Bayesian Analysis (0000) **00**, Number 0, pp. 1

Appendices: Joint Species distribution modeling: dimension reduction using Dirichlet processes

Daniel Taylor-Rodríguez*, Kimberly Kaufeld† Erin M. Schliep‡ James S. Clark§ and Alan E. Gelfand¶

1 Appendix A: Sampling algorithm

The hierarchical setup described in the previous section yields the joint distribution

$$(\sigma_\varepsilon^2)^{-(\frac{nS}{2}+1)}\left\{\prod_{i=1}^{n}\exp\left[-\frac{1}{2\sigma_\varepsilon^2}||\mathbf{V}_i-\mathbf{B}\mathbf{x}_i-Q(\mathbf{k})\mathbf{Z}\mathbf{w}_i||^2\right]\times\exp\left[-\frac{1}{2}\mathbf{w}_i'\mathbf{w}_i\right]\right\}\times$$
$$|\mathbf{D_z}|^{-\frac{1}{2}}\left\{\prod_{j=1}^{N}\exp\left[-\frac{1}{2}Z_j'\mathbf{D_z}^{-1}Z_j\right]\right\}\times\left\{\prod_{l=1}^{S}\sum_{j=1}^{N}p_j\delta_j(k_l)\right\}\pi(\boldsymbol{p}|\mathbf{0},\alpha)\times$$
$$\left\{\mathcal{IW}\left(\mathbf{D_z}\Big|2+r-1,4\,\mathrm{diag}\left\{\frac{1}{\eta_1},\ldots,\frac{1}{\eta_r}\right\}\right)\times\prod_{h=1}^{r}\mathcal{IG}\left(\eta_h\Big|\frac{1}{2},\frac{1}{10^4}\right)\right\}.$$

The full conditional densities arising from the joint density above are

1) $\left[\mathbf{Z}|\mathbf{D_z},\mathbf{B},\mathbf{W},\sigma_\varepsilon^2,\mathbf{V}\right]$**:** The posterior for each row of $\mathbf{Z}$ depends on whether or not the row considered was chosen to be at least one row from $\mathbf{A}$. That is, for $j=1,2,\ldots,N$

1. If $j\notin\mathbf{k}$, sample $Z_j\sim N_r(\mathbf{0},\mathbf{D_z})$.
2. Otherwise, if $j\in\mathbf{k}$, let $S_j=\{l=1,\ldots,S \text{ s.t. } k_l=j\}$ and let $|S_j|$ denote the cardinality of S_j. Using these definitions the full conditional distribution for Z_j is given by

$$Z_j|\mathbf{D_z},\mathbf{B},\mathbf{W},\sigma_\varepsilon^2,\mathbf{V}\sim N_r\left(\boldsymbol{\mu}_{\mathbf{z}_l}\ ,\ \Sigma_{Z_j}\right)$$

*Postdoctoral Associate, SAMSI/Duke University, Research Triangle Park, NC 27709. taylor-rodriguez@samsi.info

†Postdoctoral Researcher, SAMSI/North Carolina State University, Research Triangle Park, NC 27709. kkaufeld@samsi.info

‡Assistant Professor, Department of Statistics, University of Missouri, Columbia, MO 65211. schliepe@missouri.edu

§Professor, Nicholas School of the Environment, Department of Statistical Science, Duke University, Durham, NC 27708. jimclark@duke.edu

¶Professor, Department of Statistical Science, Duke University, Durham, NC 27708. alan@stat.duke.edu

where $\Sigma_{Z_j} = \left(\frac{|S_j|}{\sigma^2_\varepsilon}\mathbf{W}^T\mathbf{W} + \mathbf{D}_\mathbf{z}^{-1}\right)^{-1}$, $\boldsymbol{\mu}_{Z_j} = \Sigma_{Z_j}\mathbf{W}^T\frac{1}{\sigma^2_\varepsilon}\sum_{l\in S_j}(\mathbf{V}^{(l)} - \mathbf{X}\beta_l)$, and finally, $\mathbf{y}^{(l)}$ and β_l are the lth column of matrix $\mathbf{Y}$ and the lth row of $\mathbf{B}$, respectively.

