or predicted distributions of 71 species on which the target depends (Schweiger et al. 2012), but these approaches are 72 restricted to unidirectional interactions (Kissling et al. 2012) and require that distributions of 73 the non-target species are known *a priori* or estimated using analogous single-species SDMs. 74 Stacked species distribution models (SSDMs) combine, or stack, multiple single-species 75 SDMs to estimate community structure and species richness (Gelfand et al. 2005; Parviainen 76 et al. 2009; Mateo et al. 2012). Yet, it has been suggested that, because they do not account 77 for species interactions, SSDMs tend to over-predict species richness (Pineda & Lobo 2009; 78 Guisan & Rahbek 2011; Calabrese et al. 2014; Thuiller et al. 2015). By accounting for 79 interactions between multiple species, the expectation is that JSDMs might allow for more 80 accurate predictions. The captured interactions might be true biotic interactions, or reflect the 81 effect of relevant missing predictors.

82 An important decision when aiming to obtain predictions from a JSDM is defining what to 83 predict, based on the JSDM's multivariate output. In a single-species SDM, the focus of 84 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.

90 Once we have a prediction, we need to evaluate it. Even ill-fitting models can generate 91 predictions, and poor predictions can hamper species management. More broadly, by 92 evaluating predictions we can use predictive performance to assess which combinations of 93 environmental variables best predict species distributions, compare performance between different modelling approaches, and assess the reliability of the predictions we generate 94 95 (Guisan & Zimmermann 2000; Lawson et al. 2014). How do we approach this for JSDMs? 96 The substantial literature on the evaluation of SDM predictions (Fielding & Bell 1997; Liu, 97 White & Newell 2009; Lawson et al. 2014) covers a wide variety of metrics, but are they 98 appropriate in a multi-species context? Are the most common metrics used for single-species 99 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 100 101 widely used in community ecology (Legendre & De Cáceres 2013). What insight can be 102 gained from these additional metrics?

103 Our article outlines different types of prediction from JSDMs. Marginal predictions are 104 environment-only predictions that average over the occurrence and co-occurrence patterns of 105 other species, and would be the most familiar to users of single-species SDMs. Joint 106 predictions simultaneously predict the occurrence of multiple species while accounting for 107 environmental responses and species correlations. Both prediction types (marginal and joint) 108 can be calculated conditional on the known occurrence state(s) of one or more species in the 109 community. We outline which evaluation methods are appropriate for different prediction 110 types. Finally, we fit a standard JSDM to a frog occurrence dataset from Melbourne, 111 Australia. Our case study is not intended to authoritatively compare JSDM performance, 112 which would require multiple datasets and modelling methods, but rather to provide practical 113 examples of the different types of JSDM predictions and evaluation metrics.

114 **2.** Materials and Methods:

115 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 presence of species A and B denote 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:

$$Pr(A,B) = Pr(A|B) \cdot Pr(B)$$

$$= Pr(B|A) \cdot Pr(A)$$
(1)

121 where Pr(A|B) is the (conditional) probability that species A is present given species B is 122 present, and Pr(A) is the marginal probability of species A being present.

123 The two-species scenario can be extended to more species to consider the relationship 124 between the joint probability of occurrence and the probability of presences (or absences) of a 125 subset of the species conditional on the presence (or absence) of the complement of species. 126 In a four-species scenario:

$$Pr(A,B,C,D) = Pr(A,B|C,D) \cdot Pr(C,D)$$
(2)  
= Pr(A|B,C,D) \cdot Pr(B,C,D)

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).

