Inferring joint species distribution models using variational Bayes: example on the Borneo forest

Journée AppliBUGS, Agroparistech

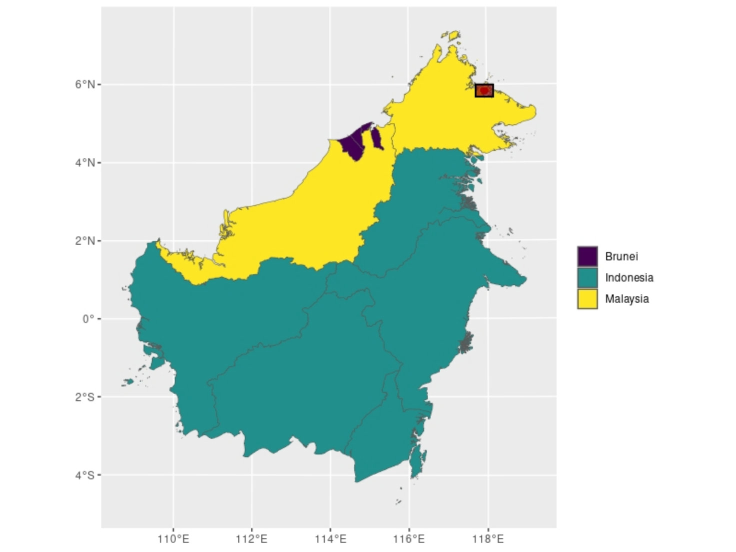
Pierre Gloaguen[1] Eric Parent Giacomo Sellan Achille Thin

19 décembre 2023

[1] Université Bretagne Sud

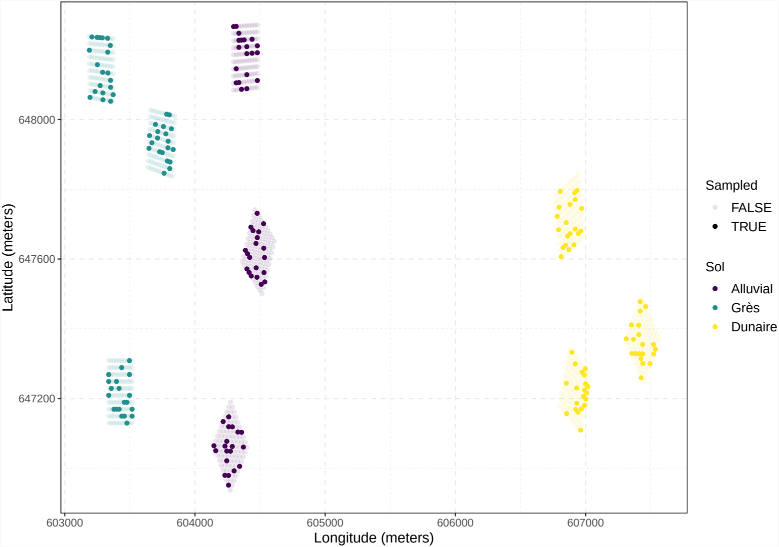
Data set and questions

Study area: Borneo forest



Experimental design

- 900 sites where trees abundances are recorded;
- 180 sites where soil chemistry is recorded.



Abundance data

- $n = 180$ sampling sites, $p \approx 200$ plant species are counted, giving a matrix $\mathbf{Y}$ **abundance** data;

	Dehaasia caesia	*Polyalthia canangioides*	*Dipterocarpus acutangulus*	*Aglaia glabriflora*
Site 1	7	0	0	0
Site 2	7	0	7	0
Site 3	7	0	0	0
Site 4	7	0	0	0
Site 5	6	0	0	0
Site 6	6	0	0	0
Site 7	6	0	0	0
Site 8	6	0	1	0
Site 9	6	0	4	0
Site 10	6	0	0	0