2) $\left[\mathbf{W}|\mathbf{B},\mathbf{A},\sigma^2_\varepsilon,\mathbf{V}\right]$**:** As the rows of $\mathbf{W}$ are independent, the posterior for each of the rows is

$$\mathbf{w}_i|\mathbf{B},\mathbf{A},\sigma^2_\varepsilon,\mathbf{V}_i \sim N_r\left(\Sigma_\mathbf{W}\mathbf{A}\frac{1}{\sigma^2_\varepsilon}(\mathbf{V}_i - \mathbf{B}\mathbf{x}_i)\right)$$

where $\Sigma_\mathbf{W} = \left(\frac{1}{\sigma^2_\varepsilon}\mathbf{A}^T\mathbf{A} + I_r\right)^{-1}$

3) $\left[\mathbf{k}|\boldsymbol{p},\mathbf{B},\mathbf{Z},\sigma^2_\varepsilon,\mathbf{V}\right]$**:** For the vector of labels $\mathbf{k}$, the full conditional is

$$\left[\mathbf{k}|\boldsymbol{p},\mathbf{B},\mathbf{Z},\sigma^2_\varepsilon,\mathbf{V}\right] = \prod_{l=1}^{q}\left\{\sum_{j=1}^{N} p_{lj}\delta_j(k_l)\right\}$$

with $p_{lj} \propto p_j \times \exp\left[-\frac{1}{2\sigma^2_\varepsilon}||\mathbf{V}^{(l)} - \mathbf{X}\beta_l - \mathbf{W}Z_j||^2\right]$.

4) $[\boldsymbol{p}|\mathbf{k}]$**:** The full conditional posterior for $\boldsymbol{p}$, given conjugacy of the $\mathcal{GD}$ distribution with multinomial sampling, the draws of $\boldsymbol{p}$ are

$$\begin{aligned} p_1 &= \xi_1, \\ p_j &= (1-\xi_1)\cdots(1-\xi_{j-1})\xi_j, \text{ for } j = 2,3,\ldots,N-1, \\ p_N &= 1 - \sum_{j=1}^{N-1} p_j, \end{aligned}$$

with $\xi_j \overset{ind}{\sim} \text{Beta}\left(\frac{\alpha}{N} + \sum_{l=1}^{S} I_{(k_l=j)}, \frac{N-j}{N}\alpha + \sum_{s=j+1}^{N}\sum_{l=1}^{S} I_{(k_l=s)}\right)$ for $j = 1,\ldots,N-1$.

