Spatial Blind Source Separation

Klaus Nordhausen

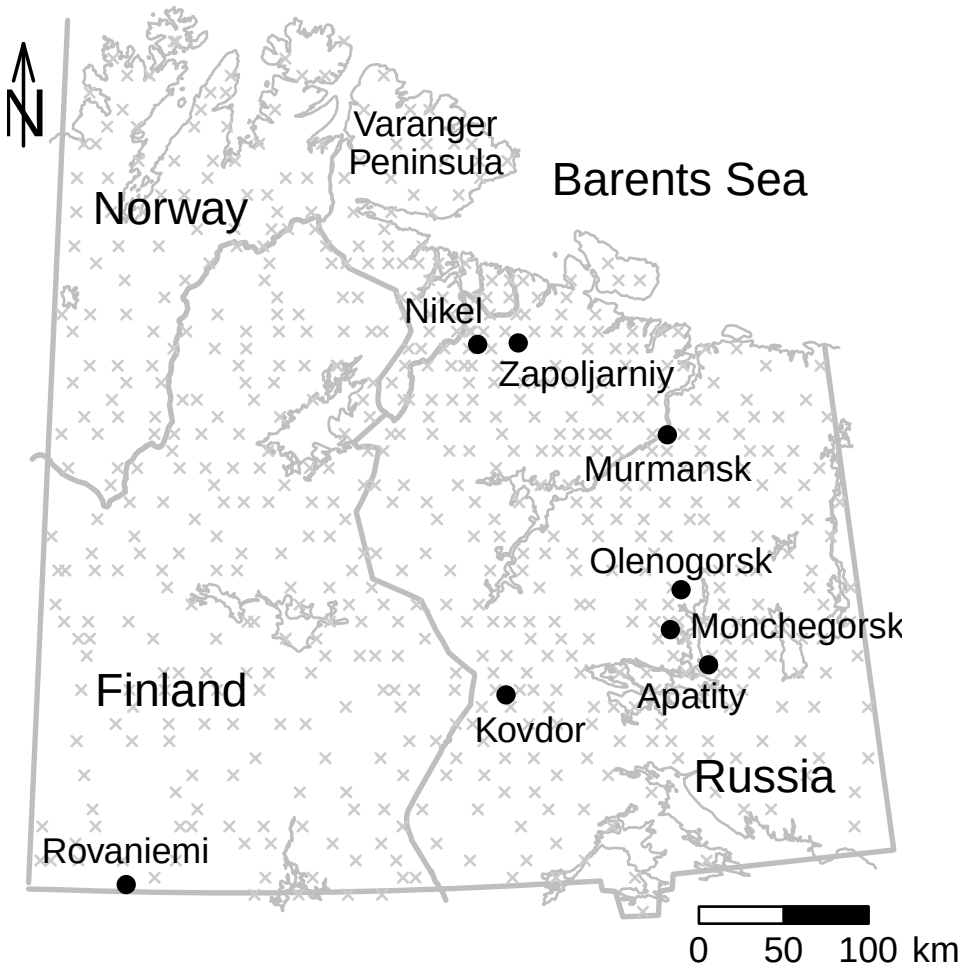
CSTAT - Computational Statistics
Institute of Statistics & Mathematical Methods in Economics
Vienna University of Technology

based on joint work with many:

Hannu Oja, Joni Virta, Francois Bachoc, Marc Genton, Anne Ruiz-Gazen, Peter Filzmoser, Clemens Reimann, Christoph Mühlmann, Mengxi Yi