Soil chemistry data

- 20 soil **covariates** are measured giving a matrix **X**

	Sol	pH	Eau	C	N	NO3	NH4	Ac	Al	Ca	Mg	K
Site 3	Alluvial	4.58	4.02	0.62	0.11	2.50	2.32	6.98	6.20	0.15	0.22	0.11
Site 6	Alluvial	4.51	3.01	0.37	0.05	4.86	4.17	3.92	3.14	0.03	0.42	0.05
Site 5	Grès	4.88	2.02	0.73	0.06	1.77	6.13	2.55	2.04	0.00	0.08	0.08
Site 7	Grès	4.72	2.07	0.52	0.04	2.16	6.88	3.17	2.65	0.00	0.09	0.07
Site 1	Dunaire	4.94	1.33	0.89	0.06	2.71	1.02	1.82	1.49	0.08	0.06	0.05
Site 2	Dunaire	4.74	1.63	0.76	0.05	0.47	0.97	1.60	1.22	0.12	0.17	0.16
Site 4	Dunaire	4.80	0.80	0.87	0.04	0.00	1.50	1.21	0.86	0.08	0.05	0.03
Site 8	Dunaire	5.04	1.45	0.89	0.06	1.05	1.38	2.16	1.57	0.07	0.05	0.05
Site 9	Dunaire	4.82	1.27	0.83	0.05	0.00	1.38	0.96	0.71	0.10	0.18	0.08
Site 10	Dunaire	4.76	1.68	0.95	0.07	0.00	2.32	3.20	2.70	0.10	0.07	0.08

- Additionnally, plant phylogeny and some species' traits can be obtained. . .

Joint species distribution models

Joint species distribution modelling

A classical statistical approach

- **Y** is a matrix of **counts** ⇒ **Poisson distribution**;

  $$\mathbf{Y} \sim \mathcal{P}\text{oisson}(\exp(\mathbf{Z})).$$

  where **Z** is a matrix having the same dimensions as **Y** (the exponential is taken entrywise).
- **Z** will be a **linear predictor**;
- **X** will be seen as **features** for this predictor, and will be linked to **Z**;

Model on the linear predictor

- **Z** is a matrix $n \times p$ (# of sites × # of species), modelling the **intensity** of presence of species per unit;
- We suppose it is random, with **Normal** distribution;
- A Matrix Normal random variable is characterized by:
  - Its expected value (mean intensity) **M**;
  - Its covariance between rows (sites) Σ_{sites} (matrix $n \times n$);
  - Its covariance between columns (species) $\Sigma_{species}$ (matrix $p \times p$);

$$\mathbf{Z} \sim \mathcal{MN}\left(\mathbf{M}, \overset{\text{rowwise cov.}}{\Sigma_{sites}}, \underset{\text{colwise cov.}}{\Sigma_{species}}\right)$$

Model on the linear predictor

- **Z** is a matrix $n \times p$ (# of sites × # of species), modelling the **intensity** of presence of species per unit;
- We suppose it is random, with **Normal** distribution;
- A Matrix Normal random variable is characterized by:
  - Its expected value (mean intensity) **M**;
  - Its covariance between rows (sites) Σ_{sites} (matrix $n \times n$);
  - Its covariance between columns (species) $\Sigma_{species}$ (matrix $p \times p$);

$$\mathbf{Z} \sim \mathcal{MN}\left(\mathbf{M}, \overset{\text{rowwise cov.}}{\Sigma_{sites}}, \underset{\text{colwise cov.}}{\Sigma_{species}}\right)$$

Model on M, the expected log-abundance

- The expected intensity is linked to environment covariates **X**:

$$\mathbf{M} = \mathbf{X}\beta$$

  Where β is a, **unknown** $n_{cov} \times p$ (# of covariates × # of species) matrix giving the unknown **response of species to environnement**.

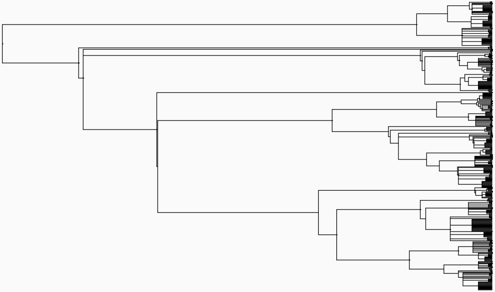
- Each species is characterized by its column in β: **its niche**.

