Generalized infinite factorization models

Antonio Canale

Joint work with Lorenzo Schiavon and David Dunson

University of Padova, Italy

canale.antonio@unipd.it

tonycanale_

Outline of presentation

Introduction

Infinite Factorization models

Generalized infinite FM and structured shrinkage priors

Simulations

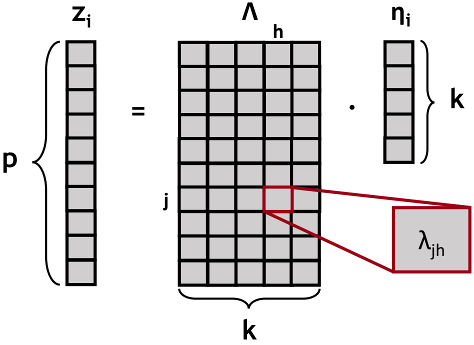
Application: Finnish bird co-occurrence data

Introduction - sparsity and interpretability

Factor Models (FM)

$$\color{red}{z_i = \Lambda \eta_i + \epsilon_i} \qquad \epsilon_i \sim f_\epsilon, \qquad \eta_i \sim f_\eta,$$

- z_i: i-th p-variate random variable;
- Λ: $p \times k$ factor loadings matrix;
- η_i: i-th vector of k latent factors.



Factor models - Thinking more generally

- Factor models express a statistical object of interest, in terms of a collection of simpler objects.
- FM as dimensionality reduction tool
- FM rooted back in psychometrics where the latent factors represent some interpretable latent trait (Spearman, 1904).
- Widely adopted and generalized: Gaussian copula FM (Murray et al., 2013), probabilistic matrix factorizations (Mnih & Salakhutdinov, 2008); functional data (Montagna et al., 2012); (Kowal and Canale, 2022)

Factor models - Thinking more generally

- Factor models express a statistical object of interest, in terms of a collection of simpler objects.
- FM as dimensionality reduction tool
- FM rooted back in psychometrics where the latent factors represent some interpretable latent trait (Spearman, 1904).
- Widely adopted and generalized: Gaussian copula FM (Murray et al., 2013), probabilistic matrix factorizations (Mnih & Salakhutdinov, 2008); functional data (Montagna et al., 2012); (Kowal and Canale, 2022)

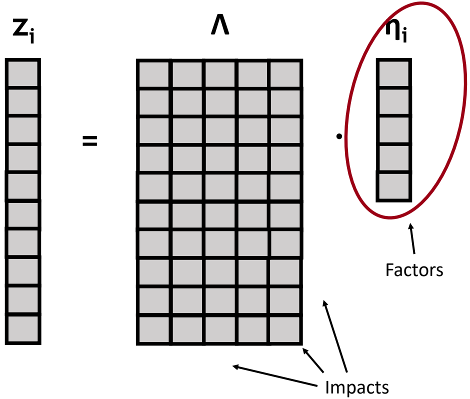
Factor models - Thinking more generally

- Factor models express a statistical object of interest, in terms of a collection of simpler objects.
- FM as dimensionality reduction tool
- FM rooted back in psychometrics where the latent factors represent some interpretable latent trait (Spearman, 1904).
- Widely adopted and generalized: Gaussian copula FM (Murray et al., 2013), probabilistic matrix factorizations (Mnih & Salakhutdinov, 2008); functional data (Montagna et al., 2012); (Kowal and Canale, 2022)

Factor models - Thinking more generally

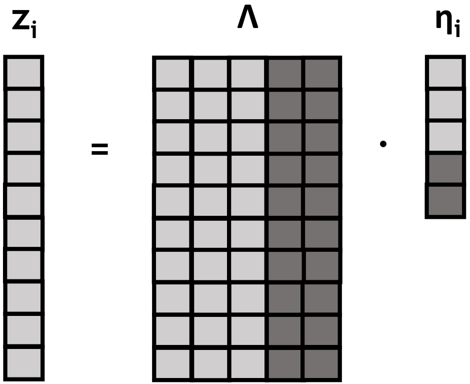
- Factor models express a statistical object of interest, in terms of a collection of simpler objects.
- FM as dimensionality reduction tool
- FM rooted back in psychometrics where the latent factors represent some interpretable latent trait (Spearman, 1904).
- Widely adopted and generalized: Gaussian copula FM (Murray et al., 2013), probabilistic matrix factorizations (Mnih & Salakhutdinov, 2008); functional data (Montagna et al., 2012); (Kowal and Canale, 2022)

Interpretation of factor models is assigning a meaning to the latent factors and then to their impact on the observed data.