Outline

Kola Data

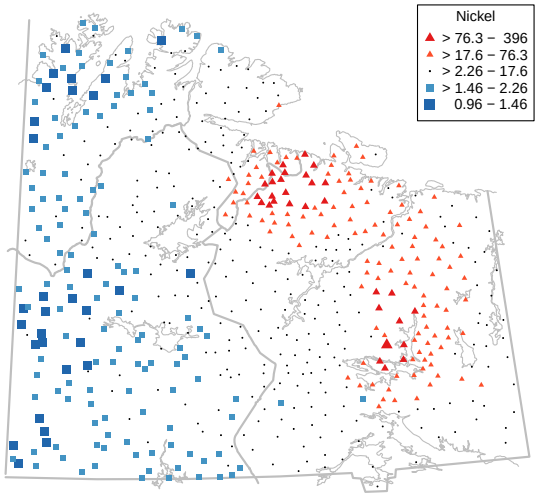


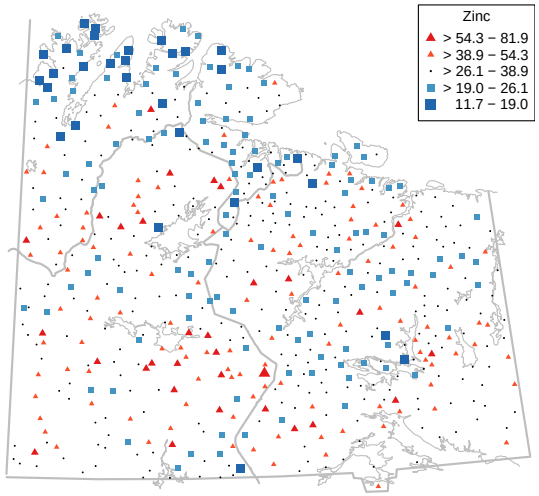
- samples of terrestrial moss
- sampled at 594 sites
- 31 chemical elements under consideration

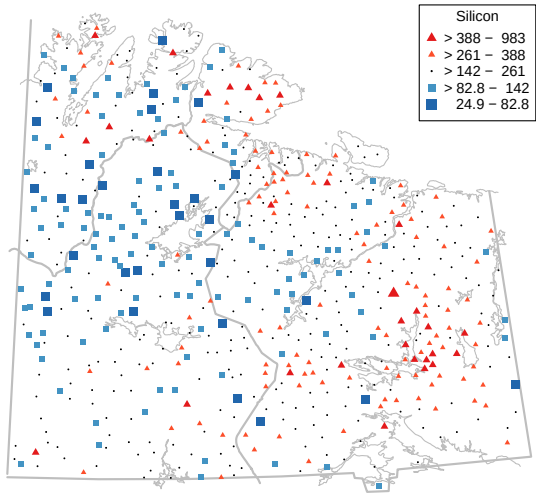
Typical Features of Geochemical Data

1. The data is multivariate.
2. The data is compositional, which means that correlations tend to be negative.
3. The measurements are spatially dependent, which means correlations between variables taken at sites close to each other tend to be high.

Maps for selected chemical elements







Spatial statistics terminology

Consider a random field $z = z(s)$ defined on a domain $\mathcal{S}$, i.e. $s \in \mathcal{S}$.
Some terminology needed later:

- z is of second order: $\text{var}(z_s) < \infty$ for all $s \in \mathcal{S}$,
- z is stationary in the weak sense: $E(z(s_1)) = E(z(s_2))$ is constant and the covariance function is translation invariant so that $\mathbf{COV}(z(s_1), z(s_2)) = C(s_1 - s_2)$ for all $s_1, s_2 \in \mathcal{S}$. C is called then the covariance function.
- z is isotropic in the wide sense: its expectation and its variance are invariant under rotation.

For a stationary and isotropic random field, the covariance function can be written: $C(x) = C(||x||)$.

Covariance functions

Covariance functions are a corner stone of spatial statistics and there exist many covariance functions for weakly stationary and isotropic random fields. Two popular choices are:

- A popular class of covariance model is the **Matérn** class:

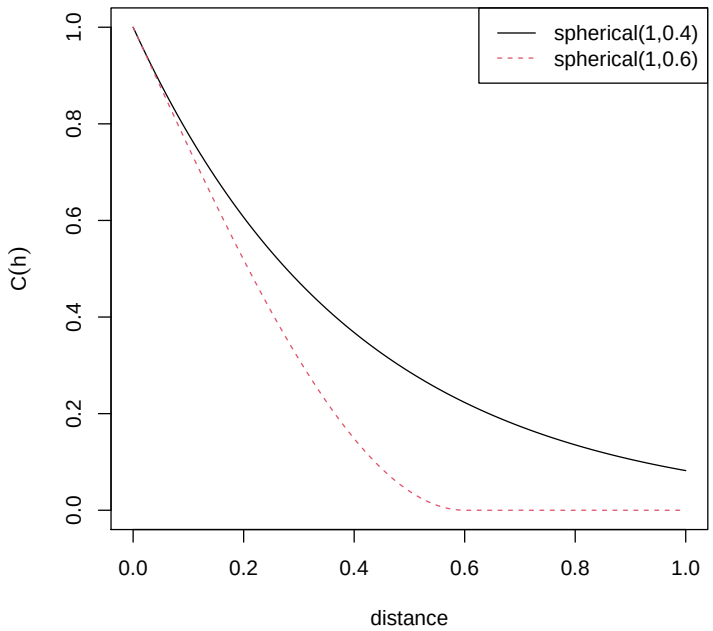
$$C(h) = \sigma^2 \frac{1}{\Gamma(h)} \left(\frac{\theta h}{2}\right)^2 2K_\nu(\theta h), \quad \text{where} \quad \nu > 0, \theta > 0$$