132 Most occurrence-based JSDMs are built on the foundation of the Chib and Greenberg (1998) 133 multivariate probit regression model. In what follows, we focus on the multivariate probit 134 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 135 136 (present or absent) of species j, for j = 1,...,J, at a site i, for i = 1,...,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 137 138 (see Figure 1a for a visual representation). This latent variable is not to be confused with the 139 latent variable/latent factor concept of latent factor models (Warton et al. 2015; Wilkinson et 140 al. 2019). The model is as follows:

141

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

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

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

$$e_{i} \sim MVN(\mathbf{0},\mathbf{R})$$
(3)

where the latent variable,  $z_{ij}$ , is the sum of the linear predictor,  $\mu_{ij}$ , and the correlated residual 143 error,  $e_{ij}$ . The linear predictor is the product of the measured environmental variables  $X_{i,j}$ , and 144 their corresponding regression coefficients  $\beta_{,j}$ , as in standard generalised linear models. 145 146 Correlations in the residual error  $e_i$  are captured in **R**, a symmetric and positive-definite matrix; its diagonal elements are 1 and its off-diagonal elements - the residual correlations 147 between species - are restricted between -1 and 1. The elements of R reflect species co-148 149 occurrence patterns not described by the environmental predictors (i.e. species interactions, or 150 missing predictors). Standard deviations, and in turn variances, are constrained to equal 1 in 151 probit regression, thus covariance and correlation matrices are equivalent. Because the 152 variance of the latent variable distribution remains constant, the probability of presence for a 153 single species is controlled only by the mean value of the distribution,  $\mu$  (compare Figures 1a 154 and 1b).

Author



155

156 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 157 158 variable,  $z_A$ , with mean,  $\mu_A$ , of 0.5 for species A. Pr(A) is equal to the area under the curve where  $z_A > 0$ , shown here in grey. b) The normally distributed latent variable,  $z_B$ , with mean, 159  $\mu_B$ , of -1 for species B. Pr(B) is equal to the area under the curve where  $z_B > 0$ , shown here 160 161 in grey. c) The multivariate, normally-distributed latent variable,  $z_{ij}$ , for a two species scenario. The mean of the distribution,  $\mu_{ij}$ , on each species' respective axis is the same as 162 163 their independent distributions in a) and b) and there is positive correlation of 0.75 between 164 them. The contours of the probability distribution, the grey ellipses, indicate probability 165 density values of 0.1, 0.3, 0.5, 0.7, and 0.9. The numbers in the four corners are the 166 probabilities of the multivariate latent variable integrated in that quadrant, e.g. in the upper 167 right quadrant; there is a probability of 0.16 that both species will occur at the site. d) The 168 multivariate, normally-distributed latent variable from c) truncated on the known occurrence 169 state of species A.

2041210x, 2021, 3, Downloaded from https://besjoumals

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 multi-dimensional 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 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:

$$= \Pr(z_A > 0, z_B > 0) = \int_0^\infty \int_0^\infty f(z_A, z_B) dz_A dz_{B,sdfsd}$$
(4)

176 Where  $f(z_A, z_B)$  is the joint probability distribution for the bivariate normal. The double 177 integral calculates the probability in the region where both  $z_A$  and  $z_B$  are greater than zero 178 (i.e. both species are present). In this two-species case, this probability can be visualised as 179 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:

183

$$Pr(A,B) = \int_{0}^{\infty} \int_{0}^{\infty} f(z_A | z_B) \cdot f(z_B) dz_A dz_B$$

$$= \int_{0}^{\infty} \int_{0}^{\infty} f(z_B | z_A) \cdot f(z_A) dz_B dz_A$$
(5)

Here the conditional probability distribution  $f(z_A|z_B)$  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 JSDMs. 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.

206

#### 207 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 et al 2016a, 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 generalised linear model. 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.

218 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., Ovaskainen *et al* 2016a, Norberg *et al* 2019).

225 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 226 probability value for each possible realisation of the community composition (each 227 228 realisation is a vector of 1s and 0s indicating presence/absence of each species; Figure 2ii). With J species, the number of possible community assemblages is  $2^{J}$ . This number increases 229 230 very quickly with J. Whilst with three species there are eight possible assemblages, with ten 231 species there are 1024, and with twenty species over a million. It is therefore generally 232 infeasible to compute and store the probabilities of all possible assemblages when evaluating 233 the predictions. One alternative is to use the model to simulate community assemblages by 234 taking random draws from the joint probability distribution. The frequencies of the simulated 235 assemblages reflect the probability of plausible assemblages. Also, in some cases we are only 236 interested in a subset of the probabilities. For instance, for model evaluation purposes, we 237 only need to compute joint predictions for the species assemblages observed in held-out data.