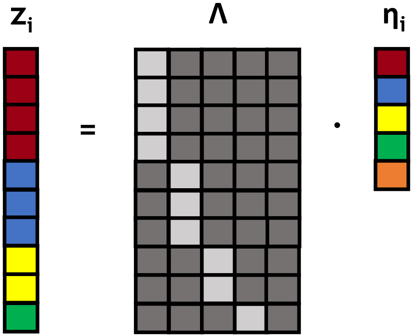
Interpretation of loadings matrix and factors is strongly favoured by

- a limited number *k* of factors;



Interpretation of loadings matrix and factors is strongly favoured when

- each factor has an impact only on a small group of components of z_i.



Definition

A **sparse** array is an array in which most of the elements are equal to zero.

- **Zeros in the last columns** of $\Lambda \Rightarrow$ reducing the number of factors k
- **Pattern of zeros** in $\Lambda \Rightarrow$ factors influence only small groups of components

Definition

A **sparse** array is an array in which most of the elements are equal to zero.

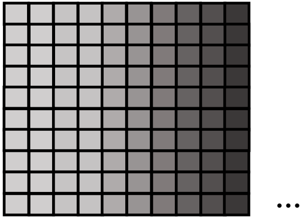
- **Zeros in the last columns** of $\Lambda \Rightarrow$ reducing the number of factors k
- **Pattern of zeros** in $\Lambda \Rightarrow$ factors influence only small groups of components

Infinite Factorization models

Bayesian approach introduced by Bhattacharya and Dunson (2011).

Infinitely many factors, with the **impact** of these factors **decreasing** with the factor index.

Accomplished with **increasing shrinkage priors**, that allow to approximate the IFM through a finite number of factors.



Λ

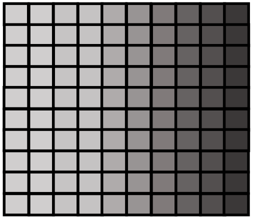
No structure **constraints** are imposed on the number of factors or on the sparsity pattern!

Infinite factor models (IFM)

Bayesian approach introduced by Bhattacharya and Dunson (2011).

Infinitely many factors, with the **impact** of these factors **decreasing** with the factor index.

Accomplished with **increasing shrinkage priors**, that allow to approximate the IFM through a finite number of factors.



...

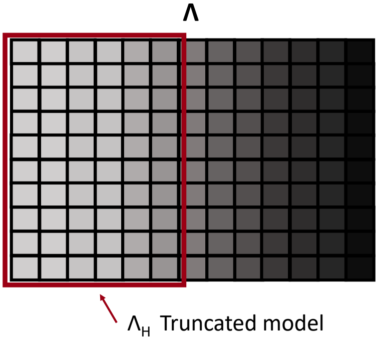
Λ

No structure **constraints** are imposed on the number of factors or on the sparsity pattern!

Posterior inference is conducted via **Monte Carlo Markov Chains**.

Truncating out the negligible **columns** of Λ, those really **close to zero** $\Rightarrow$ small number of latent factors.

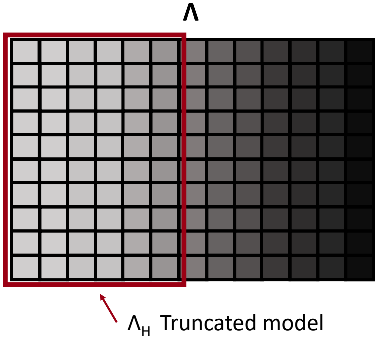
Priors on loadings elements with sufficiently mass concentration **around zero** $\Rightarrow$ Sparse pattern on Λ.



Posterior inference is conducted via **Monte Carlo Markov Chains**.

Truncating out the negligible **columns** of Λ, those really **close to zero** $\Rightarrow$ small number of latent factors.

Priors on loadings elements with sufficiently mass concentration **around zero** $\Rightarrow$ Sparse pattern on Λ.



Current Infinite Factor Models

- Multiplicative gamma process (**MGP**) - Bhattacharya & Dunson, 2011.
- Cumulative shrinkage process (**CUSP**) - Legramanti et al., 2020.

Problems:

1. lack of careful consideration of the within component sparsity structure
2. no accommodation for grouped variables and other non-exchangeable structure.

- Multiplicative gamma process (**MGP**) - Bhattacharya & Dunson, 2011.
- Cumulative shrinkage process (**CUSP**) - Legramanti et al., 2020.

