Appendices

A Latent factors models

Latent factor models (also called 'latent variable models') were developed in the JSDMs framework to reduce the dimension of the residual correlation matrix $\boldsymbol{R}$, whose number of elements increases quadratically with the number of species (Warton et al., 2015). Following the same notation of Box 1, these models write:

$$\begin{aligned} y_{ij} &= \mathrm{I}(z_{ij} > 0), \\ \boldsymbol{z}_i = \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + e_{ij}, &\quad e_{ij} \overset{\text{iid}}{\sim} N(0,1), \quad \omega_{ij} \overset{\text{iid}}{\sim} N(0,1). \end{aligned} \tag{A.1}$$

where I() is the indicator function and $N(0,1)$ is the standard univariate Gaussian distribution. The $S \times r$ matrix $\boldsymbol{\Lambda}$ is the *factor loading* matrix, and the r-dimensional Gaussian random vectors $\boldsymbol{\omega}_i$ are called *latent factors*. When the number of latent factors r is smaller than the number of species (usually $r \ll S$), the dimension of the residual correlation matrix decreases, as we now have $\boldsymbol{R} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T$. Latent variables can be interpreted as 'missing covariates', and a shared (or opposite) response to these missing covariates (through similar or opposite latent loadings), introduces the residual correlation among taxa.

Importantly, latent variables change the probability of presence of species, since we now have that the probability of presence of species i at site j is $p_{ij} = \Phi(\boldsymbol{\beta}_j\boldsymbol{x}_i^T + \boldsymbol{\Lambda}_j\boldsymbol{\omega}_i^T)$, where $\boldsymbol{\Lambda}_j$ is the j-th row of the factor loadings matrix and Φ is the cumulative distribution function of the standard Gaussian distribution. However, this is only true when the latent variables $\boldsymbol{\omega}_i$ are known. For out-of-sample predictions, that are fundamental to understand the distribution of species under new, climate change driven, scenarios, one has to integrate out these latent variables. This is not only computationally costly, but also reduces the advantages of latent factor models, whose prediction is now only based on the environmental variables, as in the Multivariate probit model described in the main text.

In Appendix B we prove that in the Gaussian setting the regression coefficients do not change when we assume the residuals to be correlated. The same proof does not hold for latent factor models, and one might find differences in the estimated regression coefficients due to the different structure of the model. However, we want to stress here how those discrepancies are not driven by the effect of other species. In fact, as it is the case in every regression, in the multivariate probit model the regression coefficient B_{jk} is linked to the partial correlation between species j and the environmental covariate k, with the effect of the other environmental covariates being removed. Importantly, the effect of the presence-absence of the other species is not removed, which is the reason why both SDMs and JSDMs can only fit the realized niche (as also discussed in the last section of Appendix B). With latent factor models, when estimating β_{jk}, we also remove the effect of the latent variables, that might be related to missing covariates, but are not themselves representative of the explicit presence/absence of other species. Therefore, there is no reason why latent factor models should better approximate the fundamental niche.

To conclude, if latent factor models solve the scale problem of JSDMs (but scale badly with the number of sites, Pichler and Hartig, 2020) and might provide computationally driven changes in the estimated probability of presence and estimated niches, they still do not solve the problem of taking into account the effects of other species. See Chapter 7 of Ovaskainen and Abrego (2020) for a complete description of latent factor models and their correct interpretation.

References

Ovaskainen, O. and Abrego, N. (2020). *Joint Species Distribution Modelling: With Applications in R*. Ecology, Biodiversity and Conservation. Cambridge University Press.

Pichler, M. and Hartig, F. (2020). A new method for faster and more accurate inference of species associations from novel community data.

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., and Hui, F. K. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology and Evolution*, 30(12):766 – 779.

B Computational details and proofs

To better understand the fundamental differences between SDMs and JSDMs, an interesting angle is to compare the species environmental niches they estimate. Since considering a link function leads to intractable calculations, we focus on the particular case of continuous data (e.g. biomass), where the latent variable in the equation coincides with the data. We first compute the posterior mean of the regression coefficients in the Bayesian settings for SDMs (Section B.1) and JSDMs (Section B.2). In Section B.3, we then compute for both models the maximum likelihood estimator of the regression coefficients.

We rewrite the model defined in Box 1: $i = 1, \dots, n$ is the index for sites, $j = 1, \dots, S$ the index for species, $k = 1, \dots, K$ is the index for the environmental covariates, $\boldsymbol{Y}$ is the matrix whose lines are the observed vectors of species at the different sites, while the lines of $\boldsymbol{X}$ are the measured environmental covariate at each site and $\boldsymbol{E}$ is the matrix containing the error terms (i.e. $\epsilon_{ij} = [E]_{ij}$). We now call $\boldsymbol{B}$ the $S \times K$ matrix of the regression coefficients, whose rows $\boldsymbol{B}_j$ are the species-specific responses to the environment. Noticed that we changed from notation of Box 1 concerning to the regression coefficients, to guarantee a better readability of the following computations. Therefore the models write:

$$\underbrace{\boldsymbol{Y}}_{n\times S} = \underbrace{\boldsymbol{X}}_{n\times K}\underbrace{\boldsymbol{B}^T}_{K\times S} + \underbrace{\boldsymbol{E}}_{n\times S}. \tag{A.2}$$

where $\boldsymbol{B}^T$ is the transpose of $\boldsymbol{B}$. The consequent likelihood of the data is:

$$P(\boldsymbol{Y} \mid \boldsymbol{B}, \boldsymbol{\Sigma}, \boldsymbol{X}) = (2\pi)^{-nS/2}|\boldsymbol{\Sigma}|^{-n/2}e^{-1/2\mathrm{tr}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B}^T)\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B}^T)^T}.$$

Where $\boldsymbol{\Sigma}^{-1}$ is the inverse of matrix $\boldsymbol{\Sigma}$ and tr() is the trace operator. When the $\boldsymbol{\Sigma}$ matrix is diagonal, i.e. $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \dots, \sigma_S^2)$, each species j is modelled independently from the others (i.e. SDMs), that is:

$$\boldsymbol{y}_j \mid \boldsymbol{B}_j, \sigma_j, \boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{B}_j^T, \sigma_j^2\boldsymbol{I}). \tag{A.3}$$

where $\boldsymbol{y}_j$ the j^{th} column of $\boldsymbol{Y}$ and $\boldsymbol{I}$ is the identity matrix.

Notice that with Gaussian data we have no identifiability issue, and there are no constraints on the variances of both SDMs and JSDMs.

Since we work in the Bayesian framework, inference is related to the posterior distribution of the parameters (i.e. the probability distribution of the parameters given the observed data) that we want to estimate, here the regression coefficients of the estimated environmental niche $\boldsymbol{B}$. To do so, we first need to define the prior distribution of the parameters for both SDMs and JSDMs. To be able to carry out analytical computations, we need to work under particular priors, called conjugate priors (Christensen et al., 2010).

B.1 Species Distribution Models

We compute the posterior distribution of the regression coefficients in the case of SDMs. In order to better visualize the difference between SDMs and JSDMs, we hereafter call $\boldsymbol{\beta}_j = \boldsymbol{B}_j^T$ the species-specific vector of responses to the environment estimated by SDMs. Since we model each species independently, we compute here below the estimated regression coefficients for a given species, and we therefore drop the j index for the sake of clarity. For such model the likelihood is:

$$P(y_1, \dots, y_n \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n/2}e^{-1/(2\sigma^2)\sum_{i=1}^n (y_i - \boldsymbol{\beta}^T x_i)^2}. \tag{A.4}$$

To carry out analytical computations on the posterior distribution, we choose to consider a Normal - inverse Gamma prior for the distribution of the regression coefficients $\boldsymbol{\beta}$ and the variance σ, that are conjugate priors for the likelihood of above.

$$\begin{aligned} P(\boldsymbol{\beta}, \sigma) &= P(\boldsymbol{\beta} \mid \sigma)P(\sigma) \\ \boldsymbol{\beta}|\sigma &\sim N_K(\boldsymbol{\beta}_0, \sigma^2\boldsymbol{V}_0) \\ \sigma^2 &\sim IG(a_0, b_0), \end{aligned}$$

where N_K is the multivariate K-dimensional Gaussian and distribution$\boldsymbol{V}_0$ represents the covariance matrix across the regression coefficients of the different covariates conditional on σ^2 and a_0 and b_0

are the shape and scale parameters of the inverse gamma distribution (i.e. the hyperprior of the variance σ^2). This is a commonly used model in the Bayesian literature, and the full computation of the posterior distribution is described in Christensen et al. (2010). The posterior distribution of the regression coefficients is thus:

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X} \sim t_K(a_n, \boldsymbol{\beta}_n, b_n \boldsymbol{V}_n), \tag{A.5}$$

where t_K denotes the multivariate Student t-distribution with parameters $a_n, b_n, \boldsymbol{\beta}_n, \boldsymbol{V}_n$ whose values are:

$$\begin{aligned}
\boldsymbol{\beta}_n &= (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}(\boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{V}_0^{-1}\boldsymbol{\beta}_0) \\
\boldsymbol{V}_n &= (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1} \\
a_n &= a_0 + n \\
b_n &= \frac{1}{a_n}(a_0 b_0 + \boldsymbol{\beta}_0^T\boldsymbol{V}_0^{-1}\boldsymbol{\beta}_0 + \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{\beta}_n^T\boldsymbol{V}_n^{-1}\boldsymbol{\beta}_n).
\end{aligned}$$

Therefore, we obtain that:

$$\mathbb{E}[\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X}] = \boldsymbol{\beta}_n = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}(\boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{V}_0^{-1}\boldsymbol{\beta}_0).$$

The credible intervals of the regression coefficients depend on their variance, that is given by:

$$\mathrm{Var}[\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X}] = \frac{a_n}{a_n - 2} b_n \boldsymbol{V}_n = \frac{a_0 + n}{a_0 + n - 2}\frac{1}{a_n}(a_0 b_0 + \boldsymbol{\beta}_0^T\boldsymbol{V}_0^{-1}\boldsymbol{\beta}_0 + \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{\beta}_n^T\boldsymbol{V}_n^{-1}\boldsymbol{\beta}_n)(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}. \tag{A.6}$$

B.2 Joint species Distribution Models

We now consider model (A.2). If we call $\boldsymbol{y}_i$ the rows of Y, the likelihood of such model can be rewritten as:

$$P(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \mid \boldsymbol{B}, \boldsymbol{\Sigma}, \boldsymbol{X}) = (2\pi)^{-nS/2}|\boldsymbol{\Sigma}|^{-n/2} e^{-1/2 \sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{B}\boldsymbol{x}_i)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{B}\boldsymbol{x}_i)},$$

using the property that $\mathrm{tr}(\boldsymbol{X}^T\boldsymbol{Y}) = \mathrm{vec}(\boldsymbol{X})^T\mathrm{vec}(\boldsymbol{Y})$, where $\mathrm{vec}(\boldsymbol{X})$ denotes the vectorization of $\boldsymbol{X}$.

Again, we need to choose a prior distribution over $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ that leads to a tractable posterior. Following Rowe (2002), Chapter 8, we choose the following conjugate prior that is consistent with the prior of the previous model:

$$\begin{aligned}
P(\boldsymbol{B}, \boldsymbol{\Sigma}) &= P(\boldsymbol{B} \mid \boldsymbol{\Sigma})P(\boldsymbol{\Sigma}) \\
\boldsymbol{B} \mid \boldsymbol{\Sigma} &\sim MN_{S,K}(\boldsymbol{B}_0, \boldsymbol{\Sigma}, \boldsymbol{V}_0) \\
\boldsymbol{\Sigma} &\sim IW(\nu_0, \boldsymbol{S}_0),
\end{aligned}$$

where $MN_{S,K}()$ is the matrix normal distribution. The definition of such a distribution in that $\boldsymbol{B} \sim MN_{S,K}(\boldsymbol{B}_0, \boldsymbol{\Sigma}, \boldsymbol{V}_0)$ if $\mathrm{vec}(\boldsymbol{B}) \sim N_{S\times K}(\mathrm{vec}(\boldsymbol{B}_0), \boldsymbol{V}_0 \otimes \boldsymbol{\Sigma})$, where $\otimes$ denotes the Kronecker product and $N_{S\times K}$ is the multivariate normal distribution with dimension given by the product between S and K.

Therefore, $\boldsymbol{V}_0$ (a $K \times K$ matrix)is the prior covariance matrix across the elements of the same column of $\boldsymbol{B}$, while $\boldsymbol{\Sigma}$ (a $S \times S$ matrix) is the covariance matrix across the elements of the same row.

In other words, $\boldsymbol{\Sigma}$ is the covariance matrix of the regression coefficients across all the species, for each covariate k: $\boldsymbol{B}_k \sim N_S(\boldsymbol{B}_{0,k}, \boldsymbol{\Sigma})$. Instead, $\boldsymbol{V}_0$ is the prior covariance matrix of the regression coefficients across the different covariates that for the regression coefficients of species j (i.e. the j-th line of the matrix $\boldsymbol{B}$, that we call $\boldsymbol{B}_j$): $\boldsymbol{B}_j \sim N_S(\boldsymbol{B}_{0,j}^T, \boldsymbol{V}_0)$.

To guarantee that the priors on the regression coefficients are consistent across SDMs and JSDMs, we have that the rows of $\boldsymbol{B}_0$ are equal to $\boldsymbol{\beta}_{0,j}^T$ of SDMs in section B.1 (i.e. $\boldsymbol{B_0} = [\boldsymbol{\beta}_{0,1}, \ldots, \boldsymbol{\beta}_{0,S}]^T$).

The computations of the posterior distribution of the regression coefficients $\boldsymbol{\beta}$ are given by Rowe (2002), who obtains:

$$\boldsymbol{B} \mid \boldsymbol{Y}, \boldsymbol{X} \sim t_{S,K}(n + \nu_0 - S + 1, \boldsymbol{B}_n, \boldsymbol{G}, (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1})$$

where $t_{S,K}$ is the matrix t-distribution with parameters $n + \nu_0 - S + 1, \boldsymbol{B}_n, \boldsymbol{G}, (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}$ and:

$$\boldsymbol{G} = \boldsymbol{S}_0 + \boldsymbol{Y}^T\boldsymbol{Y} + \boldsymbol{B}_0\boldsymbol{V}_0^{-1}\boldsymbol{B}_0^T + (\boldsymbol{Y}^T\boldsymbol{X} + \boldsymbol{B}_0\boldsymbol{V}_0^{-1})(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}(\boldsymbol{Y}^T\boldsymbol{X} + \boldsymbol{B}_0\boldsymbol{V}_0^{-1})^T$$
$$\boldsymbol{B}_n = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}(\boldsymbol{Y}^T\boldsymbol{X} + \boldsymbol{B}_0\boldsymbol{V}_0^{-1}).$$

Therefore we have that :

$$\mathbb{E}[\boldsymbol{B} \mid \boldsymbol{Y}, \boldsymbol{X}] = \boldsymbol{B}_n = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}(\boldsymbol{Y}^T\boldsymbol{X} + \boldsymbol{B}_0\boldsymbol{V}_0^{-1}).$$

This is equivalent to say that for each line $\boldsymbol{B}_j$ of $\boldsymbol{B}$ it holds that:

$$\mathbb{E}[\boldsymbol{B}_j^T \mid \boldsymbol{y}_j, \boldsymbol{X}] = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}(\boldsymbol{X}^T\boldsymbol{y}_j + \boldsymbol{V}_0^{-1}\boldsymbol{B}_{0,j}^T)$$

which is identical to the $\boldsymbol{\beta}_{j_n}$, the posterior mean of the regression coefficients in the independent case, in Appendix B.1. In other words, the two posterior means are equivalent between SDMs and JSDMs. The correlation among the residuals does not change the posterior mean of the regression coefficients, when the data are normally distributed. Importantly, in a frequentist setting where we obtain the regression coefficients $\hat{\boldsymbol{\beta}}$ by maximising the likelihood of the data, the estimates for SDMs and JSDMs are exactly the same (Appendix B.3).
To study the credible intervals of these regression coefficients, we focus on the marginal distribution of the regression coefficients of each species, given by the lines $\boldsymbol{B}_j$ of $\boldsymbol{B}$. Each $\boldsymbol{B}_j^T$ follows a generalized multivariate t-distribution, whose variance is given by:

$$\mathrm{Var}[\boldsymbol{B}_j^T \mid \boldsymbol{Y}, \boldsymbol{X}] = \frac{G_{jj}(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{V}_0^{-1})^{-1}}{n + \nu_0 - S + 1}.$$

This expression is really complicated due to the term G_{jj} (the j-th element on the diagonal of the scale parameter of the matrix t-distribution), however we can recognize some of the formulations that we had in (A.6). In particular, we see that, compared to (A.6), the degrees of freedom are decreased by the number of species, which would imply in general a higher variance of the regression coefficients, and thus, larger confidence intervals. However the term G_{jj} might counterbalance such an effect and we can't state that one method will always provide larger confidence intervals than the other.