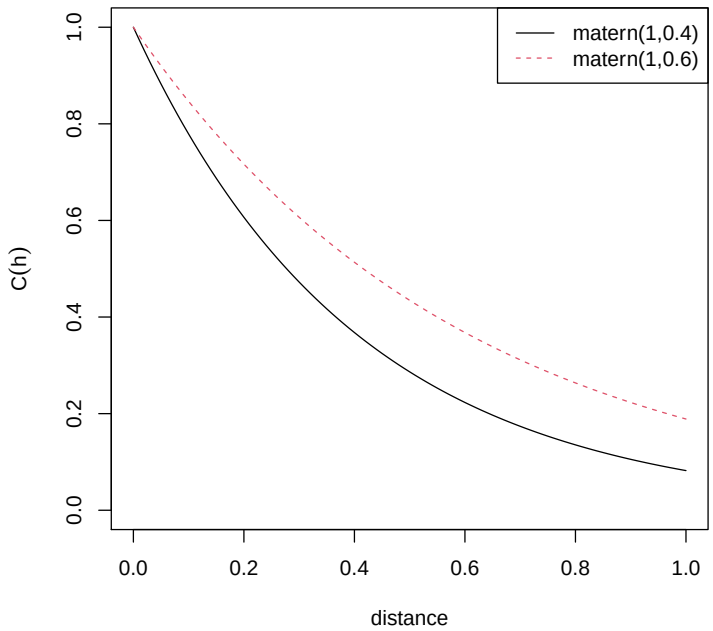
  and the function K_ν is a modified Bessel function of order ν.

- Another popular class is the **spherical** covariance model:

$$C(h) = \begin{cases} \sigma^2(1 - \frac{3}{2}\frac{h}{\phi} + \frac{1}{2}\left(\frac{h}{\phi}\right)^3) & \text{if} \quad h \leq \phi \\ 0 & \text{otherwise} \end{cases}$$