238

239

#### 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 243 more informed prediction of the unknown occurrence state of the remaining species. This 244 corresponds to predicting plant species A-D in the community simultaneously when we know 245 the occurrence state of species E, and leveraging that knowledge to constrain the possible 246 predicted assemblages (e.g., Taylor-Rodriguez *et al* (2017)).

247 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 248 249 positive if the species is present, or negative if it is absent (see Figure 1d for a visual 250 representation). Because probability distributions integrate to one, truncating this distribution 251 alters the probabilities of the remaining community assemblage possibilities. For instance, in 252 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  $Pr(B \mid A)$ 253 254 = 0.23 (Figure 1d).

#### 255 2.3.4. Conditional marginal prediction

256 Conditional marginal predictions, which are simultaneously conditional on and marginal to 257 the occurrence states of the other species (Figure 2v), are another type of prediction that can 258 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 259 260 makes use of the updated marginal distribution of the remaining species. This corresponds to 261 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 262 263 independent of the remaining species in the community B-D. This prediction type is also 264 simple to represent; like marginal predictions, it can be stored as a vector of probabilities for each species. 265

266

#### 267 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 theyfocus (see Table 1).

Threshold-independent metrics evaluate continuous predicted probabilities against observed presence-absence data. A widely used threshold-independent metric used in singlespecies 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.

280 Threshold-dependent metrics compare binary predictions against observed presence-281 absence data. This requires simplifying predicted values (which may be probabilities or some 282 other metric) to binary outcomes, considering them presences if they exceed a set threshold 283 value, or absences otherwise. Then, a confusion matrix contrasts observed and predicted 284 occurrence states, and metrics derived from it. Examples include precision, sensitivity, and 285 true/false positive/negative rates. The value of thresholding continuous predictions has been 286 debated in the SDM literature (Liu et al. 2005; Freeman & Moisen 2008; Guillera-Arroita et 287 al. 2015), for prediction and evaluation (Lawson et al. 2014). How to determine the threshold 288 value when binary conversion is wanted is also debated. It is common to set the threshold at 289 an arbitrary value of 0.5 (Freeman & Moisen 2008), which provides an obvious decision 290 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 291 292 threshold to the observed prevalence of the species in question (Hanberry& He 2013). In the 293 multi-species context of JSDMs, a logical extension of this debate is whether to define 294 community-wide or species-specific thresholds. However, Lawson et al (2014) showed that 295 by using a probabilistic confusion matrix we can calculate threshold-dependent types of 296 metrics without the need to threshold probabilistic predictions. To avoid any issues with the 297 choice of threshold impacting our analysis, we followed this idea and used the probabilistic 298 confusion matrix approach when calculating our threshold-dependent metrics.

299 **Community dissimilarity indices** are widely used in community ecology to quantify the 300 dissimilarity between two realisations of species assemblages. Examples of common metrics 301 are Bray-Curtis dissimilarity and Jaccard distance (Chao *et al.* 2004). These metrics compare 302 predicted assemblages with observed assemblages and thus probabilistic predictions require 303 thresholding or samples from a binomial distribution before these metrics can be evaluated on 304 them. 305 **Species richness metrics** consider the ability of models to predict a single (but widely 306 studied) aspect of community composition – the number of species present. We have 307 evaluated species richness difference, defined as predicted minus observed richness, as it 308 directly compares predictions with observed richness values.

Likelihood metrics assess model fit by computing the probability of observing a given 309 community assemblage, assuming the model is 'correct' in its structure and parameter 310 311 estimates. For reasons of numerical stability, it is common to work with the log of the 312 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-313 computing the probability of observing that species' presence/absence observations- and then 314 315 combines these into a single metric, assuming the species' distributions to be independent 316 (i.e. the log-likelihoods are summed up). We can also define a joint log-likelihood that 317 assesses all species simultaneously as an assemblage at each site, accounting for the 318 correlation structure encoded in the JSDM formulation.