Problems:

1. lack of careful consideration of the within component sparsity structure
2. no accommodation for grouped variables and other non-exchangeable structure.

Generalized infinite FM and structured shrinkage priors

Generalized Infinite Factorization (GIF)

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh})$$

$$\theta_{jh} = \tau_0$$

- $\tau_0 \sim f_{\tau_0}$: global scale;

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh})$$

$$\theta_{jh} = \tau_0$$

- $\tau_0 \sim f_{\tau_0}$: global scale;

Generalized Infinite Factorization (GIF)

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh})$$

$$\theta_{jh} = \tau_0 \, \gamma_h$$

- $\tau_0 \sim f_{\tau_0}$: global scale;
- $\gamma_h \sim f_{\gamma_h}$: column scale;

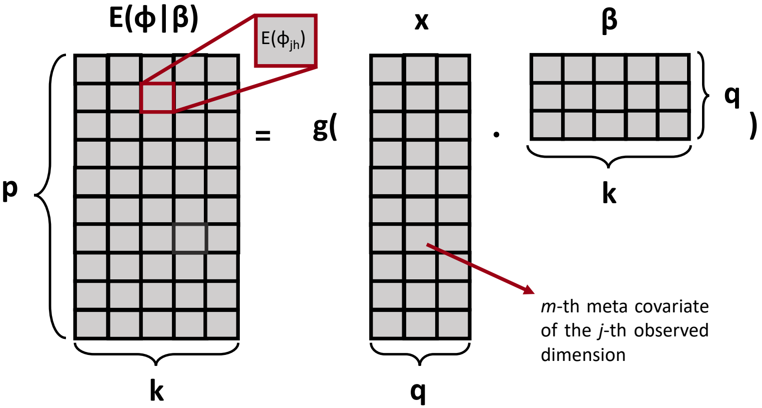
Generalized Infinite Factorization (GIF)

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh})$$

$$\theta_{jh} = \tau_0 \, \gamma_h \, \phi_{jh}$$

- $\tau_0 \sim f_{\tau_0}$: global scale;
- $\gamma_h \sim f_{\gamma_h}$: column scale;
- $\phi_{jh} \sim f_{\phi_j}$: local scale. That depends on meta covariates: $E(\phi_{jh}) = g(x_j^\top \beta_h)$

Exogenous information about the sparsity structure



$$E(\phi_{jh} \mid \beta_h) = g(x_j^\top \beta_h), \qquad \beta_h = (\beta_{1h}, \ldots, \beta_{qh})^\top, \qquad \beta_{mh} \sim f_\beta$$

Bird species occurrence example (1)

- **y**: occurrence of **p species** in **n** different **environments**;
- η: **k** latent **factors**;
- Λ: **impact of the latent factors** on the species occurrence;
- **x**: q **species characteristics** (taxonomy, size, migratory strategy...), providing similarities between different species.

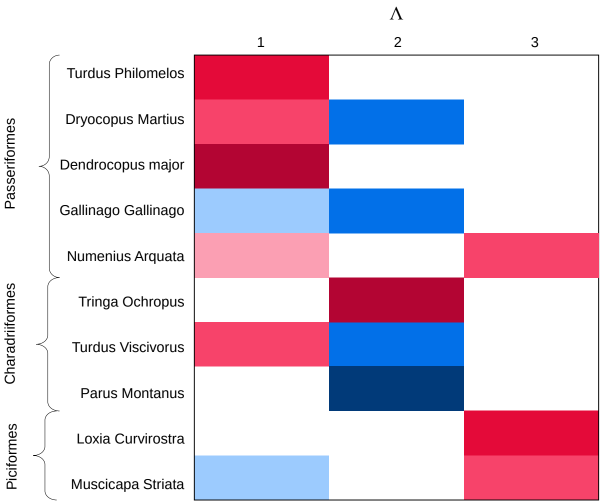
Considering x indicating the phylogenetic order of each species.
If the h-th factor does not impact the occurrence of the species j ($\lambda_{jh} = 0$), it could not even impact the other species s belonging to the same order of j ($\lambda_{sh} = 0$).

Bird species occurrence example (1)

- **y**: occurrence of **p species** in **n** different **environments**;
- η: **k** latent **factors**;
- Λ: **impact of the latent factors** on the species occurrence;
- **x**: q **species characteristics** (taxonomy, size, migratory strategy...), providing similarities between different species.

Considering x indicating the phylogenetic order of each species.
If the h-th factor does not impact the occurrence of the species j ($\lambda_{jh} = 0$), it could not even impact the other species s belonging to the same order of j ($\lambda_{sh} = 0$).