B.3 Results in the frequentist approach

In the frequentist framework, we aim to find the maximum likelihood estimator (MLE) of the parameters of the models. In the case of independent species (SDMs), the MLE of the regression coefficients is well known. Indeed, the MLE and Best Unbiased Linear Estimator (BLUE) of $\boldsymbol{\beta}_j$ of model (A.3) is:

$$\hat{\boldsymbol{\beta}}_{j,\mathrm{MLE}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}_j.$$

If we consider model (A.2) with a full covariance matrix Σ, the MLE and BLUE is (e.g. Rencher, 2002):

$$\hat{\boldsymbol{\beta}}_{\mathrm{MLE}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y},$$

which implies that for each species j we obtain that $\boldsymbol{\beta}_j$, the j^{th} line of $\boldsymbol{\beta}$ is :

$$\hat{\boldsymbol{\beta}}_{j,\mathrm{MLE}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}_j.$$

The equivalence of the estimated regression coefficients in the two models is consistent with the Bayesian approach studied before.

B.4 Discussion of the results and extension to other kind of data

This result is not as surprising as it might seem. As it is the case in every regression, for both models (A.2) (JSDM) and (A.3) (SDM) the regression coefficient B_{jk} represents the partial correlation between species j and the covariate k, with the effect of the other environmental variables being

removed. Importantly, in both models the effect of the presence-absence of the other species is not removed. This is the reason why both models confound the biotic and abiotic effect in the estimation of the regression coefficients $\boldsymbol{\beta}$, and can only retrieve the realized niche. While the extensions of the proof of above to non-Gaussian data is very challenging, the argument that JSDMs (as SDMs) do not control for the effect of other species when determining the regression coefficients still holds. Therefore, we suggest that both models still confound the environment and the biotic context, regardless of the kind of data they are modelling, even thus some computationally driven small differences might arise. Indeed, when we carefully applied both models to an empirical presence-absence data, we did not find any difference in the estimated regression coefficients (Box 2 of the main text).

References

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. (2010). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians.* Chapman & Hall/CRC Texts in Statistical Science. CRC Press.

Rencher, A. (2002). *Methods of multivariate analysis.* Wiley Series in Probability and Statistics.

Rowe, D. B. (2002). *Multivariate Bayesian statistics: models for source separation and signal unmixing.* Chapman and Hall/CRC.

C Implementation details and code of Box 2

C.1 Implementation details and prior choice

The data are available from the R package **ade4** (Dray and Dufour, 2007) and include the presence/absence of 82 species on 75 sites, with different snowmelt dates, in Aravo (Valloire, France, Choler, 2005). We considered only the species with more than 4 presences, for a total of 65 species. We have considered the snowmelt date (Julian day, averaged over 1997–1999) as an environmental covariate in our study, including a quadratic term (using orthogonal polynomials to reduce correlation among the covariates):

$$\begin{aligned} y_{ij} &= \mathbb{1}(z_{ij} > 0) \\ \boldsymbol{z}_{ij} &= \beta_{j0} + \beta_{j1}\ \text{snow}_i + \beta_{j2}\ \text{snow}_i^2 + e_{ij} \end{aligned} \tag{A.7}$$

where the residuals are correlated for JSDMs, $\boldsymbol{e}_i \overset{iid}{\sim} N_S(0, \boldsymbol{R})$ and independent for SDMs, $\boldsymbol{e}_i \overset{iid}{\sim} N_S(0, \boldsymbol{I})$.

We used the priors suggested by Golding and Harris (2015). For both SDMs and JSDMs, for each element of each vector of regression coefficients $\boldsymbol{\beta}_j$ we used a diffuse normal prior with mean 0 and variance 100. This is a widely used prior which exhibits little influence on the posterior. Concerning the prior on the correlation matrix, we also followed Golding and Harris (2015) by choosing for the unidentified covariance matrix $\boldsymbol{W}$ (that is normalized to the correlation matrix $\boldsymbol{R}$ at every step of the MCMC sampler) an inverse Wishart prior:

$$\boldsymbol{W} \sim IW(\nu, \boldsymbol{V}).$$

The hyparameters ν (degrees of freedom) and $\boldsymbol{V}$ (scale matrix), were chosen as:

$$\begin{aligned} \nu &= n + 2S \\ \boldsymbol{V} &= 2S\boldsymbol{I}, \end{aligned}$$

that corresponds to a weakly informative prior on the off-diagonal elements of the correlation matrix, that are slightly shrinked towards zero. See Golding and Harris (2015) for a complete motivation on the choice of such prior.

For both SDMs and JSDMs we run a single chain of $100k$ iterations with $50k$ burn-in. Convergence was assessed visually (Figure A.1).

We computed three types of residuals for SDMs. The raw residuals as the difference between observed presence-absence and predictions, Pearson residuals as raw residuals normalised by the variance of the prediction, and residuals at the latent variable level as the difference between the inferred latent variable and the regression term. For each kind of residuals we calculated their correlation for each element of the MCMC chain, in order to have a measure of uncertainty around this estimate.