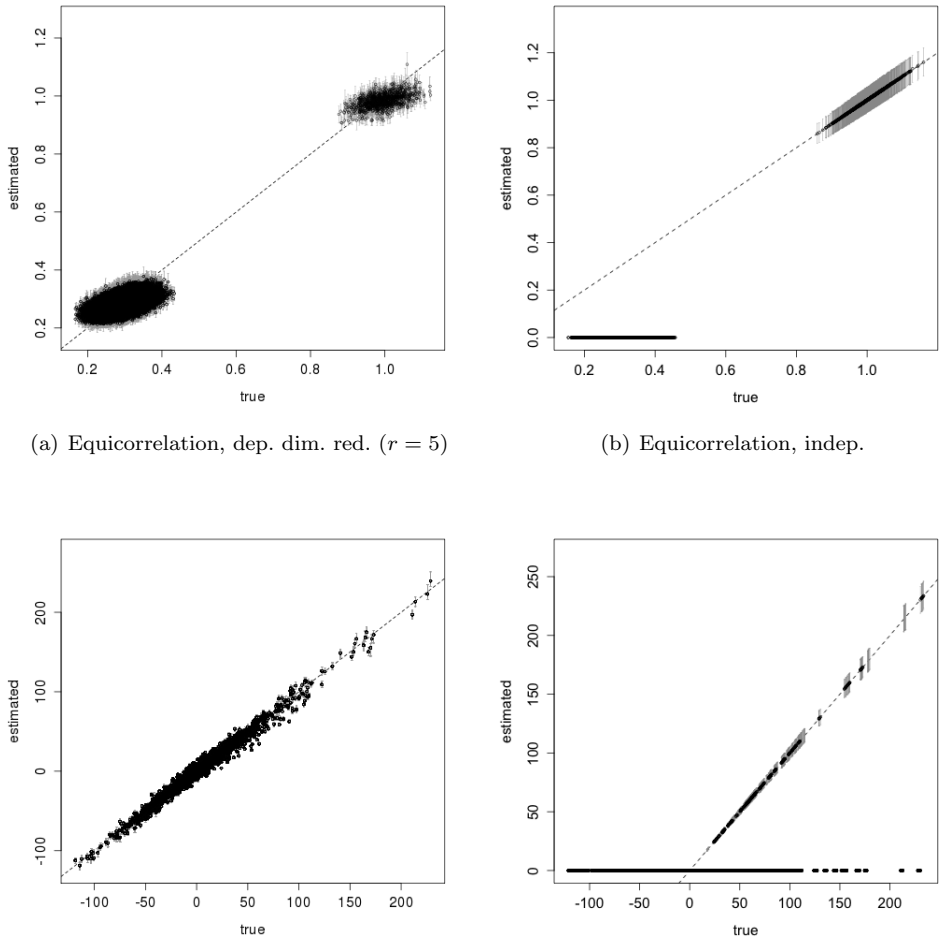
5) $\left[\sigma^2_\varepsilon|\mathbf{A},\mathbf{W},\mathbf{B},\mathbf{Y}\right]$**:** By conjugacy of the prior for σ^2_ε with the normal likelihood, the full conditional is

$$\sigma^2_\varepsilon|\mathbf{A},\mathbf{W},\mathbf{B},\mathbf{V} \sim \mathcal{IG}\left(\frac{nS+\nu}{2} + 1, \frac{\sum_{i=1}^{n}||\mathbf{V}_i - \mathbf{B}\mathbf{x}_i - \mathbf{A}\mathbf{w}_i||^2}{2} + \frac{\nu}{G^2}\right).$$

6) $[\mathbf{D}_\mathbf{z}|\mathbf{Z}]$**:** By conjugacy of the prior $\mathbf{D}_\mathbf{z}$ with the normal prior for $\mathbf{Z}$, the full conditional for $\mathbf{D}_\mathbf{z}$ is

$$\mathbf{D}_\mathbf{z}|\mathbf{Z} \sim \mathcal{IW}\left(\mathbf{D}_\mathbf{z}\middle|2 + r + N - 1, \mathbf{Z}'\mathbf{Z} + 4\,\text{diag}\left\{\frac{1}{\eta_1},\ldots,\frac{1}{\eta_r}\right\}\right).$$

2 Appendix B: Additional simulation plots from Section 6



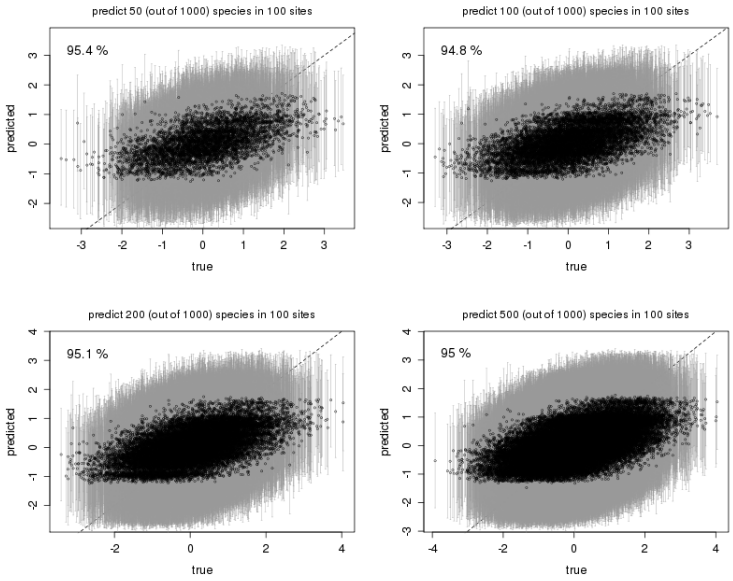
(a) Equicorrelation, dep. dim. red. ($r = 5$)

(b) Equicorrelation, indep.