Covariance functions II





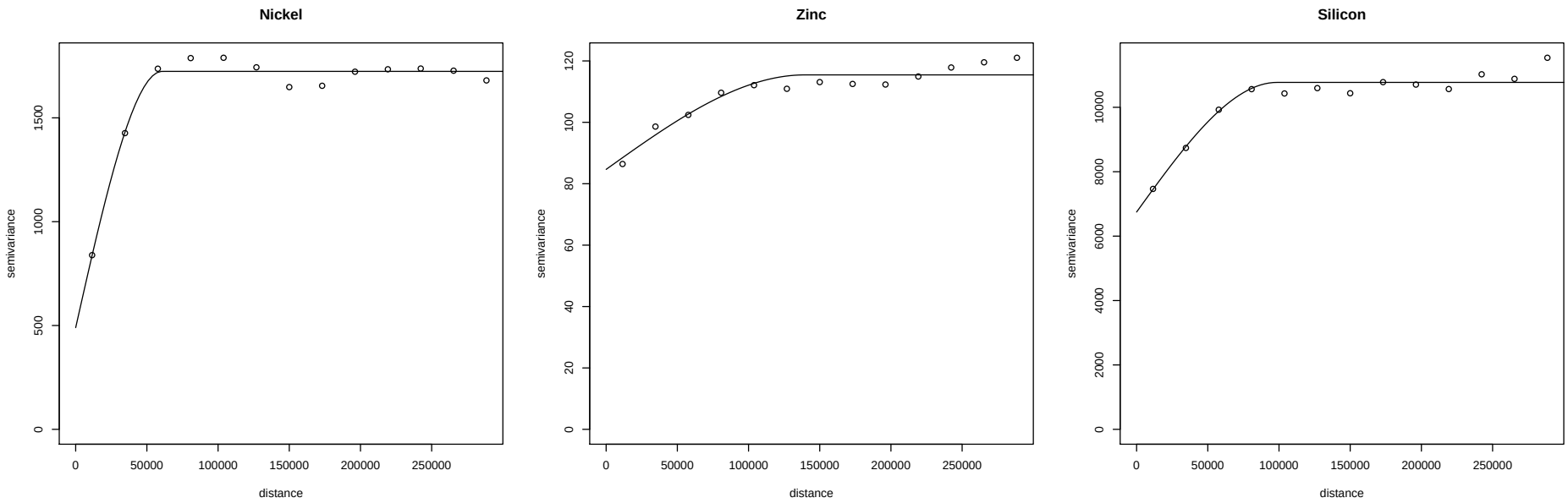
Semivariogram

Geoscientists usually however rather describe spatial dependence in the terms of a semivariogram γ.
For weakly stationary isotropic random fields we have the following relation between the covariance and the semivariogram:

$$\gamma(||s_i - s_j||) = C(0) - C(||s_i - s_j||).$$

- $\gamma(0) = 0$ and the function increases until the **sill** which is the value $\sigma^2 = \text{var}(z(s))$.
- The value h^* is called the **range** and gives the distance when observations start to be uncorrelated.
- The value h^{**} is called the **practical range** and is the distance at which the semivariogram achieves 95% of the sill.
- There may be a discontinuity at 0 which is called the **nugget effect** and is linked with measurement errors or/and micro-variability.

Semivariograms for selected chemical elements



Empirical semivariograms for three chemicals from the Kola data togehter with fitted spherical semivariograms functions.

Multivariate spatial data

Specifying multivariate spatial covariance functions is however much more difficult as these should have many specific properties and fulfil many conventions.
Multivariate spatial models are particular challenging as many parameters need to be fitted.
The maybe most popular multivariate spatial model for p-variate data, the **linear model of coregionalization (LMC)** is based on K univarite spatial correlation functions ρ_k, $k = 1, \ldots, K$ and yields the covariance matrix

$$C(h) = \sum_{k=1}^{K} \rho_k(h) T_k,$$

with $k < p$ and T_k's being non-negative definite $p \times p$ matrices.

Blind Source Separation Problem

The blind source separation problem in its most basic form assumes the model

$$\mathbf{x} = \mathbf{\Omega z},$$

where $\mathbf{x}$ is an observable p-variate vector, $\mathbf{\Omega}$ a full rank $p \times p$ mixing matrix and $\mathbf{z}$ an unobservable p-variate vector, the so-called sources. The goal of blind source separation (BSS) is to estimate $\mathbf{\Omega}$ in order to recover $\mathbf{z}$.

Clearly the BSS problem is not solvable without further assumptions on $\mathbf{z}$. Several approaches have been suggested for this purpose in the literature.

Motivation of BSS

One of the most common assumptions in BSS is that the components in $\mathbf{z}$ are uncorrelated or even independent.
The motivations behind BSS is then:

- The components of $\mathbf{z}$ have a (physical) meaning.
- Only a few components are actually of interest (dimension reduction).
- Especially under the independence assumption the components of $\mathbf{z}$ can be considered individually simplifying the modelling.

BSS is well established for iid data and multivariate time series.

In the following we suggest BSS for spatial data.

Spatial blind source separation model

Let $\mathbf{x}(s) = (x_1(s), \ldots, x_p(s))^\top$ be p-variate random field defined on a domain $\mathcal{S}$.

We say that $\mathbf{x}(s)$ follows the spatial BSS (SBSS) model if at any location s $\mathbf{x}(s)$ is a linear mixture of a p-variate latent field $\mathbf{z}(s) = (z_1(s), \ldots, z_p(s))^\top$, i.e.,

$$\mathbf{x}(s) = \mathbf{\Omega}\mathbf{z}(s), \tag{1}$$

where $\mathbf{\Omega}$ is an unknown $p \times p$ full rank matrix. We make the following assumptions:

(A1): $\mathbf{E}(\mathbf{z}(s)) = \mathbf{0}$ for $s \in \mathcal{S}$;

(A2): $\mathbf{COV}(\mathbf{z}(s)) = \mathbf{E}(\mathbf{z}(s)\mathbf{z}(s)^\top) = \mathbf{I}_p$;

(A3): $\mathbf{COV}(\mathbf{z}(s_1), \mathbf{z}(s_2)) = \mathbf{E}(\mathbf{z}(s_1)\mathbf{z}(s_2)^\top) = \mathbf{D}(s_1, s_2)$, where $\mathbf{D}$ is a diagonal matrix whose diagonal elements depend only on $s_1 - s_2$.