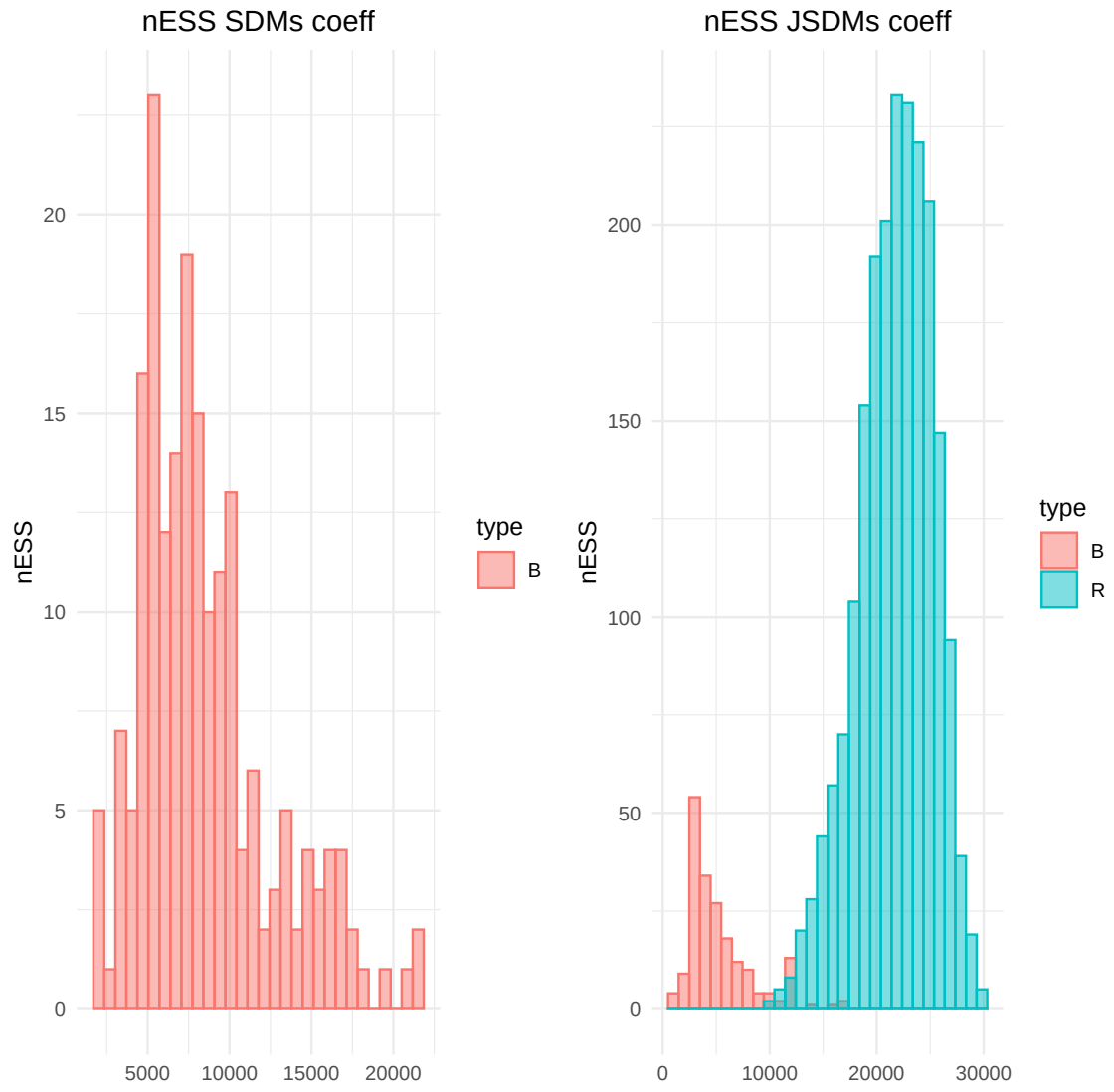

Figure A. 1: Effective sample size (nESS) for SDMs (left) and JSDMs (right). Regression coefficients nESS are represented in red, the nESS for the elements of the residual correlation matrix (for JSDMs only) are represented in green.

C.2 Code

```
list=rm(list=ls())

library("ade4")
library("BayesComm")

data(aravo)

spe = aravo$spe
env = aravo$env

#set Y
spePA = spe>0
spePA = apply(spePA, FUN= as.numeric, MARGIN = 2)
```

```
spePA_low = spePA[,-which(colSums(spePA) < 5)] #keeping species with more than 4
    records
Y=as.matrix(spePA_low)
nsp= ncol(Y)
# 67 species, 75 sites

X = cbind(poly(env$Snow,2))
np = ncol(X) + 1

nits=1e5
nburn = 5e4

#SDM fitting: R is the identity matrix
SDM = BC(Y, X, model = "environment", its = nits, burn = nburn, thin = 1, verbose
    =2)


## To compute residual correlation in SDMs. We will consider three ways of
    computing them. First, the residuals at the latent variable level (as JSDMs
    do), than the residuals at the observed level, both raw residuals (i.e. the
    difference between observed and predicted) and Pearson residuals (raw
    residuals divided by the variance of the prediction)

#Chains of regression coefficients
B_s = vector()
for(i in 1:ncol(Y)){

  B_s = cbind(B_s, SDM$trace$B[[i]])

}

#Correlation of residuals at the latent level
R_s = array(dim = c(nrow(B_s), nsp , nsp))

for(j in 1:dim(SDM$trace$z)[1]){

  mu = cbind(rep(1,nrow(X)), X) %*% matrix(as.vector(B_s[j,]),nrow=np,ncol=nsp)
  R_s[j, , ] = cor(as.matrix(SDM$trace$z[j,,]) - mu)

}

# Correlation of Raw residuals
R_s_Obs = array(dim = c(nrow(B_s), nsp , nsp))

for(j in 1:dim(SDM$trace$z)[1]){

  R_s_Obs[j, , ] = cor(Y- pnorm(SDM$trace$z[j,,]) )

}

#Pearson residuals
R_s_ObsP = array(dim = c(nrow(B_s), nsp , nsp))

for(j in 1:dim(SDM$trace$z)[1]){

  R_s_ObsP[j, , ] = cor( (Y- pnorm(SDM$trace$z[j,,])) /sqrt(pnorm(SDM$trace$z[j
      ,,])*(1-pnorm(SDM$trace$z[j,,])))) 
```

```
}
## Fit JSDM, R is a correlation matrix

JSDM = BC(Y, X, model = "full", its = nits, burn = nburn, thin = 1, verbose=2)
```

References

Choler, P. (2005). Consistent shifts in alpine plant traits along a mesotopographical gradient. *Arctic Antarctic and Alpine Research - ARCT ANTARCT ALP RES*, 37:444–453.

Dray, S. and Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software, Articles*, 22(4):1–20.

Golding, N. and Harris, D. J. (2015). *BayesComm: Bayesian Community Ecology Analysis*. R package version 0.1-2.

D Supplementary figures for Box 2

D.1 Niche estimation

Estimated regression coefficients for all species. JSDMs' estimates are in green, SDMs' estimates in red. For each species, the estimate of the intercept is represented in the last line, while the two coefficients for the snow melt day are in the first and second line.

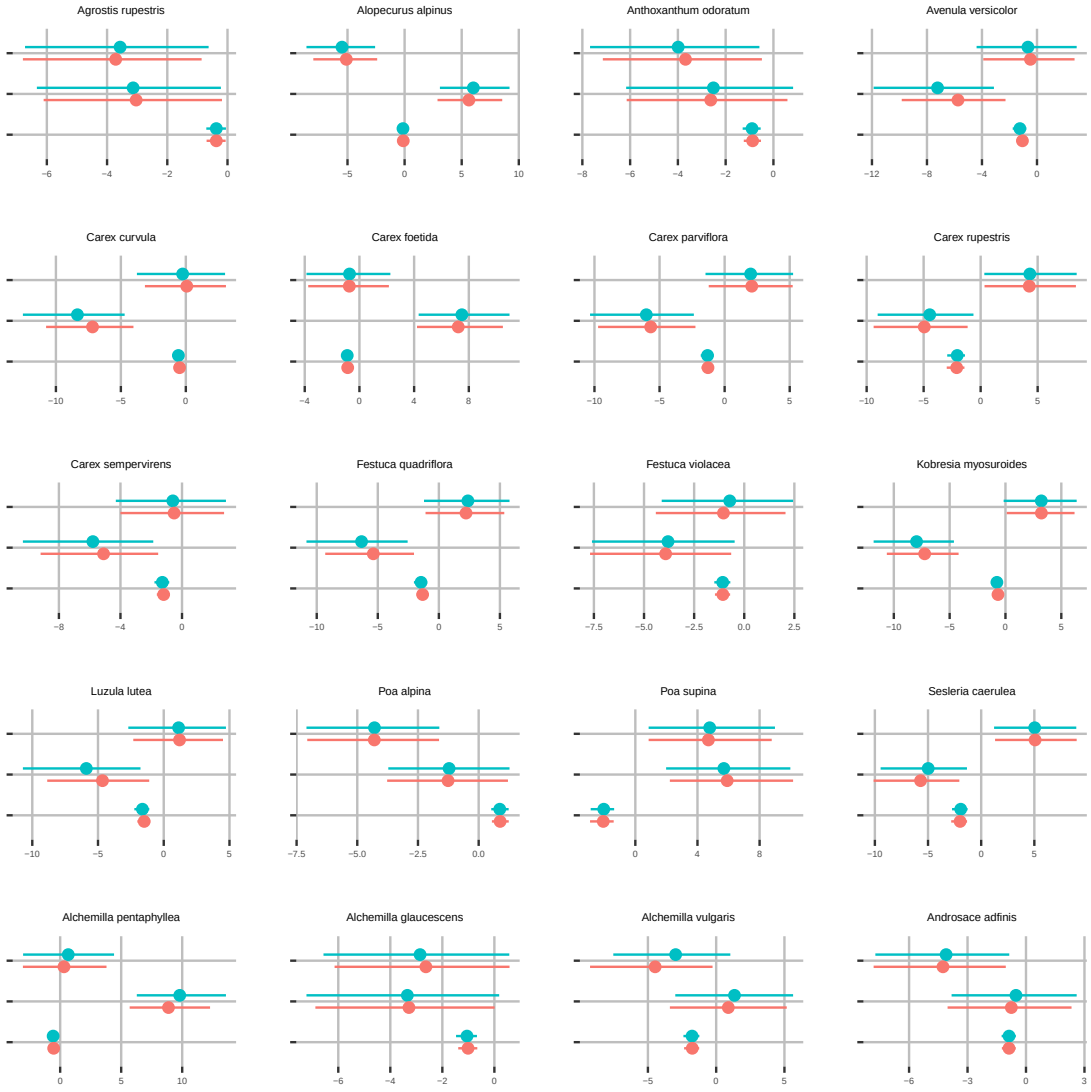

Figure A. 2: Posterior mean and 95% credible intervals for species from 1 to 20 (in alphabetical order). JSDMs' estimates are in green, SDMs' estimates in red. For each species, the estimate of the intercept is represented in the last line, while the two coefficients for the snow melt day are in the first and second line.