Bird species occurrence example (2)



Theoretical prior properties

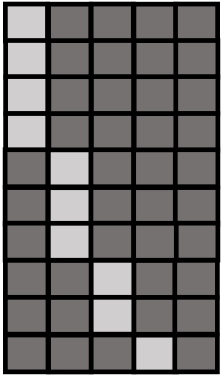
We define desirable properties for the GIF class including

- Increasing shrinkage ($\text{var}(\lambda_{jh}) < \text{var}(\lambda_{j(h-1)})$ for any h)
- Robustness to large signals (not overshrinking)
- Asymptotic increasing sparsity (for $p \to \infty$ the sparsity rate increases)

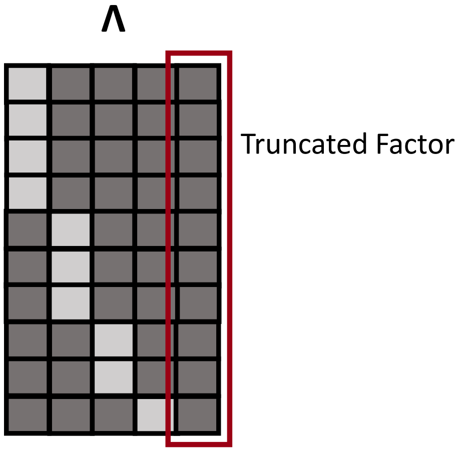
We provide conditions for the properties to hold.

Practical consequences: sparsity and interpretability

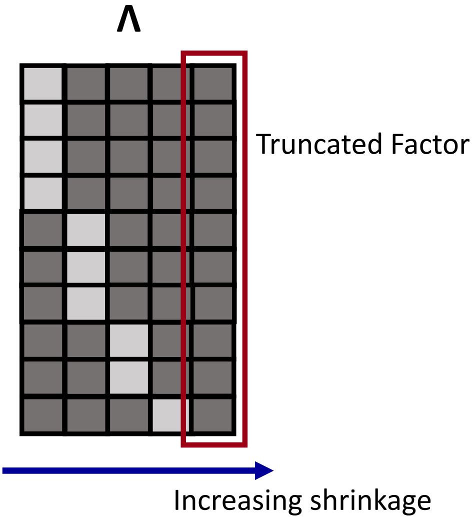
Λ



Coming back to the origin: sparsity and interpretation

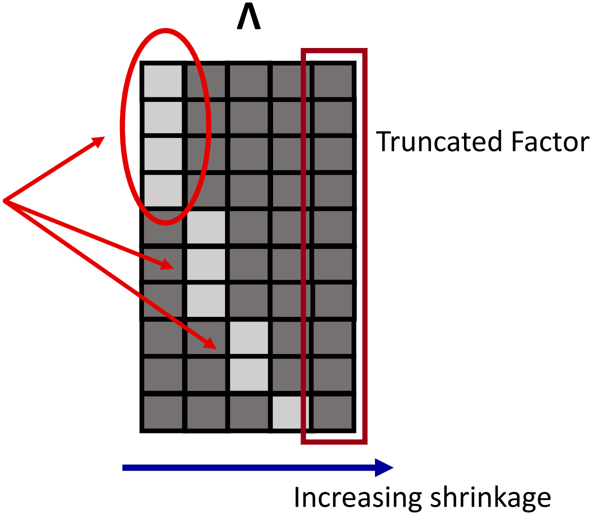


Practical consequences: sparsity and interpretability

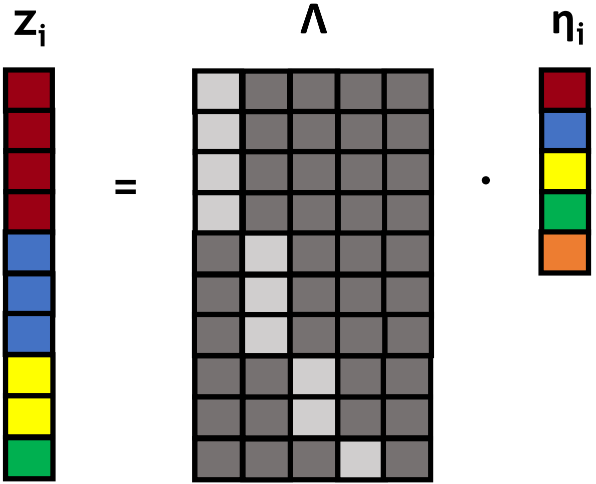


Practical consequences: sparsity and interpretability

Groups of similar components according to ***X***



Practical consequences: sparsity and interpretability



$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh}) \qquad \theta_{jh} = \tau_0 \, \gamma_h \, \phi_{jh}$

