Clustering species with residual covariance matrix in Joint Species Distribution models

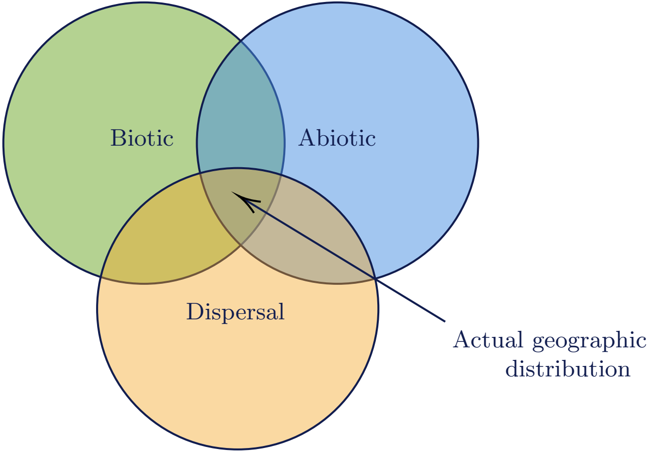
Daria Bystrova

Statify, Inria Grenoble Rhône-Alpes, Université Grenoble-Alpes, France

Applications du Bayesian Unified Group of Statisticians
25 June, 2021

Contents

Species distribution



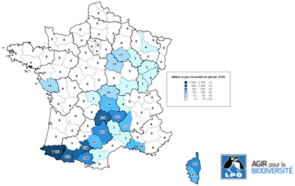
The three factors that determine the actual distribution of a species [Soberon and Peterson, 2005].

Species distribution models (SDMs)

Species

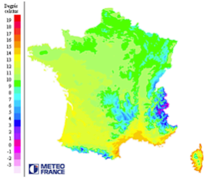
Species Distribution Data

$\{y\}_{i=1}^{n}$



Environmental covariates (characteristics)

$\{x\}_{i=1}^{n}$

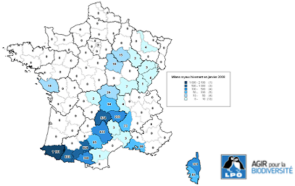


Species distribution models (SDMs)

Species

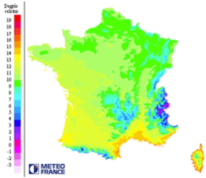
Species Distribution Data

$\{y\}_{i=1}^{n}$



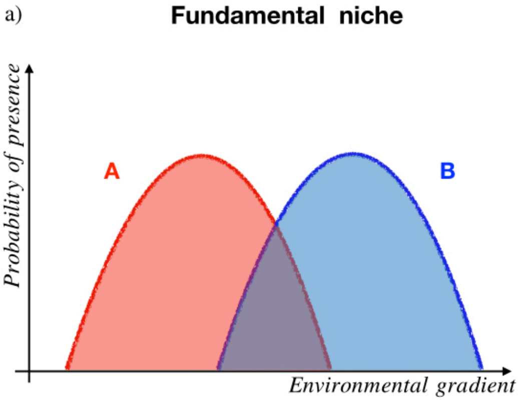
Environmental covariates (characteristics)

$\{x\}_{i=1}^{n}$



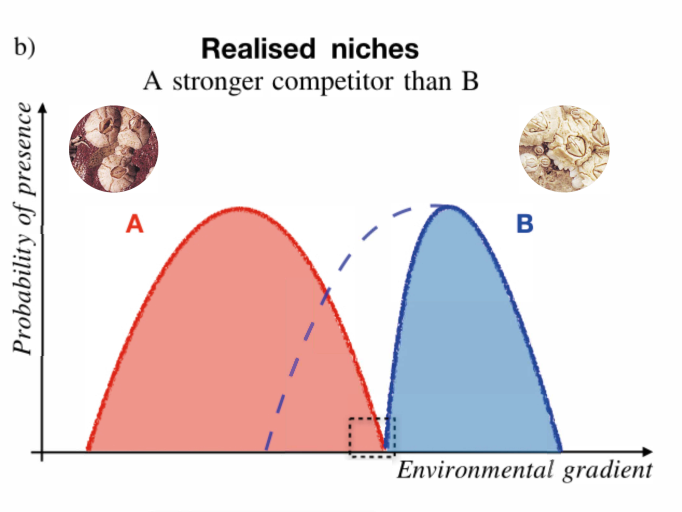
$$P(y_i = 1) = \Phi(\beta^T x_i)$$

Species distribution models: fundamental niche



[Poggiato et al., 2021]

Species distribution models: realized niche



[Poggiato et al., 2021]