Structuring the niches

- Suppose we have access to other data about species:
- Species traits in a matrix **T**:

Espece	TxCroissance	Densite	Hauteur
Strychnos borneensis	0.008	0.750	19.749
Dysoxylum indet	0.027	0.585	8.588
Memecylon indet	0.013	0.783	8.692
Cratoxylum cochinchinense	0.025	0.670	9.894
Sterculia stipulata	0.027	0.365	10.087

- Phylogeny, giving a correlation matrix **C**:



Structuring the niches

- The matrix β stacks the niches of species (vector of responses to environnement);
- Assume that:
  - The **traits** might affect this response to environment (i.e. similar traits lead to similar niche);
  - The response to environment might be correlated between species, because of **phylogeny**.

Structuring the niches

- The matrix β stacks the niches of species (vector of responses to environnement);
- Assume that:
  - The **traits** might affect this response to environment (i.e. similar traits lead to similar niche);
  - The response to environment might be correlated between species, because of **phylogeny**.

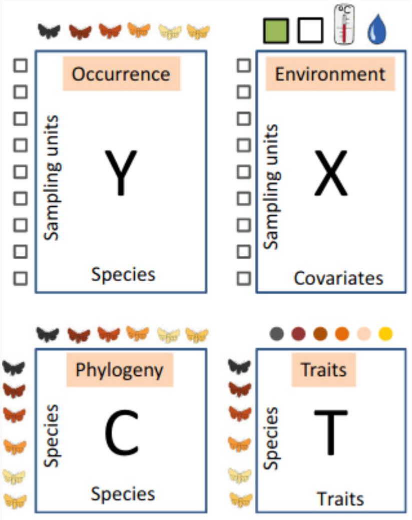
Formally, β is assumed to be a Matrix Normal random variable such that:

$$\boldsymbol{\beta} \sim \mathcal{MN}\left(\boldsymbol{\Gamma}\mathbf{T}', \eta^2\mathbf{I}_{n_{cov}}, \rho\mathbf{C} + (1-\rho)\mathbf{I}_p\right)$$

- $\boldsymbol{\Gamma}$ is a $n_{cov} \times n_t$ (# of covariates × # of traits) describing how the response to environment is structured by the traits; **Do the species niches are correlated to species traits?**
- **C** is the correlation matrix induced by the **phylogeny**;
- $0 \leq \rho \leq 1$ is the importance weight of **phylogeny** in the columns correlation of $\boldsymbol{\beta}$.

In summary: modelling fixed effects

Retrieving the nice framework of Ovaskainen et Abrego (2020)



Modelling residuals

- So far:

$$\begin{aligned}
\mathbf{Y} &\sim \mathcal{P}\text{oisson}(\exp(\mathbf{L})) && \text{Abundance distribution}\\
\mathbf{Z} &\sim \mathcal{MN}\left(\mathbf{X}\beta, \Sigma_{sites}, \Sigma_{species}\right) && \text{Model for the presence intensities}\\
\beta &\sim \mathcal{MN}\left(\boldsymbol{\Gamma}\mathbf{T}', \eta^2\mathbf{I}_{n_{cov}}, \rho\mathbf{C} + (1-\rho)\mathbf{I}_p\right) && \text{Model for the niches}
\end{aligned}$$

- What about Σ_{sites} (the covariance between intensities in sampling sites)?
  - Classical spatial structure can be added (as in geostatistics);
- What about $\Sigma_{species}$ (the covariance between intensities of species)?
  - If environment explained all, residual species intensities would be independant!
  - However, some species **cooccurence** might remain!
  - We might want to model the structure of this covariance matrix.

Modelling residual cooccurence

- $\Sigma_{species}$ is $p \times p$, thus resulting in $\frac{p(p+1)}{2}$ free parameters which can quickly becomes large;
- One can impose a *low rank* structure over $\Sigma_{species}$;

Modelling residual cooccurence

- $\Sigma_{species}$ is $p \times p$, thus resulting in $\frac{p(p+1)}{2}$ free parameters which can quickly becomes large;
- One can impose a *low rank* structure over $\Sigma_{species}$;

Probabilistic PCA approach

- We will write (in the spirit of PCA):

$$\Sigma_{species} = \text{diag}\boldsymbol{\sigma}^{\mathbf{2}} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{T},$$

where $\boldsymbol{\Lambda}$ is a matrix of size $p \times q$, where $q < p$.