- 319
- ----

320 Table 1: Summary of evaluation metrics for JSDM prediction	320	Table 1: Sur	nmary of e	valuation	metrics t	for JSDI	M 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	Х	Х				
Area under the Receiver Operating Characteristic curve (AUC)		Х		Х			
Bray-Curtis dissimilarity	X	Х			Х		
Canberra Index	X	Х			Х		
Cohen's Kappa	X		Х				
Diagnostic odds ratio	X	Х	Х				
$F_1$ score	X	Х	Х				
False discovery ratio	X	Х	Х				

False negative rate	X	Х	X				
False omission rate	X	Х	X				
False positive rate	Х	Х	X				
Gower Index	X	Х			Х		
Gower Index (alternative)	X	Х			Х		
Jaccard distance	X	Х			Х		
Kendall rank correlation coefficient		Х		Х			
Kulczynski Index	X	Х			Х		
Log Likelihood – Independent		Х					Х
Log Likelihood – Joint		Х					Х
Mountford Index	X	Х			Х		
Mean error (bias)		Х		Х			
Mean square error		Х		Х			
Negative likelihood ratio	X	Х	X				
Negative predictive performance	X	Х	X				
Pearson correlation coefficient		Х		Х			
Positive likelihood ratio	X	Х	X				
Positive predictive performance / Precision	X	Х	X				
R <sup>2</sup> / Coefficient of determination	X	Х		Х			
Raup-Crick dissimilarity	X	Х	X				
Root mean square error		Х		Х			
Spearman rank correlation coefficient		Х		Х			
Species richness difference	X	Х				Х	
Sum of squared errors		Х		Х			
True negative rate / Specificity	X	Х	X				
True positive rate / Sensitivity	X	Х	X				
Youden's J statistic	Х	Х	Х				

#### *2.5. Case study*

We illustrate the application of different JSDM prediction types and the appropriate 322 323 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 324 dataset contains 9 species, 104 waterbodies (sites), and 3 measured covariates (area, road 325 density, presence of vertical wall), and was previously analysed in Pollock et al (2014) and 326 327 Wilkinson et al (2019). Species prevalence ranged from 0.02-0.52, and were mainly positively correlated with each other except for a single species negatively correlated with the 328 329 rest. In our analysis, we standardised the two continuous variables and used five-fold random 330 cross validation for model evaluation.

331 We chose to use a standard multivariate probit regression JSDM (Chib& Greenberg 1998) 332 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 333 334 coefficients and an inverse Wishart prior, with n + 2I degrees of freedom and scale matrix I, 335 on the correlation coefficients. The model was fit by MCMC using a Gibbs sampler 336 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 337 338 examined via visual assessment of trace plots for the posterior distributions. Model fitting and 339 predictions were undertaken on The University of Melbourne's Spartan HPC infrastructure 340 (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.

348 While we have defined 35 metrics that can be calculated to assess different aspects of JSDM 349 predictions, we present only a subset in this analysis for illustrative purposes. More detail on 350 the metrics, including how they are calculated, which prediction types they are appropriate 351 for, and how to interpret them can be found in Appendix S1. Threshold-dependent and 352 threshold-independent metrics are calculated on a per-species basis, whereas community 353 dissimilarity and species richness metrics are calculated per-site. Most metrics can 354 theoretically be calculated for either species or sites (as they are just comparing two binary 355 vectors: observations and predictions) so this split is based on how they are historically used 356 in ecological literature. Metrics were evaluated once for each of the 1000 posterior samples 357 drawn.

358

**359 3. Results** 

360 The patterns of predicted probabilities of presence for each species obtained from the 361 marginal and conditional marginal predictions both broadly matched the community 362 assemblage observed in the held-out testing data (Figure 3). The AUC for most species using 363 the marginal predictions was typically high (>0.88) with a standard deviation of <0.05 across 364 all posterior samples. Exceptions were L. ewingii and Litoria peroni, which were predicted 365 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 366 predictions conditioned on the known occurrence state of L. ewingii returned higher AUC 367 368 values by a mean of 0.03 (Fig. 3). The largest AUC gain was 0.08 for Litoria verreauxi. For 369 most other species-level evaluation metrics, we found the conditional marginal prediction to 370 outperform the marginal prediction, but in most cases the differences were relatively minor. We did observe some exceptions to this, such as an increased  $R^2$  value for L. verreauxi of 371 372 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 0.05-0.1 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.