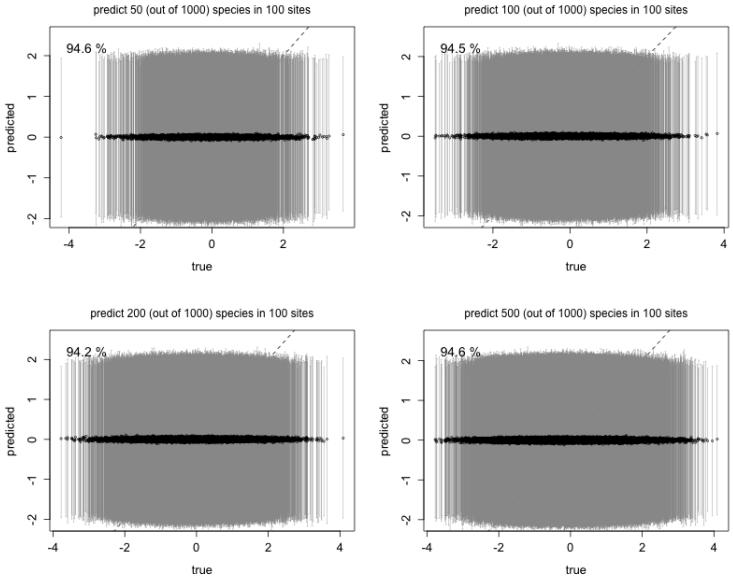
(c) Clustered, dep. dim. red. ($r = 5$)

(d) Clustered, indep.

Figure 1: True vs estimated covariance parameters for one out-of-sample dataset with the equicorrelated and clustered covariance structures for $n = 3000, S = 1000$ comparing the dimension reduced dependent approach (with $r = 5$) and the independence model

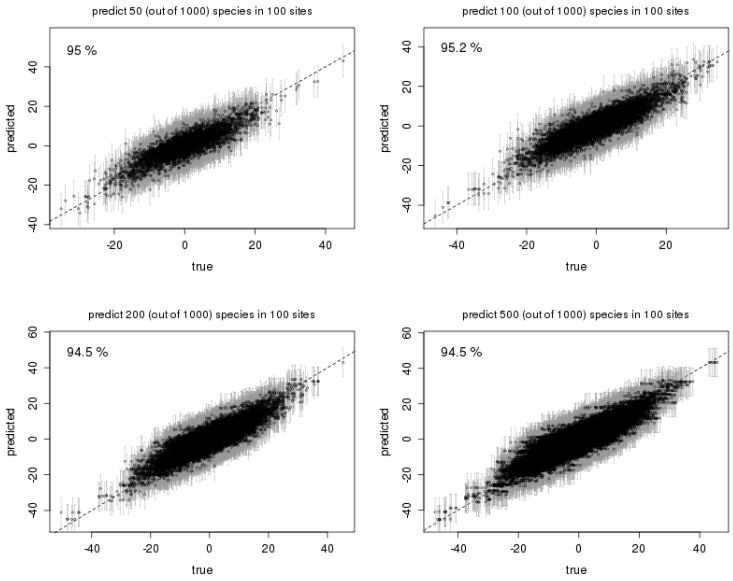


(a) Dependence with dimension red. ($r = 5$), equicorrelated covariance structure