Central GIF equations

$\tau_0 = 1, \qquad \gamma_h = \vartheta_h \rho_h, \qquad \vartheta_h^{-1} \sim \mathrm{Ga}(a_\theta, b_\theta),$

Power law tail column scale

$\rho_h = \mathrm{Ber}\left(1 - \pi_h\right), \qquad \pi_h = \sum_{l=1}^{h} w_l, \qquad w_l = v_l \prod_{m=1}^{l-1}(1 - v_m), \qquad v_m \sim \mathrm{Be}(1, \alpha),$

Increasing shrinkage via cumulative stick-breaking process (Legramanti et al. 2020)

$\phi_{jh} \mid \beta_h \sim \mathrm{Ber}\{\mathrm{logit}(X_j^\top \beta_h)\} \log(p)/p \qquad \beta_h \sim N_q(0, \sigma_\beta^2 I_q),$

Meta covariates inclusion that impacts the sparsity pattern

Structured Increasing Shrinkage prior

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh}) \qquad \theta_{jh} = \tau_0 \, \gamma_h \, \phi_{jh}$$

Central GIF equations

$$\tau_0 = 1, \qquad \gamma_h = \vartheta_h \rho_h, \qquad \vartheta_h^{-1} \sim \mathrm{Ga}(a_\theta, b_\theta),$$

Power law tail column scale

$$\rho_h = \mathrm{Ber}\left(1 - \pi_h\right), \qquad \pi_h = \sum_{l=1}^{h} w_l, \qquad w_l = v_l \prod_{m=1}^{l-1}(1 - v_m), \qquad v_m \sim \mathrm{Be}(1, \alpha),$$

Increasing shrinkage via cumulative stick-breaking process (Legramanti et al. 2020)

$$\phi_{jh} \mid \beta_h \sim \mathrm{Ber}\{\mathrm{logit}(X_j^\top \beta_h)\} \log(p)/p \qquad \beta_h \sim N_q(0, \sigma_\beta^2 I_q),$$

Meta covariates inclusion that impacts the sparsity pattern

Structured Increasing Shrinkage prior

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh}) \qquad \theta_{jh} = \tau_0 \, \gamma_h \, \phi_{jh}$$

Central GIF equations

$$\tau_0 = 1, \qquad \gamma_h = \vartheta_h \rho_h, \qquad \vartheta_h^{-1} \sim \mathrm{Ga}(a_\theta, b_\theta),$$

Power law tail column scale

$$\rho_h = \mathrm{Ber}\left(1 - \pi_h\right), \qquad \pi_h = \textstyle\sum_{l=1}^{h} w_l, \qquad w_l = v_l \prod_{m=1}^{l-1}(1 - v_m), \qquad v_m \sim \mathrm{Be}(1, \alpha),$$

Increasing shrinkage via cumulative stick-breaking process (Legramanti et al. 2020)

$$\phi_{jh} \mid \beta_h \sim \mathrm{Ber}\{\mathrm{logit}(X_j^\top \beta_h)\} \log(p)/p \qquad \beta_h \sim N_q(0, \sigma_\beta^2 I_q),$$

Meta covariates inclusion that impacts the sparsity pattern

Structured Increasing Shrinkage prior

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh}) \qquad \theta_{jh} = \tau_0 \, \gamma_h \, \phi_{jh}$$

Central GIF equations

$$\tau_0 = 1, \qquad \gamma_h = \vartheta_h \rho_h, \qquad \vartheta_h^{-1} \sim \mathrm{Ga}(a_\theta, b_\theta),$$

Power law tail column scale

$$\rho_h = \mathrm{Ber}\left(1 - \pi_h\right), \qquad \pi_h = \sum_{l=1}^{h} w_l, \qquad w_l = v_l \prod_{m=1}^{l-1}(1 - v_m), \qquad v_m \sim \mathrm{Be}(1, \alpha),$$

Increasing shrinkage via cumulative stick-breaking process (Legramanti et al. 2020)

$$\phi_{jh} \mid \beta_h \sim \mathrm{Ber}\{\mathrm{logit}(X_j^\top \beta_h)\} \log(p)/p \qquad \beta_h \sim N_q(0, \sigma_\beta^2 I_q),$$