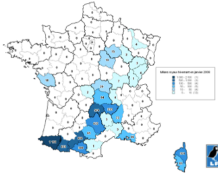
Species distribution models

Species

$j = \{1, \ldots, S\}$

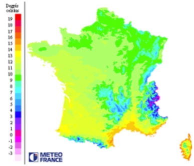
Species Distribution Data

$\{y_{ij}\}_{i=1}^{n}$



Environmental covariates (characteristics)

$\{x\}_{i=1}^{n}$



Joint species distribution models (JSDMs)

Formally: Consider $j = 1, .., S$ species, $i = 1, .., n$ sites and $\boldsymbol{x}_i = \{x_{ik}\}_{k=1}^{K}$ K environmental covariates. **Response variable** $y_{ij} \in \{0, 1\}^S$ is modelled as follows:

$$y_{ij} = \mathrm{I}(z_{ij} > 0)$$
$$\boldsymbol{z}_i = \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{e}_i$$
$$\boldsymbol{e}_i \overset{\text{iid}}{\sim} N_S(0, \boldsymbol{R}),$$

- $\boldsymbol{\beta} \in \mathbb{R}^{K \times S}$, $\boldsymbol{\beta_j}$ represent species-specific response to the environment.
- $\boldsymbol{R}$ correlation matrix: reflects species co-occurrence pattern not explained by selected abiotic covariates.

Existing approaches: Pollock et al. [2014],Clark et al. [2017] Ovaskainen et al. [2017],Harris [2015], Vanhatalo et al. [2020].

Problems:

- inter-species dependencies captured by $\boldsymbol{R} \neq$ species interactions.

Joint species distribution models (JSDMs)

Formally: Consider $j = 1, .., S$ species, $i = 1, .., n$ sites and $\boldsymbol{x}_i = \{x_{ik}\}_{k=1}^K$ K environmental covariates. **Response variable** $y_{ij} \in \{0, 1\}^S$ is modelled as follows:

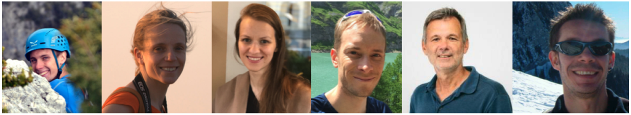
$$y_{ij} = \mathrm{I}(z_{ij} > 0)$$
$$\boldsymbol{z}_i = \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{e}_i$$
$$\boldsymbol{e}_i \overset{\text{iid}}{\sim} N_S(0, \boldsymbol{R}),$$

- $\boldsymbol{\beta} \in \mathbb{R}^{K \times S}$, $\boldsymbol{\beta_j}$ represent species-specific response to the environment.
- $\boldsymbol{R}$ correlation matrix: reflects species co-occurrence pattern not explained by selected abiotic covariates.

Existing approaches: Pollock et al. [2014],Clark et al. [2017] Ovaskainen et al. [2017],Harris [2015], Vanhatalo et al. [2020].

Problems:

- inter-species dependencies captured by $\boldsymbol{R} \neq$ species interactions.
- computationally heavy as models have $O(S^2)$ parameters



On the interpretations of Joint Species Distribution Models
Giovanni Poggiato, Tamara Münkemüllerer, Daria Bystrova, Julyan Arbel, James S. Clark and Wilfried Thuiller. *Trends Ecol. Evol.* 2021

Questions

- Can JSDMs improve the estimation of species fundamental niches?
- What can the residual correlation matrix tell us about biotic interactions?
- When and why do JDSMs outperform SDMs?

Clustering species with residual covariance matrix in Joint Species Distribution models.

Daria Bystrova, Giovanni Poggiato, Billur Bektas, Julyan Arbel, James S.Clark, Alessandra Guglielmi and Wilfried Thuiller. *Front. Ecol. Evol.* 2021

Questions

1. Can prior knowledge, combined with dimension reduction on the structure of the residual covariance matrix, improve model inference in JSDM?
2. Can estimated clusters be interpreted in ways that help us understand species communities?

Motivation

- models have $O(S^2)$ parameters $\implies$ dimension reduction approaches.
- large number of parameters in $\boldsymbol{R} \implies$ difficulties in interpretation
- use prior knowledge in JSDMs

Existing approaches:

- Latent variable models(LVM) [Warton et al., 2015], GJAM [Taylor-Rodriguez et al., 2017], HMSC [Ovaskainen et al., 2016], BORAL [Hui, 2016];
- efficient parallel sampling Chen et al. [2018], sjSDM [Pichler and Hartig, 2020]

Our proposal: a novel framework that allows for a clustering of residual associations that makes use of prior information.