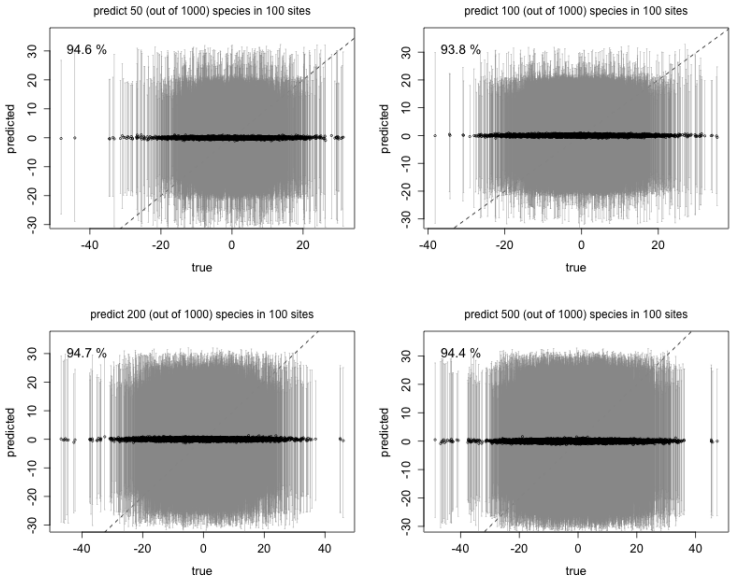


(b) Independence, equicorrelated covariance structure

Figure 2: Mean and 95% credible set for the conditional prediction for 50, 100, 200 and 500 species conditional on 950, 900, 800 and 500 species, respectively, in 100 plots with true data generated from equicorrelated covariance structure for $n = 3000, S = 1000$. Comparison between the dimension reduced dependent approach (with $r = 5$) and the independence model. On each of the plots the empirical coverage for the prediction intervals is provided.



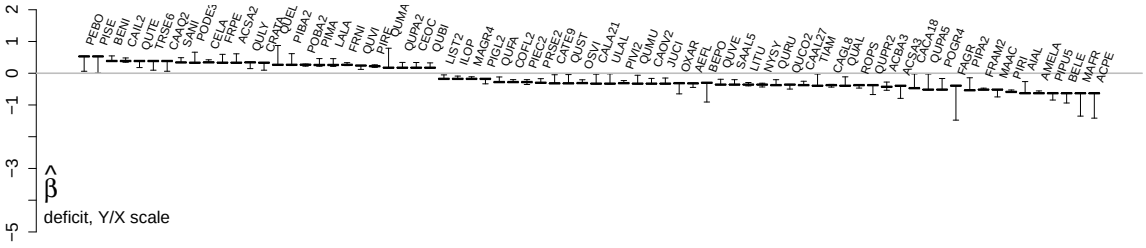
(a) Joint with dimension red. ($r = 5$), clustered covariance structure



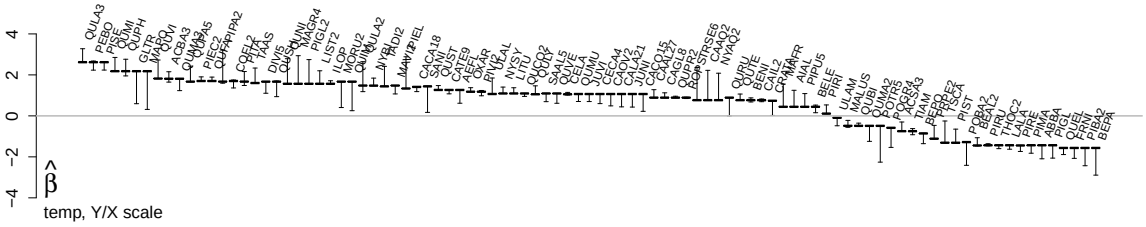
(b) Independence, clustered covariance structure

Figure 3: Mean and 95% credible set for the prediction of 50, 100, 200 and 500 species conditional on 950, 900, 800 and 500 species, respectively, in 100 plots with true data generated from clustered covariance structure for $n = 3000, S = 1000$. Comparison between the dimension reduced dependent approach (with $r = 5$) and the independence model. On each of the plots the empirical coverage for the prediction intervals is provided.

3 Appendix C: Fitted regression coefficients for the FIA data



(a) deficit



(b) temperature

Figure 4: Mean and 95% credible sets for the "significant" fitted regression coefficients in the FIA data.