Meta covariates inclusion that impacts the sparsity pattern

Simulations

Simulation scenarios

- We compare the performance of our proposal with current approaches (Bhattacharya & Dunson, 2011; Rockova & George, 2016; Legramanti et al., 2020)
- Scenario (a) increasing shrinkage FM (no local sparsity; Scenario (b) locally sparse FM (no increasing shrinkage); Scenario (c) is a) + b); Scenario (d) is b) + c) + metacovariate-dependence in sparsity
- Performance measures: LPML, posterior mean of k (estimated number of columns of Λ), MSE of Ω

	(p,k)	MGP		CUSP		SIS	
		$Q_{0.5}$	IQR	$Q_{0.5}$	IQR	$Q_{0.5}$	IQR
LPML	(16,4)	-28.68	0.42	-28.68	0.43	-28.65	0.41
	(32,8)	-60.08	0.45	-60.09	0.45	-60.07	0.49
	(64,12)	-117.68	0.56	-117.75	0.53	-117.88	0.56
	(128,16)	-225.04	1.04	-225.13	1.04	-228.76	1.47
$E(H_a \mid y)$	(16,4)	8.17	1.44	4.00	0.00	4.00	0.00
	(32,8)	10.68	0.33	8.00	0.00	8.00	0.00
	(64,12)	14.16	1.09	12.00	0.00	12.00	0.00
	(128,16)	17.03	0.47	16.00	0.00	18.00	0.02

Figure 1: LPM;L and estimated latent dimension (k) in Scenario (a) —worst case for the proposed method)

Results (2)

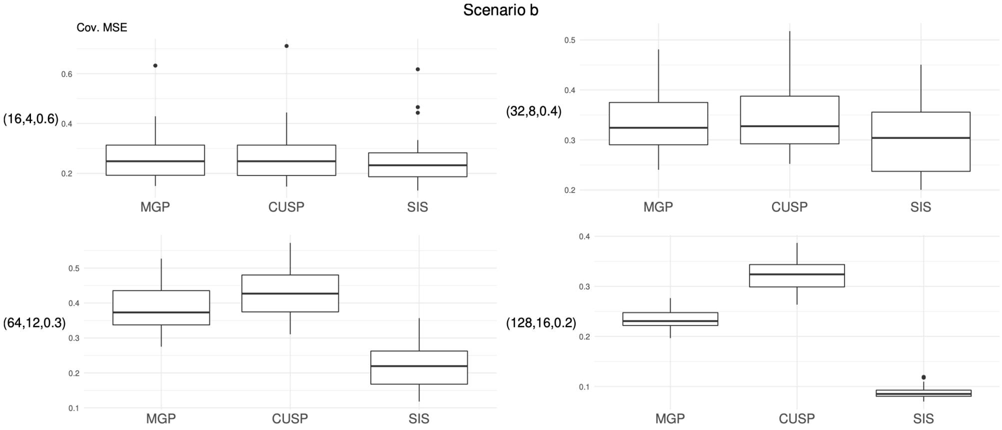


Figure 2: MSE for Ω for different combination of (p, k, s)

Finnish bird co-occurrence data

The co-occurrence model

- $\boldsymbol{y}$: $n \times p$ binary matrix of occurrence of **p species** in **n** different **environments**.
- $\boldsymbol{w}$: $n \times c$ **covariate matrix** including habitat type and the 'spring temperature'.
- $\boldsymbol{x}$: $p \times q$ **meta covariate matrix** including **species traits**: the species log body mass, the species migratory strategy and species superfamily.

$$y_{ij} = \mathbb{1}(z_{ij} > 0), \quad z_{ij} = w_i^T \mu_j + \epsilon_{ij}, \quad \epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^T \sim N_p(0, \Lambda\Lambda^T + I_p),$$

- $\boldsymbol{z}$: $n \times p$ underlying continuous matrix.
- Λ: loadings matrix with structured increasing shrinkage prior such that the species traits x can impact the covariance structure across species.