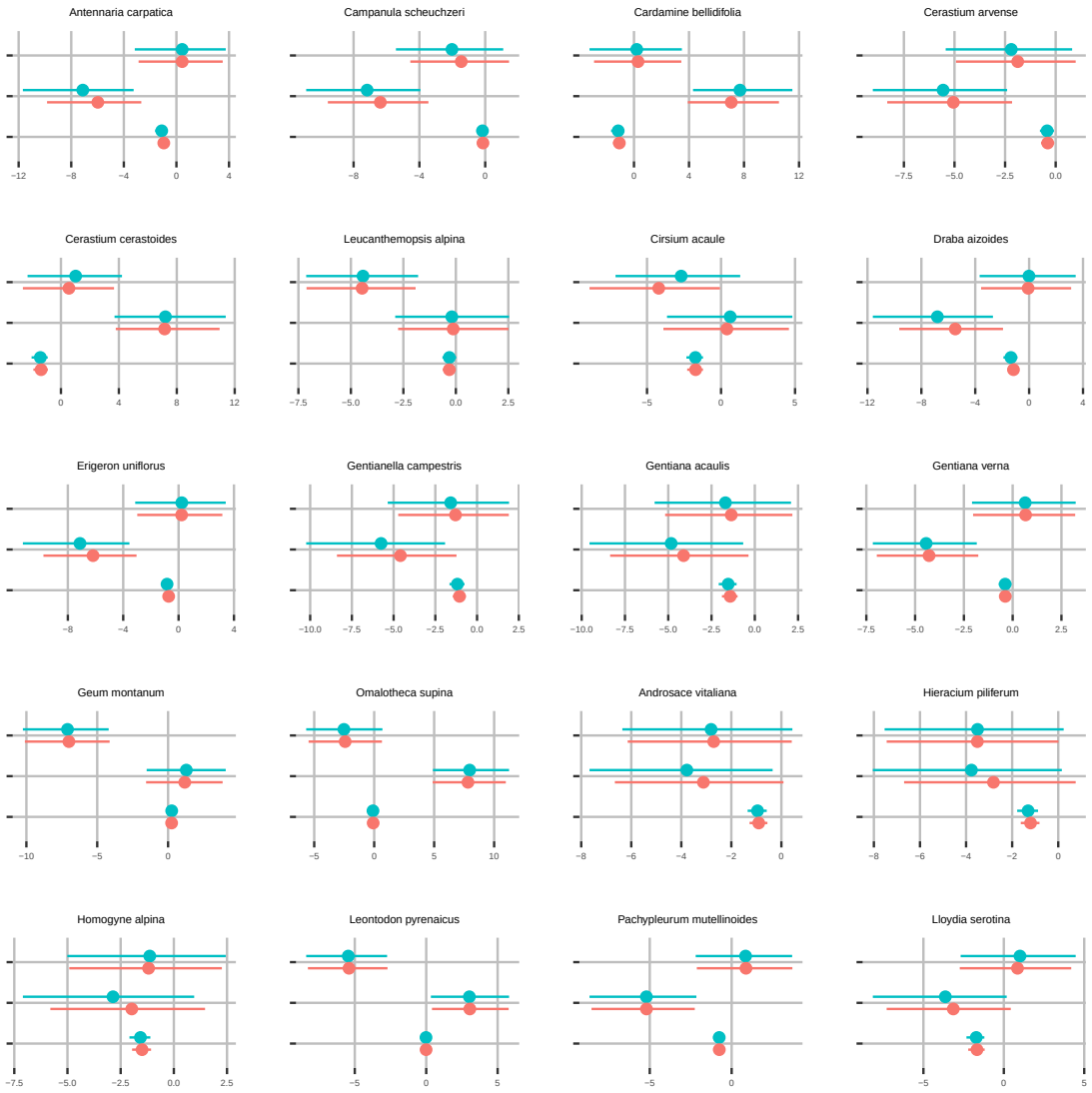

Figure A. 3: Posterior mean and 95% credible intervals for species from 21 to 40 (in alphabetical order). JSDMs' estimates are in green, SDMs' estimates in red. For each species, the estimate of the intercept is represented in the last line, while the two coefficients for the snow melt day are in the first and second line.

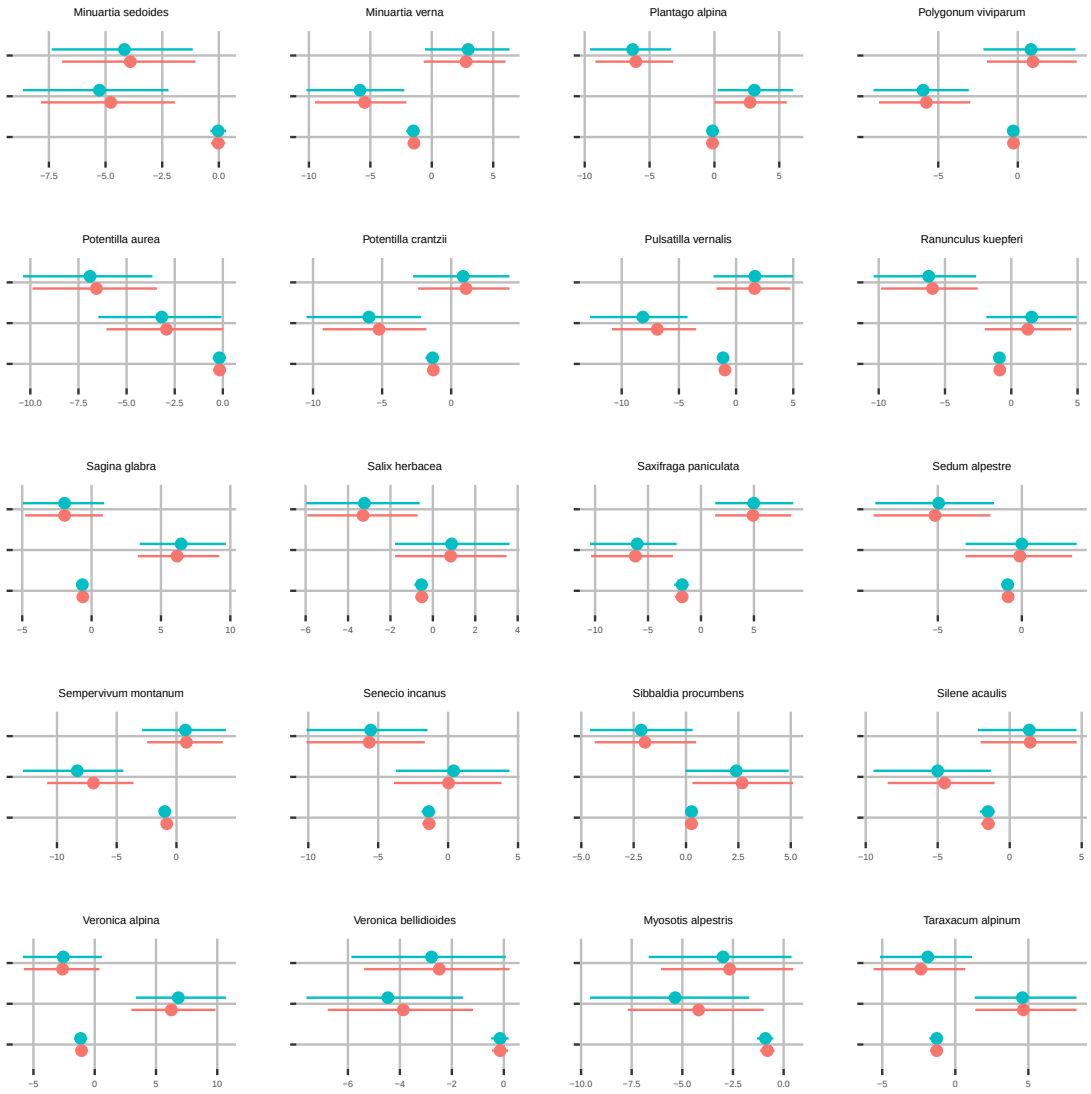

Figure A. 4: Posterior mean and 95% credible intervals for species from 41 to 60 (in alphabetical order). JSDMs' estimates are in green, SDMs' estimates in red. For each species, the estimate of the intercept is represented in the last line, while the two coefficients for the snow melt day are in the first and second line.