Spatial blind source separation model II

Assumption (A1) is made for convenience, Assumption (A2) requires uncorrelated components and fixes the scales of the latent fields and Assumption (A3) says that there are also no spatial cross-dependencies between the components.

These assumptions do however not uniquely define the model as the order and the signs of the components are not fixed. That is however common for all BSS methods and not considered a problem in practise.

In the following, let $\mathbf{COV}(z_k(s_i), z_k(s_j)) = K_k(s_i - s_j) = \mathbf{D}(s_i, s_j)_{k,k}$, where K_k denotes the stationary covariance function of z_k, for $k = 1, \ldots, p$.

Connection LMC and SBSS

There is a simple connection between the spatial blind source separation model and the linear model of coregionalization.

The covariance matrix $\mathbf{C_x}(h)$ resulting from a spatial blind source separation model is always symmetric and can be written as

$$\mathbf{C_x}(h) = \sum_{k=1}^{p} K_k(h)\mathbf{T}_k,$$

with $\mathbf{T}_k = \boldsymbol{\omega}_k\boldsymbol{\omega}_k^{\mathsf{T}}$, $\boldsymbol{\omega}_k$ being the kth column of $\boldsymbol{\Omega}$. Thus the spatial blind source separation model is a special case of the linear model of coregionalization with $r = p$ and where all coregionalization matrices $\mathbf{T}_k$, $k = 1, \ldots, p$, are rank one matrices.

Local covariance matrices

Assume n spatial locations $s_1, \ldots, s_n$.

The workhorse to solve the spatial BSS problem here will be **local covariance matrices** which are defined as

$$\mathbf{M}(f) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} f(s_i - s_j) \mathbf{E}(\mathbf{x}(s_i)\mathbf{x}(s_j)^\top),$$

where $f : \mathbb{R}^p \to \mathbb{R}$ is called a kernel function.
As a special case we denote

$$\mathbf{M}(f_0) = n^{-1} \sum_{i=1}^{n} \mathbf{E}(\mathbf{x}(s_i)\mathbf{x}(s_i)^\top).$$

Estimators of $\mathbf{M}(f)$ are denoted $\hat{\mathbf{M}}(f)$.

Kernels for spatial covariance matrices

There are many possible kernel functions for the local covariance functions. We consider here the following three:

1. Ball kernel: $B(h)(s) = I(||s|| \leq h)$ where $h > 0$ is a constant and $I(\cdot)$ denotes the indicator function.
2. Ring kernel: $R(h_1, h_2)(s) = I(h_1 < ||s|| \leq h_2)$, where $0 < h_1 \leq h_2$
3. Gaussian kernel: $G(h)(s) \equiv \exp\{-0.5(\Phi^{-1}(0.95||s||/h)^2\}$, where $\Phi^{-1}(x)$ is the quantile function of the standard normal distribution, making $G(h)$ have 95% of its total mass in the radius h ball around its center. Thus, $G(h)$ can be considered a smooth approximation of $B(h)$.

BSS unmixing matrix via simultaneous diagonalization

The first unmixing matrix functional we introduce is defined as

Definition

For any function $f : \mathbb{R}^d \to \mathbb{R}$, an unmixing matrix functional $\mathbf{\Gamma}(f)$ is defined as a functional which simultaneously diagonalizes $\mathbf{M}(f)$ and $\mathbf{M}(f_0)$ in the following way

$$\mathbf{\Gamma}(f)\mathbf{M}(f_0)\mathbf{\Gamma}(f)^\top = \mathbf{I}_p \quad \text{and} \quad \mathbf{\Gamma}(f)\mathbf{M}(f)\mathbf{\Gamma}(f)^\top = \mathbf{\Lambda}(f),$$

where $\mathbf{\Lambda}(f)$ is a diagonal matrix with diagonal elements in decreasing order.