Modelling residual cooccurence

- $\Sigma_{species}$ is $p \times p$, thus resulting in $\frac{p(p+1)}{2}$ free parameters which can quickly becomes large;
- One can impose a *low rank* structure over $\Sigma_{species}$;

Probabilistic PCA approach

- We will write (in the spirit of PCA):

$$\Sigma_{species} = \text{diag}\boldsymbol{\sigma}^{\mathbf{2}} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T,$$

where $\boldsymbol{\Lambda}$ is a matrix of size $p \times q$, where $q < p$.

- Equivalently, for the $i-th$ site the $p-$vector of log-intensity $\mathbf{Z}_i$ satisfies:

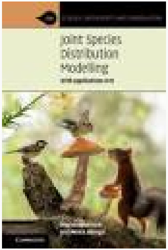
$$\mathbf{Z}_i = \beta \mathbf{X}_i^T + \Lambda \eta_i^T + \epsilon_{\mathbf{i}},$$

where:

- Λ is a $p \times q$ matrix of loads, interpreted as *responses to non-measured covariates*,
- $\eta_i \sim \mathcal{N}_q(0, \mathbf{I}_q)$ a *vector of non-measured covariates*;
- $\varepsilon_i \sim \mathcal{N}_p\left(0, \text{diag}(\sigma_j^2)_{1 \leq j \leq p}\right)$ are well-behaving residuals.

What’s new? So far, nothing!

In a bayesian inference context



- Describes all the framework in Ovaskainen et Abrego (2020);
- `R` package `Hmsc`. Use MCMC sampling, rather slow;

In a maximum likelihood scenario

- Fully described in Chiquet, Robin, et Mariadassou (2019);
- Alternative models for residuals: Chiquet, Mariadassou, et Robin (2021);
- Fully and efficiently implemented in `R` package `PLNmodels`.
- Variational EM methods: no confidence intervals;

Bayesian setting

Designing priors

$$\mathbf{Z}_i - \beta \mathbf{X}_i^T = \Lambda \eta_i^T + \epsilon_{\mathbf{i}}$$

- **Fixed effects priors** on β: gaussian priors (possibly including traits and phylogeny);
- **Variance priors** on $\operatorname{diag}(\sigma_j^2), 1 \leq j \leq p$ Inverse Gamma: Standard
- **Latent variables priors** on $\eta_i, 1 \leq i \leq n$: $\mathcal{N}_p(0, I_p)$ Standard
- **Loading priors** on the $p \times q$ matrix Λ:
  - Incite columns of Λ to become lighter and lighter as their rank increases;
  - Rationale: only few non-measured covariates are needed;

The *multiplicative gamma process shrinkage prior* of Bhattacharya et Dunson (2011) allows for conjugate scheme and penalize high rank columns of Λ.

The multiplicative gamma process shrinkage prior

- Idea: Penalize high rank columns of the $p \times q$ matrix Λ;
  - Let, for $1 \leq j \leq p$ and $1 \leq h \leq q$, $\phi_{j,h} \overset{\text{ind}}{\sim} \mathcal{G}\text{amma}\left(\frac{\nu+1}{2}, \frac{\nu+1}{2}\right)$.
  - Let, for $1 \leq h \leq q$, $\delta_h \overset{\text{ind}}{\sim} \mathcal{G}\text{amma}(\alpha, 1)$ such that $\alpha > 1$ (thus $\mathbb{E}\left[\delta_h\right] > 1$);
  - Then set as prior:

$$\Lambda_{j,h} | \phi_{j,h}, \delta_{1:h} \overset{\text{ind}}{\sim} \mathcal{N}\left(0, \phi_{j,h}^{-1} \prod_{\ell=1}^{h} \delta_\ell^{-1}\right).$$

- When h increases, the last columns of matrix Λ tend to collapse towards 0 (their prior mean), because precision of column h is prompted to increase as $\prod_{\ell=1}^{h} \delta_\ell^{-1}$.
- Remains the prior over α: Non informative, greater than 1.
- **Implementation** of Posterior sampling: Bayesian inference using MCMC. Done so far using the `Hmsc` `R` package. Can easily be re-implemented in `Jags` or `Stan`.