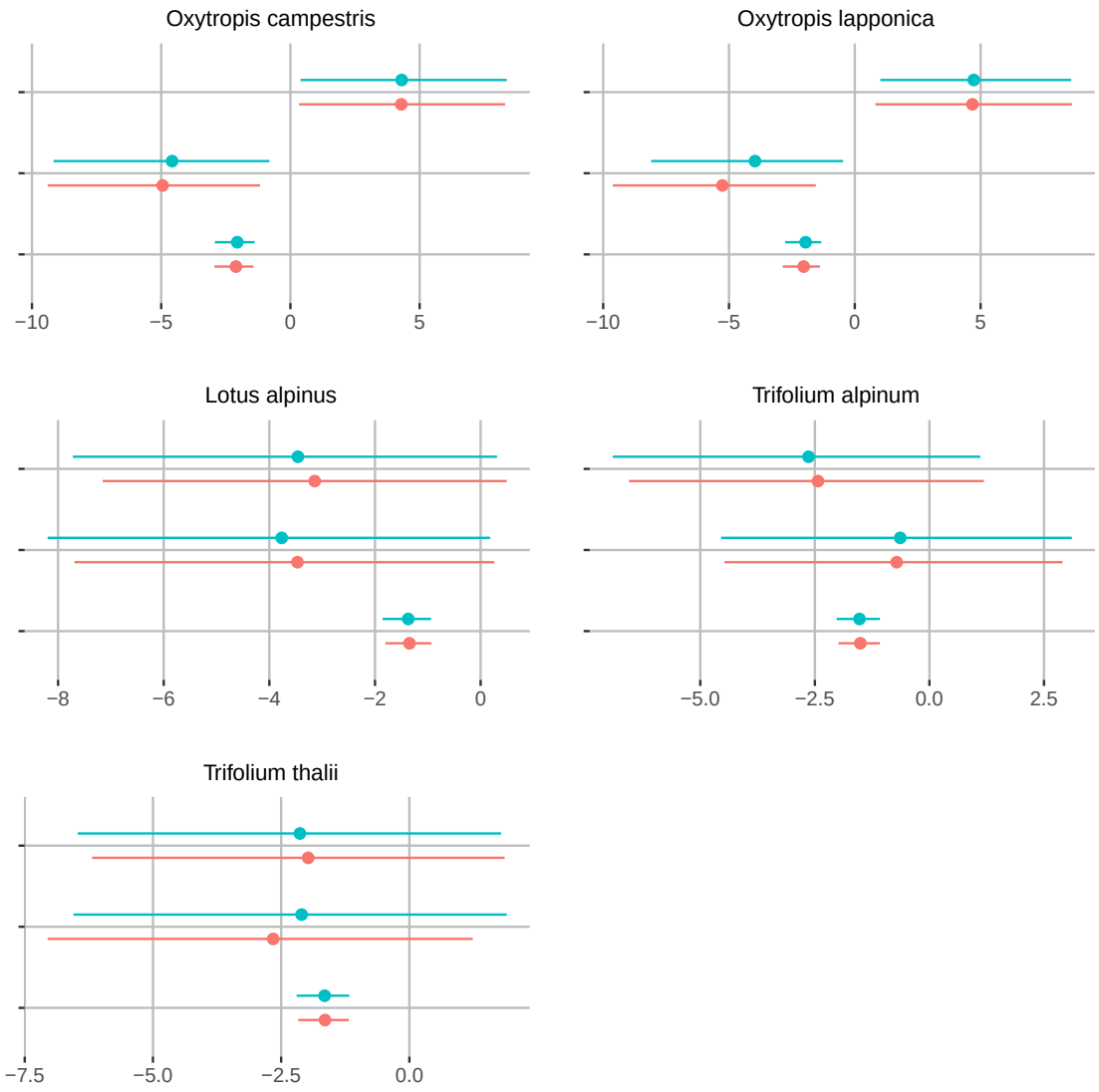

Figure A. 5: Posterior mean and 95% credible intervals for species from 61 to 65. JSDMs' estimates are in green, SDMs' estimates in red. For each species, the estimate of the intercept is represented in the last line, while the two coefficients for the snow melt day are in the first and second line.

D.2 Residual correlation

D.2.1 Raw residuals

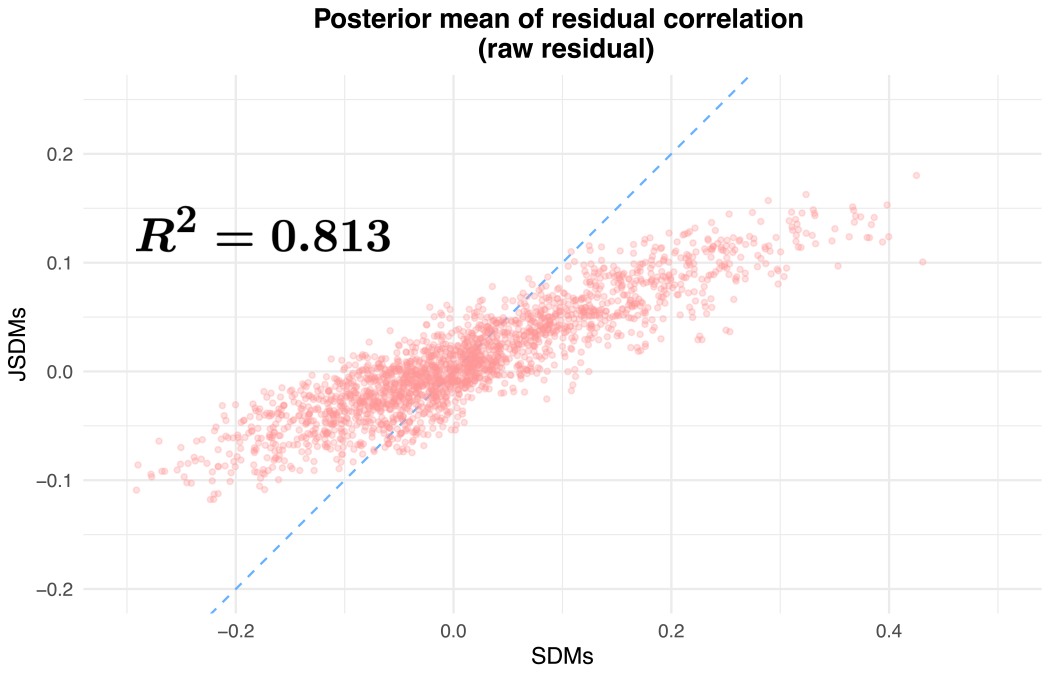

Figure A. 6: The points represent the posterior mean of the elements of the residual correlation matrix of SDMs (calculated as the raw residuals: the difference between observed presence-absence and prediction) on the x-axis and JSDMs on the y-axis. The dashed line is the 1:1 line. We have $R^2 = 0.813$, and the credible intervals are of the same sign (both negative, both positive or both overlapping zero) 86% of the times.