Latent variable models

Formally : Consider $j = 1, .., S$ species, $i = 1, .., n$ sites and $\boldsymbol{x}_i = \{x_{ik}\}_{k=1}^{K}$ K environmental covariates, $y_{ij} \in \{0, 1\}^S$ is modelled as follows:

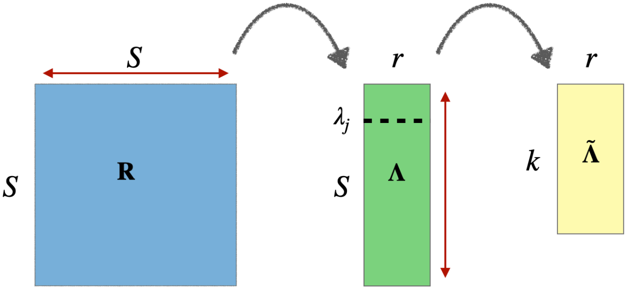
$$y_{ij} = \mathrm{I}(z_{ij} > 0),$$

$$\boldsymbol{z}_i = \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + e_i, \quad e_i \overset{\text{iid}}{\sim} N_S(0_S, \sigma_\epsilon^2 \boldsymbol{I}_S), \quad \omega_{ij} \overset{\text{iid}}{\sim} N(0, 1).$$

- $S \times r$ matrix $\boldsymbol{\Lambda}$ is the **factor loading** matrix
- r-dimensional Gaussian random vectors $\boldsymbol{\omega}_i$ are called **latent factors**
- $r \ll S$: approximating $\boldsymbol{\Sigma}$ with $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \sigma_\epsilon^2 \boldsymbol{I}_S$.

Further dimension reduction [Taylor-Rodriguez et al., 2017]:
Reducing Λ to $\tilde{\Lambda}$ by clustering rows of Λ
Clusters: species that share the same rows of Λ = species that **share the same residual correlation with respect to other species**.
→ cluster species depending on their associations with respect to other species.

Further dimension reduction



Clustering the rows $\lambda_j, j = 1, \ldots, S$ in Λ with Dirichlet process (**DP**):

$$\begin{aligned} \boldsymbol{\lambda}_j \mid G &\overset{\text{iid}}{\sim} G, \quad j = 1, \ldots, S, \\ G &\sim \text{DP}(\alpha H), \end{aligned}$$

Dirichlet process

Definition(Dirichlet process)[Ferguson, 1973]
Let H be a distribution over Θ and $\alpha > 0$. We say that G is a Dirichlet Process, namely $G \sim DP(\alpha H)$ if for any finite measurable partition $\{A_1, \ldots, A_r\}$ of Θ, we have:

$$(G(A_1), \ldots, G(A_r)) \sim Dir(\alpha H(A_1), \ldots, \alpha H(A_r)),$$

where H is called base distribution, and α the concentration parameter. The Dirichlet process (DP) is a central Bayesian nonparametric (BNP) prior.

The DP has almost surely discrete realisations[Sethuraman, 1994]:

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k},$$

where $\theta_k \sim H$

if $\mathrm{X}_1, \ldots, X_n \mid G \overset{\text{iid}}{\sim} G \quad i = 1, \ldots, n \Longrightarrow (X_1, \ldots, X_n)$ feature K_n distinct observations $(X_1^*, \ldots, X_{K_n}^*)$ with frequencies $n_1, \ldots, n_k$ such that $\sum_{i=1}^{K_n} n_i = n$.

Contribution

Clustering with BNP priors

- advantage of the BNP prior: no need to specify the exact number of clusters.
- BNP prior induce prior on the number of clusters K_n.
- In case of the DP process:
  - we can fix the features of the prior distribution on K_n using α
  - concentration parameter α has a strong effect on the posterior distribution of the number of clusters De Blasi et al. [2015].

Contribution

Clustering with BNP priors

- advantage of the BNP prior: no need to specify the exact number of clusters.
- BNP prior induce prior on the number of clusters K_n.
- In case of the DP process:
  - we can fix the features of the prior distribution on K_n using α
  - concentration parameter α has a strong effect on the posterior distribution of the number of clusters De Blasi et al. [2015].

Contribution:

1. We propose to incorporating prior knowledge on the number of species that share residual associations that improves clustering properties.
2. We introduce Pitman–Yor process, a more flexible Bayesian nonparametric prior, which is less sensitive to miscalibration than the Dirichlet process.