Proposition

The unmixing problem given by f is identifiable if and only if the diagonal elements of $\mathbf{\Omega}^{-1}\mathbf{M}(f)\mathbf{\Omega}^{-\mathbf{T}}$ are distinct.

Assumptions for asymptotic considerations I

To derive asymptotic properties of the previous unmixing matrix functional we need besides Assumptions (A1)-(A3) some further assumptions:

(A4) The coordinates $z_1, \ldots, z_p$ of $\mathbf{z}$ are stationary Gaussian processes on $\mathbb{R}^d$;

(A5) A fixed $\Delta > 0$ exists so that, for all $n \in \mathbb{N}$ and, for all $i \neq j$, $i, j = 1, \ldots, n$, $\|s_i - s_j\| \geq \Delta$;

(A5) thus implies that $\mathcal{S}$ is unbounded which means we use an increasing domain asymptotic framework.

Assumptions for asymptotic considerations II

(A6) Fixed $A > 0$ and $\alpha > 0$ exist such that, for all $x \in \mathbb{R}^d$ and, for all $k = 1, \ldots, p$,

$$|K_k(x)| \leq \frac{A}{1 + \|x\|^{d+\alpha}};$$

(A7) Assuming Assumption (A6) holds, then for the same $A > 0$ and $\alpha > 0$ we have

$$|f(x)| \leq \frac{A}{1 + \|x\|^{d+\alpha}};$$

(A8) We have

$$\liminf_{n\to\infty} \min_{i=2,\ldots,p} \left[\left\{ \mathbf{\Omega}^{-1} \mathbf{M}(f) \mathbf{\Omega}^{-\top} \right\}_{i,i} - \left\{ \mathbf{\Omega}^{-1} \mathbf{M}(f) \mathbf{\Omega}^{-\top} \right\}_{i-1,i-1} \right] > 0.$$

Properties of $\hat{\mathbf{M}}(f)$

We can show show for the properties of $\hat{\mathbf{M}}(f)$:

Proposition

Suppose $n \to \infty$ and Assumptions (A1) to (A6) hold and let $f : \mathbb{R}^d \to \mathbb{R}$ satisfy Assumption (A7).
Then $\widehat{\mathbf{M}}(f) - \mathbf{M}(f) \to 0$ in probability when $n \to \infty$.

Properties of $\hat{\mathbf{M}}(f)$ II

Proposition

Assume the same assumptions as above. Let $\mathbf{w}(f)$ *be the vector of size* $p^2 \times 1$, *defined for* $i = (a-1)p + b$, $a, b \in \{1, \ldots, p\}$, *by*

$$\mathbf{w}(f)_i = n^{1/2}\{\widehat{\mathbf{M}}(f)_{a,b} - \mathbf{M}(f)_{a,b}\}.$$

Let $\mathbf{Q}_n$ *be the distribution of* $\{\mathbf{w}(f)^\top, \mathbf{w}(f_0)^\top\}^\top$. *Then, as* $n \to \infty$,

$$d_w[\mathbf{Q}_n, \mathcal{N}\{\mathbf{0}, \mathbf{V}(f, f_0)\}] \to \mathbf{0},$$

where $\mathcal{N}$ *denotes the normal distribution.*

Properties of $\hat{\mathbf{\Gamma}}(f)$ and $\hat{\mathbf{\Lambda}}(f)$

Proposition

Assume Assumptions (A1)-(A8) hold.
For $\widehat{\mathbf{\Gamma}}(f)$ *and* $\widehat{\mathbf{\Lambda}}(f)$ *let* $\mathbf{Q}_n$ *be the distribution of*

$$n^{1/2} \begin{pmatrix} \operatorname{vect}\left\{\widehat{\mathbf{\Gamma}}(f) - \mathbf{\Omega}^{-1}\right\} \\ \operatorname{diag}\left\{\widehat{\mathbf{\Lambda}}(f) - \mathbf{\Lambda}(f)\right\} \end{pmatrix}.$$

Then, we can choose $\widehat{\mathbf{\Gamma}}(f)$ *and* $\widehat{\mathbf{\Lambda}}(f)$ *so that when* $n \to \infty$,

$$d_w\{\mathbf{Q}_n, \mathcal{N}(\mathbf{0}, \mathbf{F}_1)\} \to \mathbf{0}.$$

BSS unmixing matrix via joint diagonalization

To improve the above estimator the idea is to use more than two kernel functions of the form $f_0, f_1, \ldots, f_k$ with $k \geq 2$.