378



379

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, white = absent). The middle plot depicts the median

383 probability of presence from the posterior distributions of the marginal predictions. The right-384 hand plot depicts the median probability of presence from the posterior distributions of the 385 conditional marginal predictions. These predictions are conditional on the known occurrence 386 state of *L. ewingii* which are highlighted with the red border.

387 The community-level approach of joint predictions yields probabilities of specific 388 assemblages occurring at a site or, as we present here, one can take random draws from the 389 posterior distribution to generate plausible community assemblages at a site. In Figure 4 we 390 show random draws from the joint and conditional joint prediction types for a single site in 391 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 392 393 *L.ewingii* led to draws that were closer to the observed assemblage.



395

Figure 4: Binary community-level predictions of the frog community at a single site in the 396 397 testing dataset. Community-level predictions here are random draws from the multivariate normal distribution representing plausible community assemblages under the environmental 398 399 conditions at the site (blue = present, white = absent). The left-hand plot depicts draws of 400 plausible community assemblages from joint predictions. The right-hand plot depicts draws 401 of plausible community assemblages from joint predictions, conditional on the known 402 occurrence state of L. ewingii at the site (highlighted with the red border). The top row of

403 both plots shows the known community assemblage in the testing dataset. The site considered404 in this figure corresponds to site 8 in Figure 3.

405 Community dissimilarity metrics were generally between 0.02-0.05 higher for conditional 406 joint predictions relative to standard joint predictions. Raup-Crick dissimilarity was the 407 exception. Conditional joint predictions had values that were, on average, 0.25 lower; 408 however, with a standard deviation of 0.37 across all predictions, this metric is possible 409 struggling to calculate correctly across all draws. For the site considered in Figure 4, we 410 observed improvements of between 0.08-0.1 for the Bray-Curtis dissimilarity, Jaccard 411 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.

418 **4. Discussion** 

We have reviewed and clarified the ways in which predictions of species distributions can be 419 420 approached with JSDMs. Our review highlights the additional functionality that JSDMs 421 enable relative to simple stacking of single-species models. JSDMs have two main 422 advantages: an ability to partition the effect of measured variables and residual correlations 423 between species, and predictions of community assemblages that account for these 424 correlations. By partitioning the effect of measured variables from residual correlations, 425 which may include species interactions, JSDMs potentially enable more accurate estimates of 426 environmental drivers of species' distributions. In addition, by leveraging the information in 427 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 theirecological question.

435 We have shown that conditioning on the known occurrence status of species can improve 436 both species- and community-level predictions by exploiting the correlations estimated in 437 JSDMs (Harris 2015). The extent of this improvement is likely dependent on the strength of the correlation between the known species and other species in the community. We can 438 439 expect that a species that is highly correlated with the known species (positively or 440 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 441 species in the community, or to differing extents between species. 442

443 The JSDM tended to overpredict species richness for all prediction types that account for 444 correlations between species. For our case study of nine frog species, the JSDM predicted 445 approximately one extra species per site. Zurell et al (2019) similarly found that a JSDM 446 overpredicted species richness compared to a SSDM. As the JSDM only overpredicted 447 species richness for prediction types that account for residual correlations, the estimated 448 correlations could potentially explain why the JSDM tended to overpredict. A largely positive 449 correlation matrix between most species combined with high marginal occurrence 450 probabilities, as seen here, could cause the JSDM to predict likely assemblages in excess of 451 restrictions, such as site carrying capacities and/or dispersal limitations.

452 We have presented JSDM prediction methods using only a single dataset for illustrative 453 purposes. A more in-depth study using a wider array of datasets is required, as dataset 454 properties likely influence the relative merits of the different prediction types. For datasets in 455 which all species are reasonably prevalent, the expected performance difference between 456 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 457 458 prediction types. The value of additional information in the form of species co-occurrence 459 will also be greater when correlations between species are stronger; the known occurrence of 460 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 461 462 different prediction types when making vs extrapolative predictions, as our case study only 463 focused on the former type of evaluation and the latter has been shown to be more difficult 464 (Norberg et al. 2019).