Contribution

First approach ($\mathbf{DP_c}$):
We consider prior distribution for the precision parameter α:

$$\alpha \sim \mathrm{Ga}(\nu_1, \nu_2)$$

$\rightarrow$ use Dirichlet multinomial process [Muliere and Secchi, 2003] for approximation of DP
$\rightarrow$ calibrate prior on K_n by calibrating prior on α: $\mathbb{E}[K_S] = K^*$, where K^* prior knowledge on the number of clusters

Contribution

First approach ($\mathbf{DP_c}$):
We consider prior distribution for the precision parameter α:

$$\alpha \sim \mathrm{Ga}(\nu_1, \nu_2)$$

$\rightarrow$ use Dirichlet multinomial process [Muliere and Secchi, 2003] for approximation of DP
$\rightarrow$ calibrate prior on K_n by calibrating prior on α: $\mathbb{E}[K_S] = K^*$, where K^* prior knowledge on the number of clusters

Second approach ($\mathbf{PY_c}$):
Use more flexible Pitman–Yor process (PY):

$$\begin{aligned} \boldsymbol{\lambda}_j \mid G &\overset{\text{iid}}{\sim} G, \quad j = 1, \dots, S, \\ G &\sim \mathrm{PY}(\alpha, \sigma, H), \end{aligned}$$

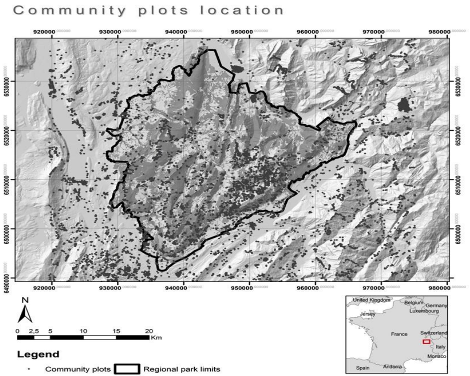
where H is the base measure, α is the concentration parameter, and σ is the discount parameter.
$\rightarrow$ use Pitman–Yor multinomial process [Lijoi et al., 2020] for approximation of PY
$\rightarrow$ calibrate prior on K_n, by calibrating parameters α, σ: $\mathbb{E}[K_S] = K^*$ and $\mathbb{V}[K_S]$ to reflect the desired level of uncertainty.

Specification of concentration parameter α

Model	**Concentration parameter α**	**Reference**
DP	Fixed (number of species)	Taylor-Rodriguez et al. [2017]
$\mathbf{DP}_c$	$Ga(\nu_1, \nu_2)$ s.t $\mathbb{E}[K_S] = K^*$	Ours
$\mathbf{PY}_c$	Fixed, s.t $\mathbb{E}[K_S] = K^*$	Ours + Lijoi et al. [2020]

K^* is the prior ecological belief on the number of groups of species with the same residual correlation structure.

Case study: The Bauges National Regional Park



The dataset contains the presences and absences of 111 plant species in 1139 sites. Thuiller et al. [2018]

- Prior knowledge on number of groups in the species interaction network: 16 Plant Functional Groups (PFGs). ($\mathbb{E}[K_S] = 16$)
- We splitted the sites into training and test set, and took covariates as in Thuiller et al. [2018].

Ecological representation of the clusters

Traits:
Landolt nutrient indicator, *Landolt lightindicator, height* (in the logarithmic scale), *specific leaf area* (SLA), *leaf dry matter content* (LDMC), *leafcarbon concentration* (LCC), and *leaf nitrogen concentration* (LNC).
Grouping traits with a similar role in the community assembly process:

- competitive effect (height, SLA,LDMC, LCC, LNC)
- tolerance to abiotic and biotic conditions (Landolt nutrient indicator, Landolt lightindicator)
- interaction via light resources (height, SLA, Landolt light indicator)
- interaction via soilresources (LNC, Landolt nutrient indicator).

$$\text{Species grouped-trait ratio} = \frac{\text{mean(distance to other species)}_{\text{within cluster}}}{\text{mean(distance to other species)}_{\text{all species}}}$$

Clustering analysis

posterior expected number of clusters:
describes the distribution of the number of clusters in Markov chain Monte Carlo (MCMC) samples.

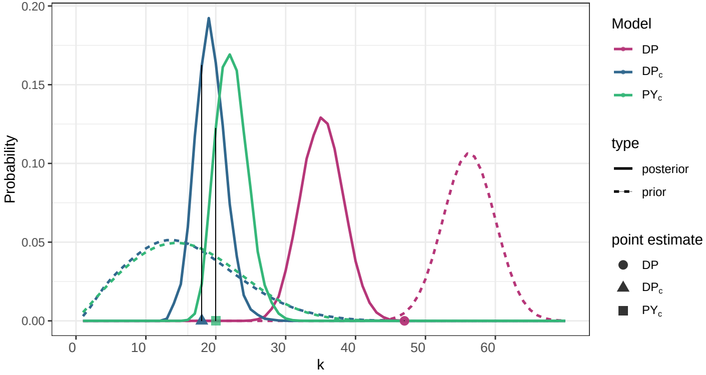
Summarize the **posterior distribution of the clusters** for **DP**, $\mathbf{DP}_c$ and $\mathbf{PY}_c$.
Optimal cluster estimate from Wade and Ghahramani [2018]:

$$c^* = \arg\min_{\hat{c}} \mathbb{E}[L(c, \hat{c}) \mid \boldsymbol{Y}_{1:n}] = \arg\min_{\hat{c}} \sum_{c} L(c, \hat{c}) p(c \mid \boldsymbol{Y}_{1:n}), \qquad (1)$$

where $p(c \mid \boldsymbol{Y}_{1:n})$ is posterior distribution of partition c.