Bayesian inference

- The target posterior has the following form:

$$\left[\mathbf{Z}, \Lambda, \boldsymbol{\sigma}^2, \boldsymbol{\eta}, \phi, \boldsymbol{\delta}, \beta | \mathbf{Y}\right] \propto [\mathbf{Y}|\mathbf{Z}][\mathbf{Z}|\boldsymbol{\eta}, \Lambda, \boldsymbol{\sigma}^2, \beta][\Lambda|\boldsymbol{\delta}, \phi][\beta][\boldsymbol{\sigma}^2][\boldsymbol{\eta}][\boldsymbol{\delta}][\phi]$$

- MCMC can be performed;
- Approximated alternative: Variational bayes inference:

Variational bayes inference

- Target distribution: $p(\theta|\mathbf{Y})$, for $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_d\}$.
- Restriction to a tractable family $q^{\lambda}(\theta)$ parameterized by λ.
- Mean field approximation: For instance:

$$q^{\lambda}(\theta) = \prod_{i=1}^{d} q^{\lambda_i}(\theta_i).$$

- Find λ by maximizing the Evidence lower bound:

$$ELBO(\lambda) = \text{argmax}_{\lambda} \mathbb{E}_{\theta \sim q^{\lambda}} \left[\log \frac{p(\mathbf{Y}, \boldsymbol{\theta})}{q^{\lambda}(\boldsymbol{\theta})} \right]$$

In our case:

$$
\begin{aligned}
\mathrm{ELBO}(\lambda) = & \mathbb{E}_q \left[\log\left([\mathbf{Y}|\mathbf{Z}]\right)\right] \\
& + \mathbb{E}_q \left[\log\left([\mathbf{Z}|\boldsymbol{\eta}, \Lambda, \boldsymbol{\sigma}^2, \beta]\right)\right] \\
& + \mathbb{E}_q \left[\log\left([\beta]\right)\right] \\
& + \mathbb{E}_q \left[\log\left([\boldsymbol{\sigma}^2]\right)\right] \\
& + \mathbb{E}_q \left[\log\left([\boldsymbol{\eta}]\right)\right] \\
& + \mathbb{E}_q \left[\log\left([\Lambda|\boldsymbol{\delta}, \phi]\right)\right] \\
& + \mathbb{E}_q \left[\log\left([\phi]\right)\right] \\
& + \mathbb{E}_q \left[\log\left([\boldsymbol{\delta}]\right)\right] \\
& - \mathbb{E}_q \left[\log q(\mathbf{Z}, \Lambda, \boldsymbol{\sigma}^2, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\delta}, \beta)\right]
\end{aligned}
$$

Optimizing the ELBO

- Coordinate ascent variational inference;
- Successive local optimizations;
- When
- Similar in the spirit as Gibbs sampling;
- By well choosing variational family, some conjugacy appears.
- In that case, at iteration t, for parameter j:

$$\log q^{\lambda_j^{(t)}}(\theta_j) = \mathbb{E}_{\boldsymbol{\theta}_{-j} \sim q^{\lambda_{-j}^{(t-1)}}} [\log p(\mathbf{Y}, \boldsymbol{\theta})]$$

Explicit conjugate results for most components

As an example consider updating $\phi_{j,h}$

The terms implying $\phi_{j,h}$ are the following:

$$\left(\frac{\nu}{2} + \frac{1}{2} - 1\right) \log \phi_{j,h} - \left(\frac{\nu}{2} + 0.5 \times \Lambda_{j,h}^2 \prod_{\ell=1}^{h} \delta_\ell\right) \phi_{j,h}.$$

Therefore, the updates of the Gamma distribution parameters are given by:

$$A^{\phi_{j,h}} = \frac{\nu}{2} + \frac{1}{2}$$

$$B^{\phi_{j,h}} = \frac{\nu}{2} + \frac{1}{2}\left(\left(M^{\Lambda_{j,h}}\right)^2 + V_{h,h}^{\Lambda_j}\right) \prod_{\ell=1}^{h} \frac{A^{\delta_\ell}}{B^{\delta_\ell}}.$$

With the notable exception of $q(Z)$

No closed form expression for updating $Z \Rightarrow$ numerically maximising:

$$\mathbb{E}_{q_Z}\left[\log[Y|Z]\right] + \mathbb{E}_{q_Z}\left[\log[Z|\eta, \Lambda, \Sigma, \beta]\right] - \mathbb{E}_q\left[\log q_Z(Z)\right]$$

For CAVI algorithm, we take $q_Z(Z) = \prod_{i,j} q_Z(Z_{i,j})$ in the normal family.