465 A small body of literature focuses on large-scale comparisons of JSDMs and single-species 466 models. These studies have included many species and model types (Zurell et al. 2019; 467 Norberg *et al.* 2019) but have not addressed all of the prediction types available to JSDMs. 468 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 469 470 draws to generate species-specific probabilities, which is an approximation of the marginal 471 distribution. Second, they used joint predictions to generate species richness metrics 472 equivalent to ours. Third, they used joint predictions and community dissimilarity metrics to 473 test predictions of species turnover between sites rather than to test community composition 474 at a site. To our knowledge, no JSDM studies have yet included conditional or conditional 475 marginal predictions, although they have been identified as desirable avenues of research 476 (Zurell et al. 2019; Norberg et al. 2019; Blanchet, Cazelles & Gravel 2020). Evaluations with 477 a broader array of JSDM implementations, prediction types, and datasets are warranted. Both 478 Norberg et al (2019) and Zurell et al (2019) included latent factor JSDMs in their 479 comparisons, whereas we use a multivariate probit model. Norberg et al (2019) found that the 480 HMSC JSDM (Ovaskainen et al. 2016b) outperformed both other JSDMs and SSDMs, while 481 Zurell et al (2019) found that the boral JSDM (Hui 2016) had similar results to those 482 described here. Zurell et al (2019) suggested that the poor performance of boral was a result 483 of how the latent factor model is used when extrapolating predictions. Prediction using latent 484 factor JSDMs can be performed in two ways: (1) marginalizing over the latent factors by 485 assigning the mean value of modelled sites to prediction sites when performing regression-486 style predictions, and (2) 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 487 488 types defined in this paper. This second method may be a better alternative to latent factor 489 model prediction than marginalizing over the latent factors.

490 **5.** Conclusion

JSDMs 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, focussed 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

498 predictions. Ecologists seeking to use these prediction methods and evaluation metrics should 499 consider which method and metric are most closely linked to the ecological question they are 500 investigating. A larger comparison of prediction types that considers different JSDM 501 implementations and multiple datasets is required to evaluate general performance of these 502 models and prediction types.

503

504 Acknowledgements

We thank Peter Vesk, Brendan Wintle, and Jian Yen for insightful discussions, and Els Van
Burm and Kristina Chyn for manuscript reviews. DW is funded by an Australian Government
Research Training Program Scholarship. NG, GGA and RT are supported by Australian
Research Council (ARC) Discovery Early Career Researcher Awards (DE180100635,
DE160100904 and DE170100601).

- 510
- 511 Conflicts of Interest

512 No conflicts of interest are declared.

513

### 514 Authors' Contributions

515 All authors conceived the ideas and methodology. DW implemented the analysis. DW led 516 writing the manuscript but all authors contributed significantly throughout and gave final 517 approval before submission.

518

### 519 Data Accessibility

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.

525 **Reference List** 

526	Araújo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling
527	species distributions under climate change. Global Ecology and Biogeography, 16, 743–
528	753.
529	Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of
530	ecological interactions. Ecology Letters, 23, 1050–1063.
531	Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014). Stacking species
532	distribution models and adjusting bias by linking them to macroecological models.
533	Global Ecology and Biogeography, 23, 99–112.
534	Chao, A., Chazdon, R.L., Colwell, R.K. & Shen, TJ. (2004). A new statistical approach for
535	assessing similarity of species composition with incidence and abundance data. Ecology
536	<i>Letters</i> , <b>8</b> , 148–159.
537	Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. <i>Biometrika</i> , 85,
538	347–361.
539	Clark, J.S., Nemergut, D., Seyednasrollah, B., Turner, P.J. & Zhang, S. (2017). Generalized
540	joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious
541	data. Ecological Monographs, 87, 34–56.
542	Elith, J. & Leathwick, J.J.R. (2009). Species Distribution Models: Ecological Explanation
543	and Prediction Across Space and Time. Annual Review of Ecology, Evolution,, 40,
544	677–697.
545	Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction
546	errors in conservation presence/absence models. Environmental Conservation, 24, 38-
547	49.
548	Freeman, E.A. & Moisen, G.G. (2008). A comparison of the performance of threshold criteria
549	for binary classification in terms of predicted prevalence and kappa. Ecological
550	<i>Modelling</i> , <b>217</b> , 48–58.
551	Gelfand, A.E., Schmidt, A.M., Wu, S., Silander, J.A., Latimer, A. & Rebelo, A.G. (2005).
552	Modelling species diversity through species level hierarchical modelling. Journal of the
553	Royal Statistical Society: Series C (Applied Statistics), 54, 1–20.
554	Golding, N. & Harris, D.J. (2015). BayesComm: Bayesian Community Ecology Analysis.
	This article is protected by copyright. All rights reserved