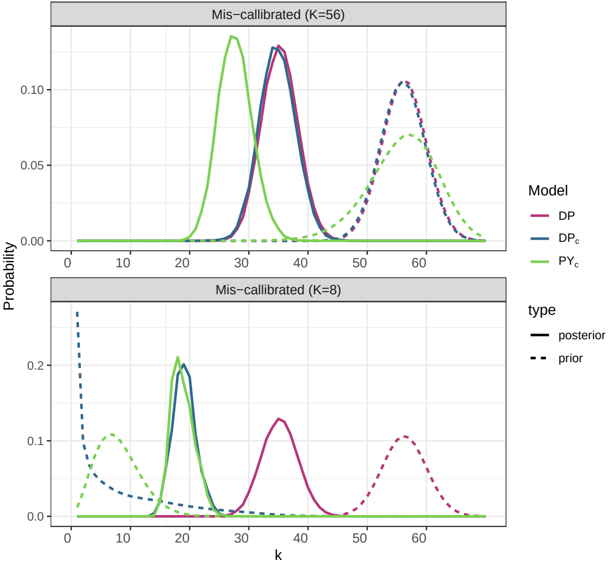
the number of clusters of the estimated clustering:
represents the number of clusters in the single partition that best represents the posterior distribution of the clusters in the MCMC samples

Results: posterior number of clusters



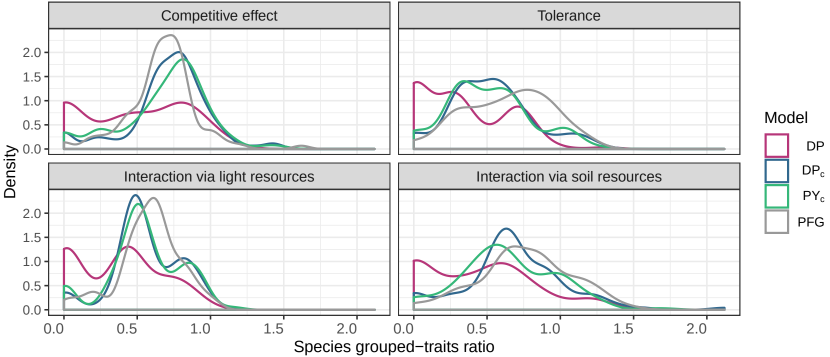
Prior distribution and posterior estimation of the number of clusters corresponding to **DP**, **DP**$_c$, **PY**$_c$ models.

Results: sensitivity



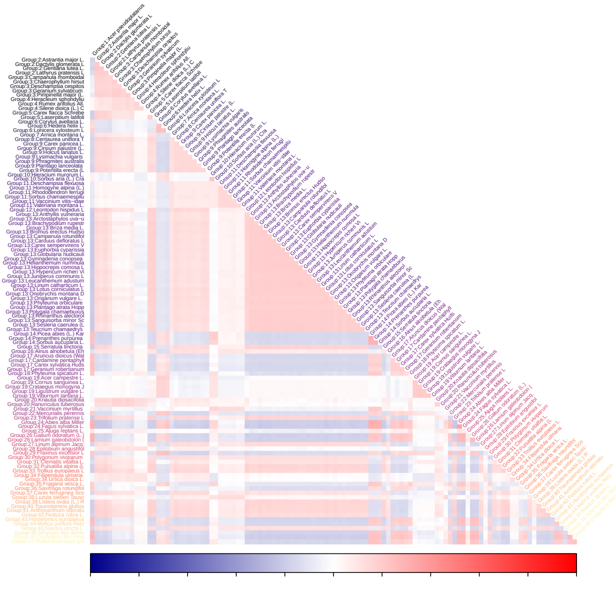
Prior and posterior distribution of the number of clusters for the models $\mathbf{DP}_c$, $\mathbf{PY}_c$.