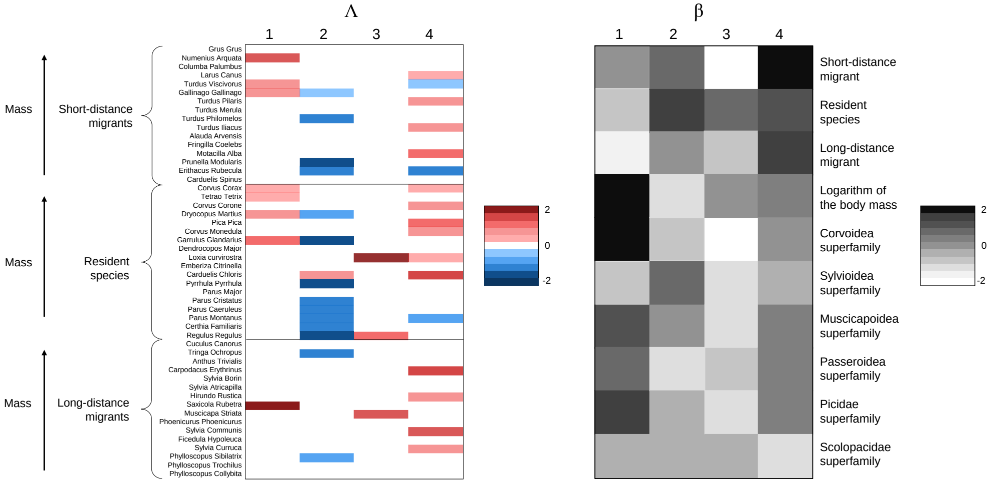
The co-occurrence model

- ***y***: $n \times p$ binary matrix of occurrence of **p species** in **n** different **environments**.
- ***w***: $n \times c$ **covariate matrix** including habitat type and the 'spring temperature'.
- ***x***: $p \times q$ **meta covariate matrix** including **species traits**: the species log body mass, the species migratory strategy and species superfamily.

$$y_{ij} = \mathbb{1}(z_{ij} > 0), \quad z_{ij} = w_i^T \mu_j + \epsilon_{ij}, \quad \epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^T \sim N_p(0, \Lambda\Lambda^T + I_p),$$

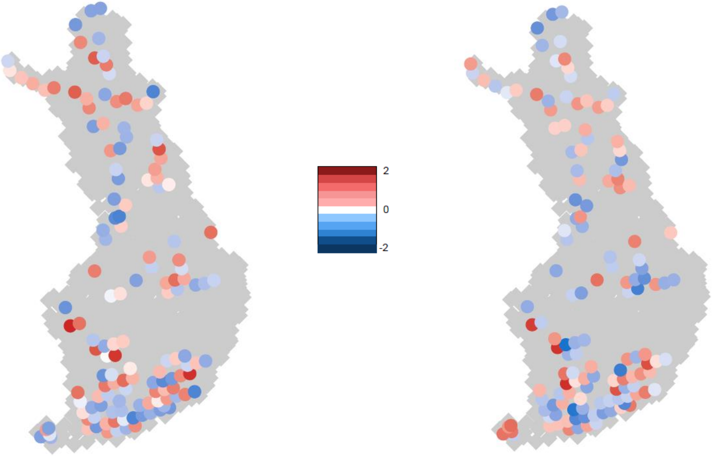
- ***z***: $n \times p$ underlying continuous matrix.
- Λ: loadings matrix with structured increasing shrinkage prior such that the species traits x can impact the covariance structure across species.

Loadings


Λ
1
2
3
4
Mass
Short-distance migrants
Mass
Resident species
Mass
Long-distance migrants
Grus Grus
Numenius Arquata
Columba Palumbus
Larus Canus
Turdus Viscivorus
Gallinago Gallinago
Turdus Pilaris
Turdus Merula
Turdus Philomelos
Turdus Iliacus
Alauda Arvensis
Fringilla Coelebs
Motacilla Alba
Prunella Modularis
Erithacus Rubecula
Carduelis Spinus
Corvus Corax
Tetrao Tetrix
Corvus Corone
Dryocopus Martius
Pica Pica
Corvus Monedula
Garrulus Glandarius
Dendrocopos Major
Loxia curvirostra
Emberiza Citrinella
Carduelis Chloris
Pyrrhula Pyrrhula
Parus Major
Parus Cristatus
Parus Caeruleus
Parus Montanus
Certhia Familiaris
Regulus Regulus
Cuculus Canorus
Tringa Ochropus
Anthus Trivialis
Carpodacus Erythrinus
Sylvia Borin
Sylvia Atricapilla
Hirundo Rustica
Saxicola Rubetra
Muscicapa Striata
Phoenicurus Phoenicurus
Sylvia Communis
Ficedula Hypoleuca
Sylvia Curruca
Phylloscopus Sibilatrix
Phylloscopus Trochilus
Phylloscopus Collybita
2
0
-2
β
1
2
3
4
Short-distance migrant
Resident species
Long-distance migrant
Logarithm of the body mass
Corvoidea superfamily
Sylvioidea superfamily
Muscicapoidea superfamily
Passeroidea superfamily
Picidae superfamily
Scolopacidae superfamily
2
0
-2