555	Golding, N., Nunn, M.A. & Purse, B. V. (2015). Identifying biotic interactions which drive
556	the spatial distribution of a mosquito community. <i>Parasites &amp; Vectors</i> , <b>8</b> , 367.
557	Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E.,
558	Mccarthy, M.A., Tingley, R. & Wintle, B.A. (2015). Is my species distribution model fit
559	for purpose? Matching data and models to applications. Global Ecology and
560	<i>Biogeography</i> , <b>24</b> , 276–292.
561	Guisan, A. & Rahbek, C. (2011). SESAM - a new framework integrating macroecological
562	and species distribution models for predicting spatio-temporal patterns of species
563	assemblages. Journal of Biogeography, 38, 1433-1444.
564	Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology.
565	Ecological Modelling, 135, 147–186.
566	Hanberry, B.B. & He, H.S. (2013). Prevalence, statistical thresholds, and accuracy
567	assessment for species distribution models. Web Ecol, 13, 13-19.
568	Harris, D.J. (2015). Generating realistic assemblages with a joint species distribution model
569	(D. Warton, Ed.). Methods in Ecology and Evolution, 6, 465–473.
570	Hui, F.K.C. (2016). boral - Bayesian Ordination and Regression Analysis of Multivariate
571	Abundance Data in r. Methods in Ecology and Evolution, 7, 744–750.
572	Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J.,
573	Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, JC.,
574	Zimmermann, N.E. & O'Hara, R.B. (2012). Towards novel approaches to modelling
575	biotic interactions in multispecies assemblages at large spatial extents. Journal of
576	<i>Biogeography</i> , <b>39</b> , 2163–2178.
577	Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014). Prevalence, thresholds
578	and the performance of presence-absence models. Methods in Ecology and Evolution, 5,
579	54-64.
580	Leathwick, J.R. & Austin, M.P. (2001). Competitive interactions between tree species in New
581	Zealand old-growth indigenous forests. <i>Ecology</i> , 82, 2560–2573.
582	Legendre, P. & De Cáceres, M. (2013). Beta diversity as the variance of community data:
583	Dissimilarity coefficients and partitioning. <i>Ecology Letters</i> , <b>16</b> , 951–963.