Results: clusters interpretation

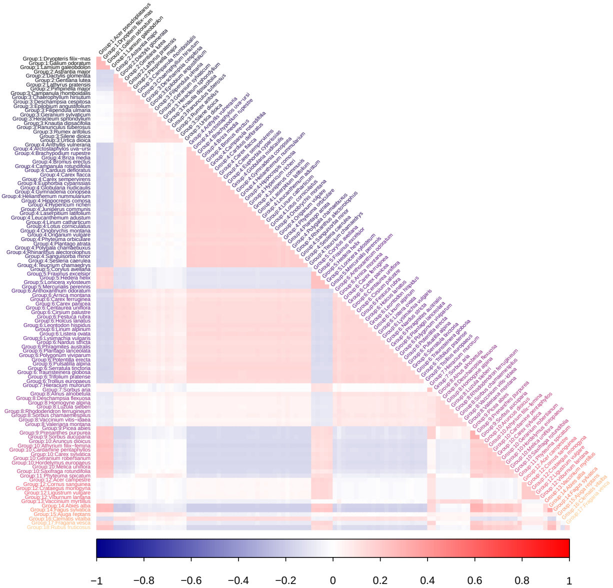


Distribution of species grouped-trait ratio. (**DP**,**DP**$_c$, **PY**$_c$)

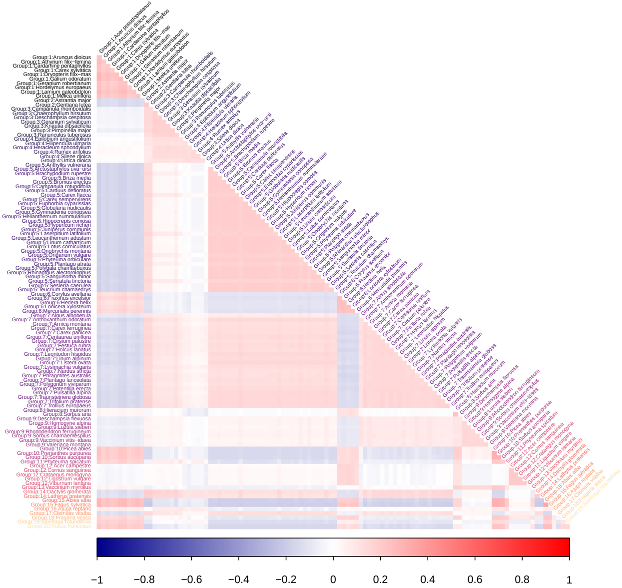
Results: residual covariance matrix DP



Results: residual covariance matrix DP_c



Results: residual covariance matrix $\mathbf{PY_c}$



Conclusion

(i) Our proposed statistical framework allows an additional but ecologically meaningful dimension reduction of JSDMs and includes prior knowledge in the residual covariance matrix.

(ii) The case study shows that the obtained clusters of species are ecologically meaningful,and correlated with functional traits.

Further directions

- we focus on estimation precision matrix with block-diagonal structure.

Links

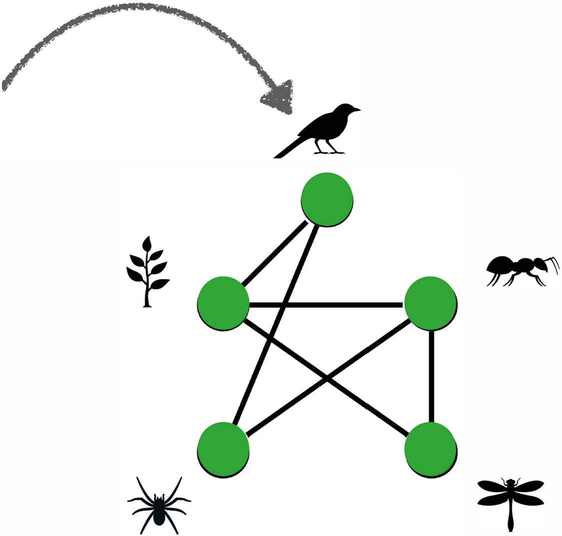
- Paper: https://www.frontiersin.org/articles/10.3389/fevo.2021.601384
- Code: https://github.com/dbystrova/GJAM_clust

Further directions

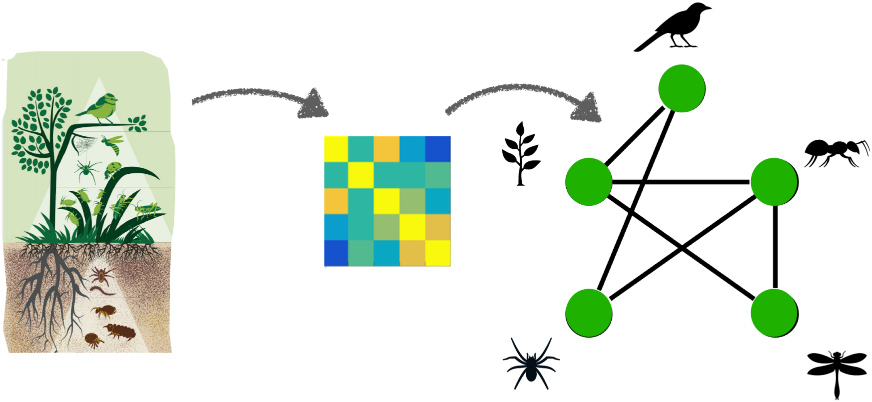
Bayesian block-diagonal graphical models via the Fiedler prior
Joint work with **Julyan Arbel** and **Mario Beraha**