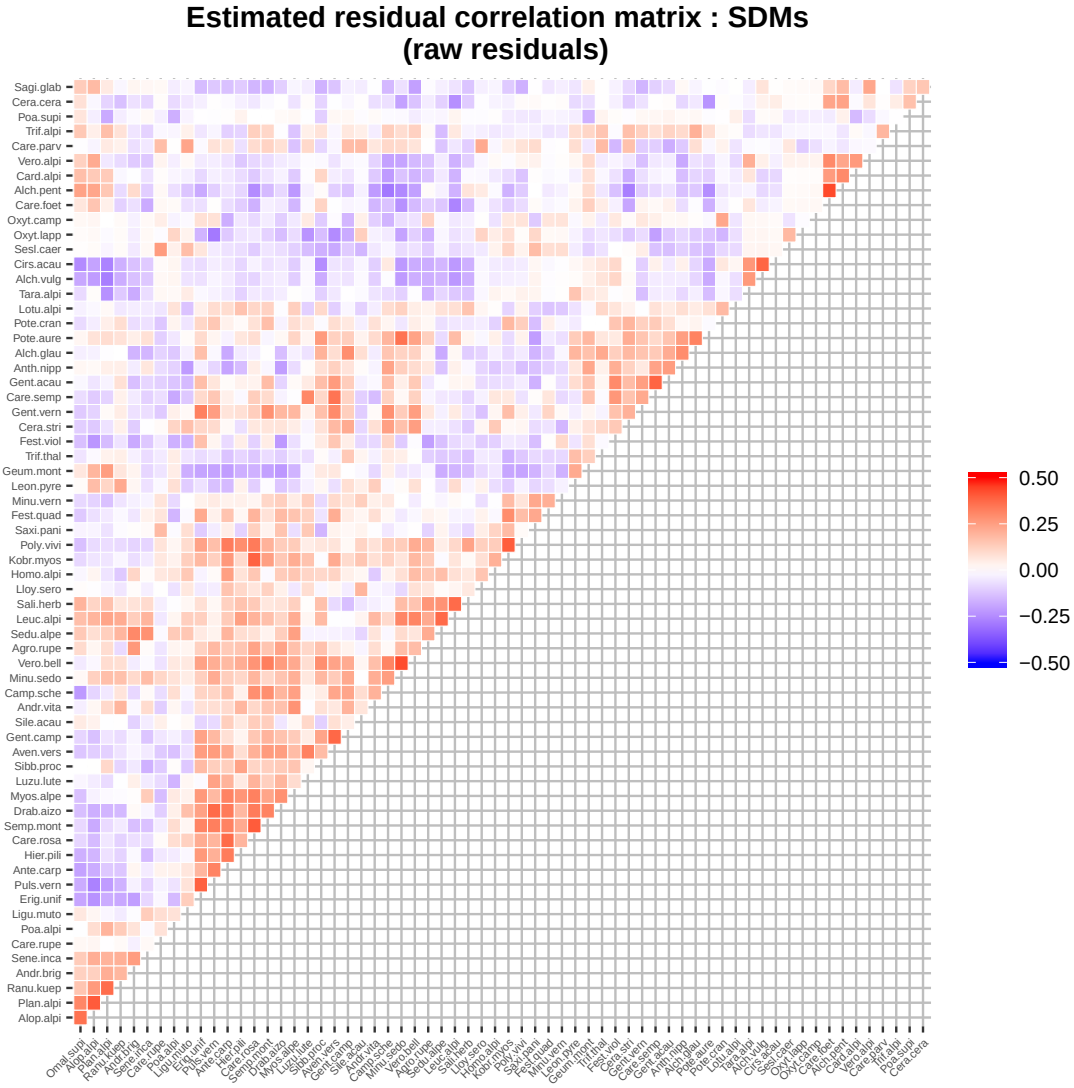

Figure A. 7: Posterior means of the residual correlation matrix for SDMs, where the residuals are calculated as the difference between the predicted probability of presence and observed occurrences.

D.2.2 Residuals at the latent variable level

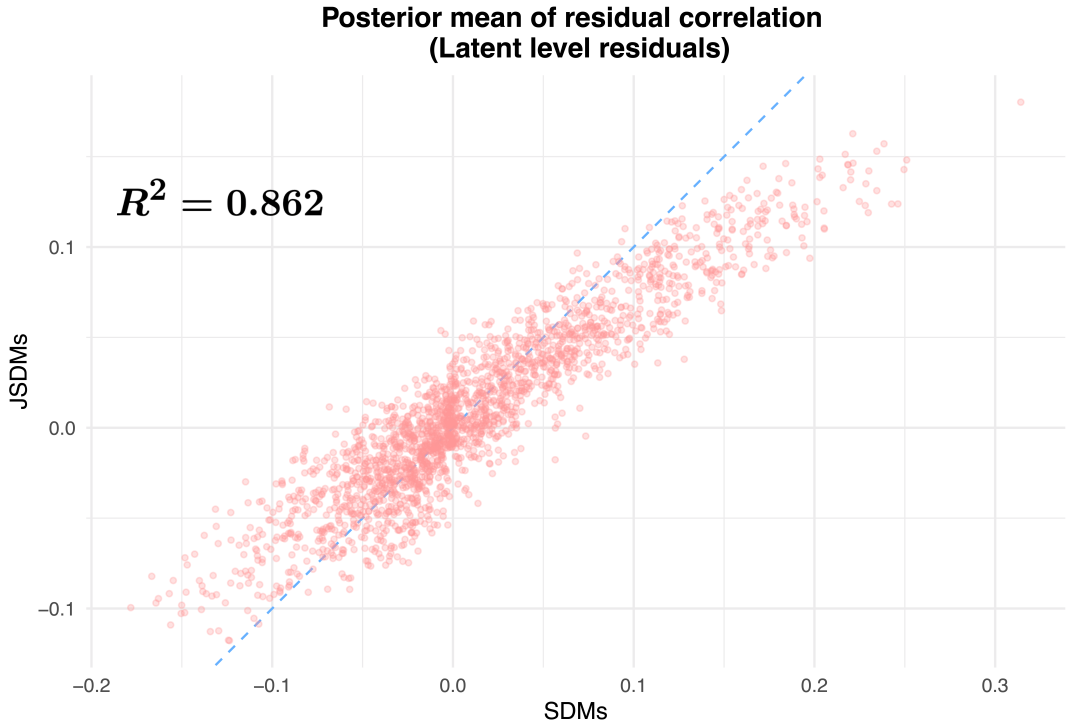

Figure A. 8: The points represent the posterior mean of the elements of the residual correlation matrix of SDMs (where the residuals are calculated at the latent variable level as the difference between the inferred latent variable and the regression term) on the x-axis and JSDMs on the y-axis. The dashed line is the 1:1 line. We have $R^2 = 0.862$, and the credible intervals are of the same sign (both negative, both positive or both overlapping zero) 98% of the times.