Up to constant terms (with regards to $q(Z_i j)$), for each (i,j) maximise the partial ELBO function :

Denoting $\mathbb{E}_q(Z_{i,j}) = M$ and $\mathbb{V}ar_q(Z_{i,j}) = V$:

$$Y_{i,j}M - e^{M+\frac{V}{2}} - 0.5\frac{A^{\sigma_j}}{B^{\sigma_j}}M^2 - 0.5\frac{A^{\sigma_j}}{B^{\sigma_j}}V + M \times \frac{A^{\sigma_j}}{B^{\sigma_j}}\left(M^{\eta_i}M^{\Lambda_j} + X_i M^{\beta_j}\right) + \frac{\log|V|}{2}$$

Straightforward **Implementation** through $n \times p$ calls to the R *optim* subroutine

About Z

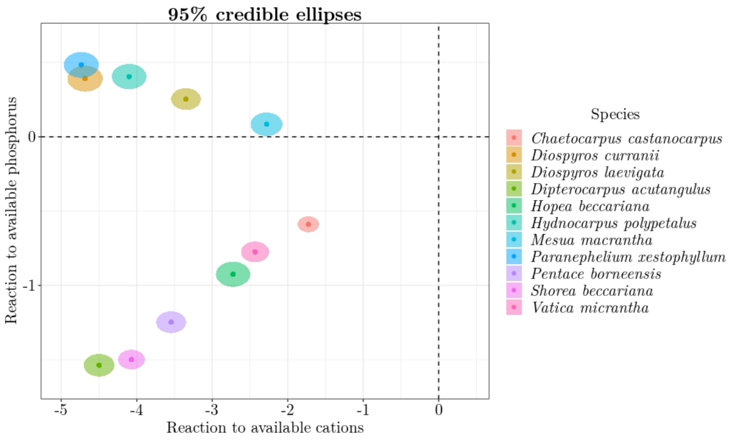
- The previous update consists in $n \times p$ optimization;
- One could image that similar $Y_{i,\dots}$ and $Xi, \dots$ should lead to similar $Z_{i,j}$, saying that posterior means and variance are *functions* of $Y_{i,\dots}$ and $Xi, \dots$;
- This leads to *amortization* (spirit of variational autoencoders);
- Actually, in our framework, this is the only brick that involves the observations distribution;
- This could lead to possible extensions for the distributions of Y (negative binomial, zero-inflated).

Application on data

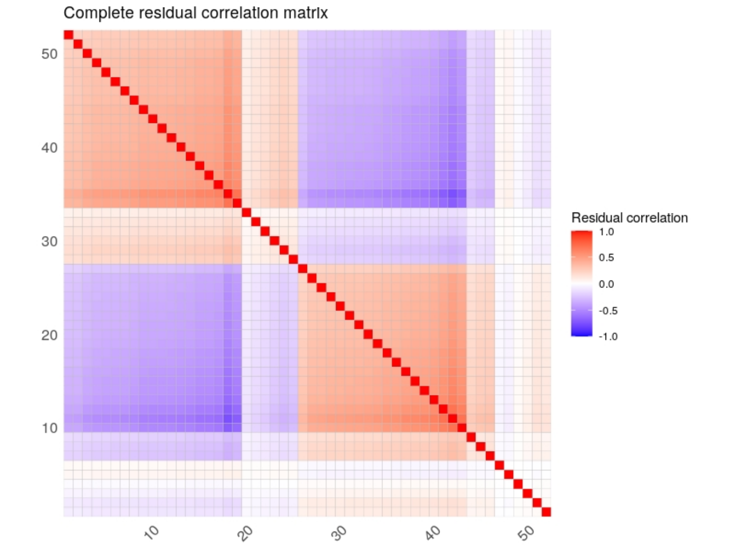
Framework

- 180 sites;
- Focus on 51 species being present relatively often;
- 18 quantitative covariates are highly correlated $\rightarrow$ transformed to 4 orthogonal and interpretable features using PCA.
- One qualitative covariate (soil typology) set aside at the beginning.