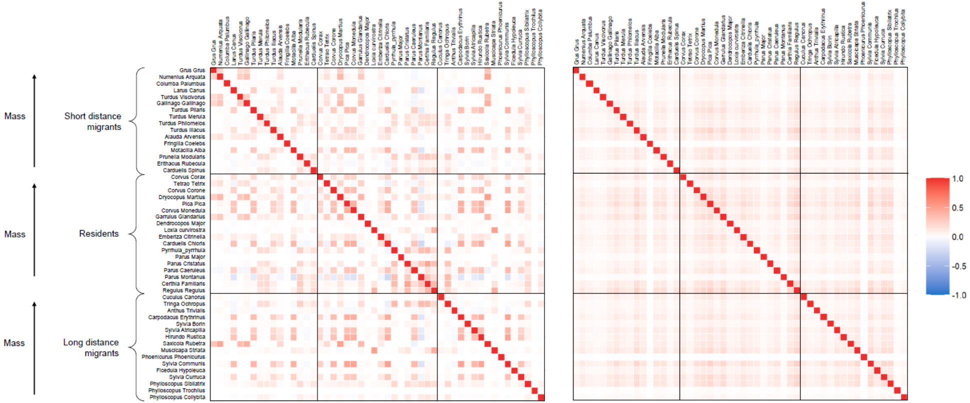
Finnish environments latent covariates

First factor **Third factor**



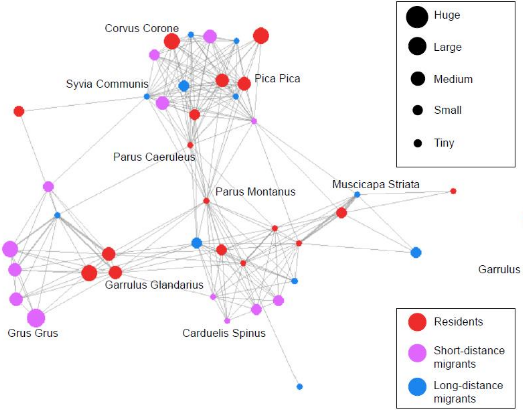
Posterior mean of correlation matrices

Structured Increasing Shrinkage prior | **Multiplicative Gamma Process**

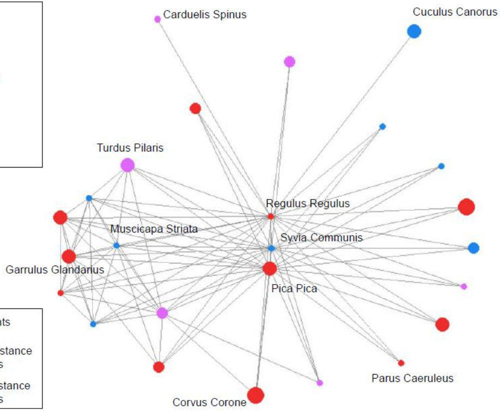


Graph representation of covariance structure

Structured Increasing Shrinkage prior



Multiplicative Gamma Process



- We introduced a class of generalized infinite factorization models
- We equipped the model with a structured increasing shrinkage prior enjoying appealing theoretical properties
- Practical gains are
- Computation is straightforward with (adaptive) Gibbs sampling
- Possible extensions in terms of probabilistic matrix factorization models

References

A. Bhattacharya and D. B. Dunson (2011). Sparse Bayesian infinite factor models. *Biometrika*
IEEE Transactions on neural networks and learning systems.

S. Legramanti, D. Durante, and D. B. Dunson (2020). Bayesian cumulative shrinkage for infinite factorizations. *Biometrika*.

A. Mnih and R. R. Salakhutdinov (2008). Probabilistic matrix factorization. In *Advances in neural information processing systems*.

S. Montagna, S. T. Tokdar, B. Neelon, and D. B. Dunson (2012). Bayesian latent factor regression for functional and longitudinal data. *Biometrics*.

J. S. Murray, D. B. Dunson, L. Carin, and J. E. Lucas (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*.

B. J. Reich and D. Bandyopadhyay (2010). A latent factor model for spatial data with informative missingness. *The annals of applied statistics*.

L. Schiavon, A. Canale, and D. B. Dunson (in press). Generalized infinite factorization models. *Biometrika*.

Thank you for your attention!

University of Padova . . .

Università degli Studi di Padova

. . . since 1222