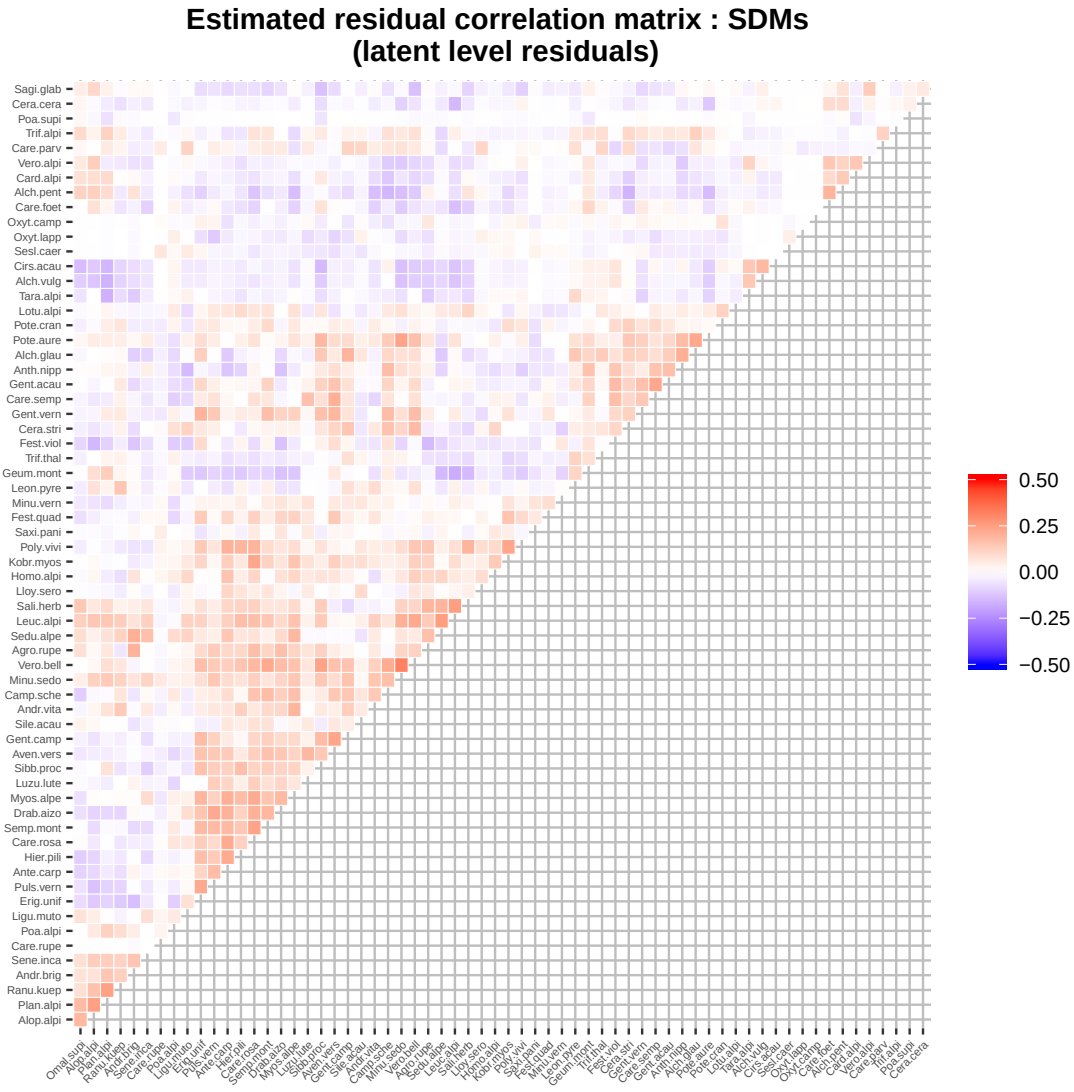

Figure A. 9: Posterior means of the residual correlation matrix for SDMs, where the residuals are calculated as the difference between the latent variable and the regression terms, as in JSDMs.

D.2.3 Pearson Residuals

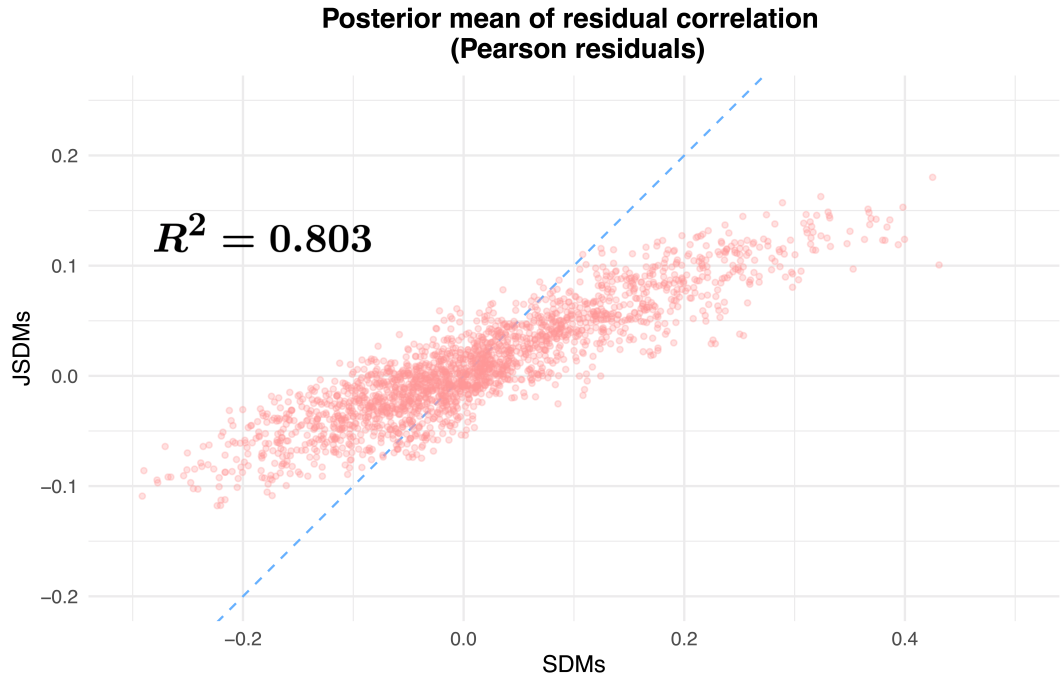

Figure A. 10: The points represent the posterior mean of the elements of the residual correlation matrix of SDMs (where the residuals are the Person residuals: the difference between observed presence-absence and prediction, normalised by the variance of the prediction) on the x-axis and JSDMs on the y-axis. The dashed line is the 1:1 line. We have $R^2 = 0.803$, and the credible intervals are of the same sign (both negative, both positive or both overlapping zero) 74% of the times.

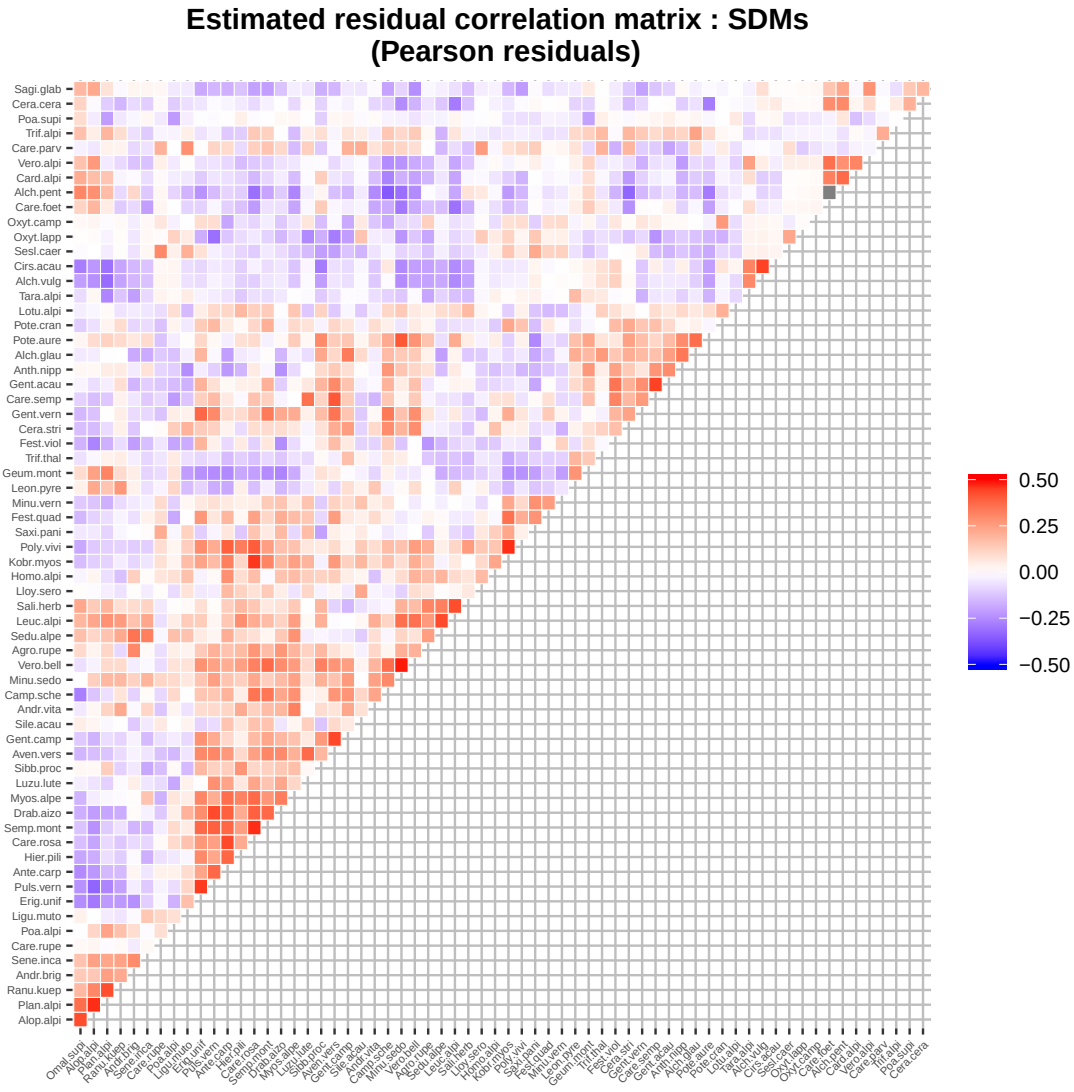

Figure A. 11: Posterior means of the residual correlation matrix for SDMs, where the residuals are calculated as the Pearson residuals.