Example of estimated niches

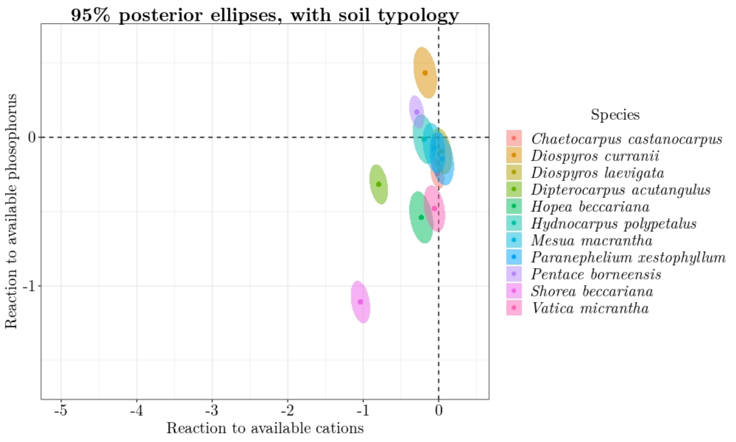


Estimated residual correlation

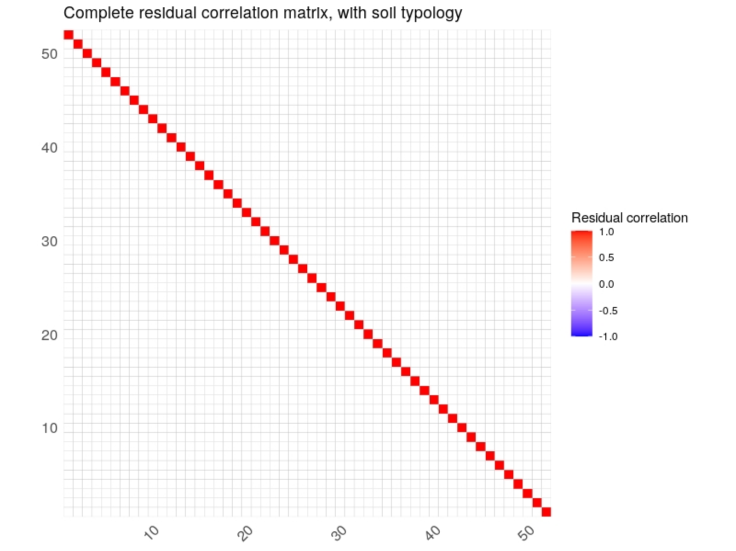


- The Λ matrix has 1 non zero column.

Addition of the soil typology as feature



Estimated residual correlation



- The Λ matrix has now 0 non zero column.

Conclusions

- “Efficient” and modular alternative to MCMC sampling for bayesian inference;
- Including different emission distribution should be straightforward;
- Alternative parameterization of the covariance would require alternative priors:
  - Quid of conjugacy?
  - Banerjee et Ghosal (2013)
- Efficiency actually depends on capacity to code, I should take lessons from PLN team;
- Implementing conjugate variational approach for parameters in PLN?
  - Would provide straightforward uncertainty quantification.

References

Banerjee, Sayantan, et Subhashis Ghosal. 2013. « Bayesian estimation of a sparse precision matrix ». *arXiv preprint arXiv:1309.1754*.

Bhattacharya, Anirban, et David B Dunson. 2011. « Sparse Bayesian infinite factor models ». *Biometrika*, 291-306.

Chiquet, Julien, Mahendra Mariadassou, et Stéphane Robin. 2021. « The Poisson-lognormal model as a versatile framework for the joint analysis of species abundances ». *Frontiers in Ecology and Evolution*. https://doi.org/10.3389/fevo.2021.588292.

Chiquet, Julien, Stephane Robin, et Mahendra Mariadassou. 2019. « Variational inference for sparse network reconstruction from count data ». In *International Conference on Machine Learning*, 1162-71. PMLR.

Ovaskainen, Otso, et Nerea Abrego. 2020. *Joint species distribution modelling: with applications in R*. Cambridge University Press.