Structure learning

Thakur [2020]



Structure learning



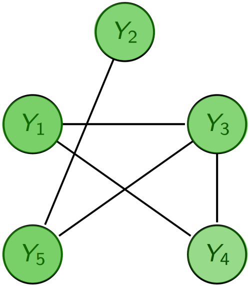
Gaussian graphical models

- associate p species with the components of a p-variate random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_p) \sim N_p(0, \boldsymbol{\Sigma})$
- n realizations of the vector $\boldsymbol{Y}$
- precision matrix $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ is associated to partial correlation matrix.

Precision matrix

A zero at element ij of **the precision matrix $\boldsymbol{\Theta}$** corresponds to conditional independence of Y_i and Y_j given the other variables Y_{-ij}

$$\boldsymbol{\Theta} = \begin{pmatrix} * & 0 & * & * & 0 \\ 0 & * & 0 & 0 & * \\ * & 0 & * & * & * \\ * & 0 & * & * & 0 \\ 0 & * & * & 0 & * \end{pmatrix} \Longrightarrow$$



Contribution

Inferential goals: estimation of the precision matrix $\boldsymbol{\Theta}$.

Problem

- Small sample setting: the number p of variables is greater than the number n of samples

Solutions:

- Graphical Lasso [Friedman et al., 2008], **Cluster Graphical Lasso**[Tan et al., 2015],
- Bayesian: Bayesian Graphical Lasso [Wang, 2012], Graphical Horseshoe [Li et al., 2019], G-Wishart prior [Mohammadi and Wit, 2015].

Our proposal:

- **Fiedler prior** shrinkage prior, useful for estimating sparse precision matrices with a **block-diagonal structure**.

Fiedler prior: motivation

Aim: take into account graph structure for prior specification.

Convenient way to capture connectivity of the graph: graph Laplacian .

Definition[Von Luxburg, 2007]: G undirected weighted graph, unnormalized Laplacian matrix $\boldsymbol{L}$ is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where

- weight matrix $\boldsymbol{W} = \{w_{ij}\}_{i,j=1}^{p}$, $w_{ij} \geq 0$
- degree matrix $\boldsymbol{D} = \text{diag}(\sum_j w_{1j}, \ldots, \sum_j w_{pj})$

Properties of Laplacian $\boldsymbol{L}$ with eigenvalues $\lambda_1 \leq \lambda_2 \leq, \ldots, \leq \lambda_p$

- $\lambda_1 = 0$
- λ_2 is called Fiedler value or algebraic connectivity: graph G is connected if and only if $\lambda_2 > 0$ (Fiedler regularization [Tam and Dunson, 2020])
- multiplicity k of the 0 eigenvalue of Laplacian equals to the number of connected components $A_1, \ldots, A_k$ [Von Luxburg, 2007].

Fiedler prior: definition

Consider partial correlation matrix $\boldsymbol{\Omega}$:
$\boldsymbol{\Omega} = \{\omega_{ij}\}_{i,j=1}^{p}$, $\omega_{ij} \in [-1, 1]$ and $\omega_{ij} = -\Sigma_{ij}^{-1} / \sqrt{\Sigma_{ii}^{-1} \Sigma_{jj}^{-1}}$.

Fiedler prior on $\boldsymbol{\Omega}$ with parameters $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)$, $\delta_j \geq 0$ is defined as

$$p(\boldsymbol{\Omega}|\boldsymbol{\delta}) = \frac{1}{Z(\boldsymbol{\delta})} \exp\left(- \sum_{j=1}^{p} \delta_j \lambda_j(\boldsymbol{\Omega}) \right)$$

where:

- $L(|\boldsymbol{\Omega}|)$ - Laplacian matrix $|\boldsymbol{\Omega}|$
- $\lambda_1(\boldsymbol{\Omega}) \leq \ldots \leq \lambda_p(\boldsymbol{\Omega})$ sorted eigenvalues of L.