584	Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005). Selecting thresholds of
585	occurrence in the prediction of species distributions. <i>Ecography</i> , 28, 385–393.
586	Liu, C., White, M. & Newell, G. (2009). Measuring the accuracy of species distribution
587	models: A review. 18th World IMACS Congress and MODSIM09 International
588	Congress on Modelling and Simulation: Interfacing Modelling and Simulation with
589	Mathematical and Computational Sciences, Proceedings, 4241–4247.
590	Mateo, R.G., Felicísimo, Á.M., Pottier, J., Guisan, A. & Muñoz, J. (2012). Do stacked
591	species distribution models reflect altitudinal diversity patterns? <i>PLoS ONE</i> , 7.
592	Meade, B., Lafayette, L., Sauter, G. & Tosello, D. (2017). Spartan HPC-Cloud Hybrid:
593	Delivering Performance and Flexibility.
594	Meier, E.S., Kienast, F., Pearman, P.B., Svenning, JC., Thuiller, W., Araújo, M.B., Guisan,
595	A. & Zimmermann, N.E. (2010). Biotic and abiotic variables show little redundancy in
596	explaining tree species distributions. <i>Ecography</i> , <b>33</b> , 1038–1048.
597	Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Araújo,
598	M.B., Dallas, T., Dunson, D., Elith, J., Foster, S.D., Fox, R., Franklin, J., Godsoe, W.,
599	Guisan, A., O'Hara, B., Hill, N.A., Holt, R.D., Hui, F.K.C., Husby, M., Kålås, J.A.,
600	Lehikoinen, A., Luoto, M., Mod, H.K., Newell, G., Renner, I., Roslin, T., Soininen, J.,
601	Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N.E., Gravel, D. &
602	Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33
603	species distribution models at species and community levels. Ecological Monographs,
604	<b>89</b> , 1–24.
605	Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016a). Using latent variable models
606	to identify large networks of species-to-species associations at different spatial scales.
607	Methods in Ecology and Evolution, 7, 549–555.
608	Ovaskainen, O., Roy, D.B., Fox, R. & Anderson, B.J. (2016b). Uncovering hidden spatial
609	structure in species communities with spatially explicit joint species distribution models.
610	Methods in Ecology and Evolution, 7, 428–436.
611	Parris, K.M. (2006). Urban amphibian assemblages as metacommunities. Journal of Animal

612 *Ecology*, **75**, 757–764.

613	Parviainen, M., Marmion, M., Luoto, M., Thuiller, W. & Heikkinen, R.K. (2009). Using
614	summed individual species models and state-of-the-art modelling techniques to identify
615	threatened plant species hotspots. Biological Conservation, 142, 2501-2509.
616	Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz,
617	N.G., Alm, T., Zimmermann, N.E. & Guisan, A. (2010). Species distribution models
618	reveal apparent competitive and facilitative effects of a dominant species on the
619	distribution of tundra plants. <i>Ecography</i> , <b>33</b> , 1004–1014.
620	Pineda, E. & Lobo, J.M. (2009). Assessing the accuracy of species distribution models to
621	predict amphibian species richness patterns. Journal of Animal Ecology, 78, 182–190.
622	Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A.
623	& McCarthy, M.A. (2014). Understanding co-occurrence by modelling species
624	simultaneously with a Joint Species Distribution Model (JSDM) (J. McPherson, Ed.).
625	Methods in Ecology and Evolution, 5, 397–406.
626	R Core Team. (2018). R: A Language and Environment for Statistical Computing.
627	Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I.,
628	Pöyry, J. & Settele, J. (2012). Increasing range mismatching of interacting species under
629	global change is related to their ecological characteristics. Global Ecology and
630	<i>Biogeography</i> , <b>21</b> , 88–99.
631	Taylor-Rodríguez, D., Kaufeld, K., Schliep, E.M., Clark, J.S. & Gelfand, A.E. (2017). Joint
632	Species Distribution Modeling: Dimension Reduction Using Dirichlet Processes.
633	Bayesian Analysis, 12, 939–967.
634	Thuiller, W., Pollock, L.J., Gueguen, M. & Münkemüller, T. (2015). From species
635	distributions to meta-communities. <i>Ecology Letters</i> , <b>18</b> , 1321–1328.
636	Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. &
637	Hui, F.K.C. (2015). So Many Variables: Joint Modeling in Community Ecology. Trends
638	in Ecology and Evolution, <b>30</b> , 766–779.
639	Wilkinson, D.P. (2019). JSDM_Prediction v0.1.0. Zenodo,
640	https://dx.doi.org/10.5281/zenodo.3514766.
641	Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2019). A

642	comparison of joint species distribution models for presence-absence data. Methods in
643	Ecology and Evolution, 10, 198–211.

## 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.,

646 Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt,

- 647 N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C.
- 648 (2013). The role of biotic interactions in shaping distributions and realised assemblages
- of species: Implications for species distribution modelling. *Biological Reviews*, 88, 15–
  30.

# 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.

- 654 Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O.
- 655 (2019). Testing species assemblage predictions from stacked and joint species
  656 distribution models. *Journal of Biogeography*, **00**, 1–13.
- 657

Author