To confront to precision matrix [Barnard et al., 2000]:

- $\boldsymbol{\Sigma}^{-1} = \boldsymbol{T\Omega T}$, where $\boldsymbol{T} = \text{diag}(\tau_1, \ldots, \tau_p)$.
- $\tau_j \overset{\text{iid}}{\sim} \text{Exp}(\eta), \quad j = 1, \ldots, p$.

Fiedler prior: properties

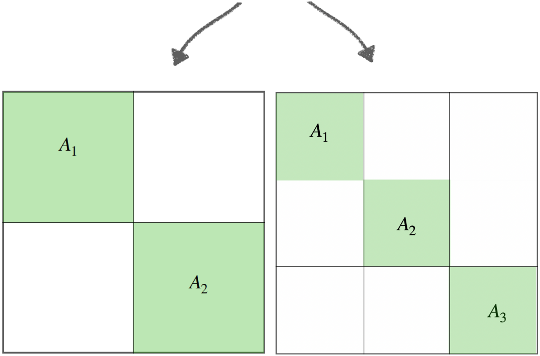
- By convention $\delta_1 = 0$
- if $\delta_2 = \ldots = \delta_p = \delta^*$:

$$\sum_{i=1}^{p} \delta_i \lambda_i(\mathbf{\Omega}) = \delta^* \sum_{i=1}^{p} \lambda_i(\mathbf{\Omega}) = \delta^* \mathrm{Tr}(L(|\mathbf{\Omega}|) = \delta^* \sum_{i \neq j} |\omega_{ij}|$$

$\implies$ Graphical Lasso for partial correlation matrices

Example:

$$\delta_2 > 0, \delta_3 = 0 \text{ or } \delta_2, \delta_3 > 0$$



References

Barnard, J., McCulloch, R., and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, pages 1281–1311.

Chen, D., Xue, Y., and Gomes, C. (2018). End-to-end learning for the deep multivariate probit model. volume 80 of *Proceedings of Machine Learning Research*, pages 932–941, Stockholmsmässan, Stockholm Sweden. PMLR.

Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., and Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs*, 87(1):34–56.

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(2):212–229.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Harris, D. J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution*, 6(4):465–473.

Hui, F. K. (2016). boral–Bayesian ordination and regression analysis of multivariate abundance data in R. *Methods in Ecology and Evolution*, 7(6):744–750.

Li, Y., Craig, B. A., and Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *JCGS*, 28(3):747–757.

Lijoi, A., Prünster, I., and Rigon, T. (2020). The Pitman–Yor multinomial process for mixture modeling. *Biometrika*, 107(4):891–906.

Mohammadi, A. and Wit, E. C. (2015). Bayesian structure learning in sparse gaussian graphical models. *Bayesian Analysis*, 10(1):109–138.

Muliere, P. and Secchi, P. (2003). Weak convergence of a Dirichlet-multinomial process. *Georgian Mathematical Journal*, 10(2):319–324.

Ovaskainen, O., Abrego, N., Halme, P., and Dunson, D. (2016). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, 7(5):549–555.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., and Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20(5):561–576.

Pichler, M. and Hartig, F. (2020). A new method for faster and more accurate inference of species associations from novel community data. *arXiv preprint arXiv:2003.05331*.

Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S., and Thuiller, W. (2021). On the interpretations of joint modeling in community ecology. *Trends in ecology & evolution*.

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., and McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model. *Methods in Ecology and Evolution*, 5(5):397–406.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650.

Soberon, J. and Peterson, A. (2005). Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity informatics*.

Tam, E. and Dunson, D. (2020). Fiedler regularization: Learning neural networks with graph sparsity. In *Proceedings of ICML 37*, volume 119, pages 9346–9355. PMLR.

Tan, K. M., Witten, D., and Shojaie, A. (2015). The cluster graphical lasso for improved estimation of gaussian graphical models. *CSDA*, 85:23–36.

Taylor-Rodriguez, D., Kaufeld, K., Schliep, E. M., Clark, J. S., and Gelfand, A. E. (2017). Joint species distribution modeling: dimension reduction using Dirichlet processes. *Bayesian Analysis*, 12(4):939–967.

Thuiller, W., Guéguen, M., Bison, M., Duparc, A., Garel, M., Loison, A., Renaud, J., and Poggiato, G. (2018). Combining point-process and landscape vegetation models to predict large herbivore distributions in space and time—A case study of Rupicapra rupicapra. *Diversity and Distributions*, 24(3):352–362.

Vanhatalo, J., Hartmann, M., and Veneranta, L. (2020). Additive multivariate gaussian processes for joint species distribution modeling with heterogeneous data. *Bayesian Anal.*, 15(2):415–447.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.

Wade, S. and Ghahramani, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis*, 13(2):559–626.

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*, 7(4):867–886.

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 & Evolution*, 30(12):766–779.