Definition

For any $f_0, f_1, \ldots, f_k$, an unmixing matrix functional $\mathbf{\Gamma}$ can be formulated as

$$\widehat{\mathbf{\Gamma}} \in \underset{\substack{\mathbf{\Gamma}:\mathbf{\Gamma}\widehat{\mathbf{M}}(f_0)\mathbf{\Gamma}^\top=\mathbf{I}_p \\ \mathbf{\Gamma} \text{ has rows } \gamma_1^\top,\ldots,\gamma_p^\top}}{\operatorname{argmax}} \sum_{l=1}^{k}\sum_{j=1}^{p}\{\gamma_j^\top \widehat{\mathbf{M}}(f_l)\gamma_j\}^2. \tag{2}$$

Proposition

The unmixing problem given by $f_0, \ldots, f_k$ is identifiable if and only if for every pair $i \neq j$, $i, j = 1, \ldots, p$, there exists $l = 1, \ldots, k$ such that $\{\mathbf{\Omega}^{-1}\mathbf{M}(f_l)\mathbf{\Omega}^{-\mathbf{T}}\}_{i,i} \neq \{\mathbf{\Omega}^{-1}M(f_l)\mathbf{\Omega}^{-\mathbf{T}}\}_{j,j}$.

Assumptions for asymptotic considerations III

As weeker assumptions are needed for the identifiability for the unmixing matrix based on joint diagonalization Assumption (A8) can be relaxed to

(A9) A fixed $\delta > 0$ and $n_0 \in \mathbb{N}$ exist so that for all $n \in \mathbb{N}$, $n \geq n_0$, for every pair $i \neq j$, $i, j = 1, \ldots, p$, there exists $l = 1, \ldots, k$, such that $|\{\mathbf{\Omega}^{-1}\mathbf{M}(f_l)\mathbf{\Omega}^{-\mathbf{T}}\}_{i,i} - \{\mathbf{\Omega}^{-1}\mathbf{M}(f_l)\mathbf{\Omega}^{-\mathbf{T}}\}_{j,j}| \geq \delta$.

Consistency of $\widehat{\mathbf{\Gamma}}$

Proposition

Let $k \in \mathbb{N}$ be fixed and Assumptions (A1)-(A7) and (A9) hold.
Let $\widehat{\mathbf{\Gamma}} = \widehat{\mathbf{\Gamma}}\{\widehat{\mathbf{M}}(f_0), \widehat{\mathbf{M}}(f_1), \ldots, \widehat{\mathbf{M}}(f_k)\}$.
Then we can choose $\widehat{\mathbf{\Gamma}}$ so that $\widehat{\mathbf{\Gamma}} \to \mathbf{\Omega}^{-1}$ in probability when n goes to infinity.

Asymptotic normality of $\widehat{\mathbf{\Gamma}}$

Proposition

Assume the same assumptions as in the previous proposition. Let $(\widehat{\mathbf{\Gamma}}_n)_{n\in\mathbb{N}}$ *be any sequence of* $p \times p$ *matrices so that for any* $n \in \mathbb{N}$, $\widehat{\mathbf{\Gamma}}_n = \widehat{\mathbf{\Gamma}}_n\{\widehat{\mathbf{M}}(f_0), \widehat{\mathbf{M}}(f_1), \ldots, \widehat{\mathbf{M}}(f_k)\}$. *Then, a sequence of permutation matrices* $(\mathbf{P}_n)$ *and a sequence of diagonal matrices* $(\mathbf{D}_n)$ *exist, with diagonal components in* $\{-1, 1\}$, *so that the distribution* $\mathbf{Q}_n$ *of* $n^{1/2}\mathrm{vect}(\check{\mathbf{\Gamma}}_n - \mathbf{\Omega}^{-1})$ *with* $\check{\mathbf{\Gamma}}_n = \mathbf{D}_n\mathbf{P}_n\widehat{\mathbf{\Gamma}}_n$ *satisfies, as* $n \to \infty$,

$$d_w\{\mathbf{Q}_n, \mathcal{N}(\mathbf{0}, \mathbf{F}_k)\} \to \mathbf{0}.$$

Simulation study I

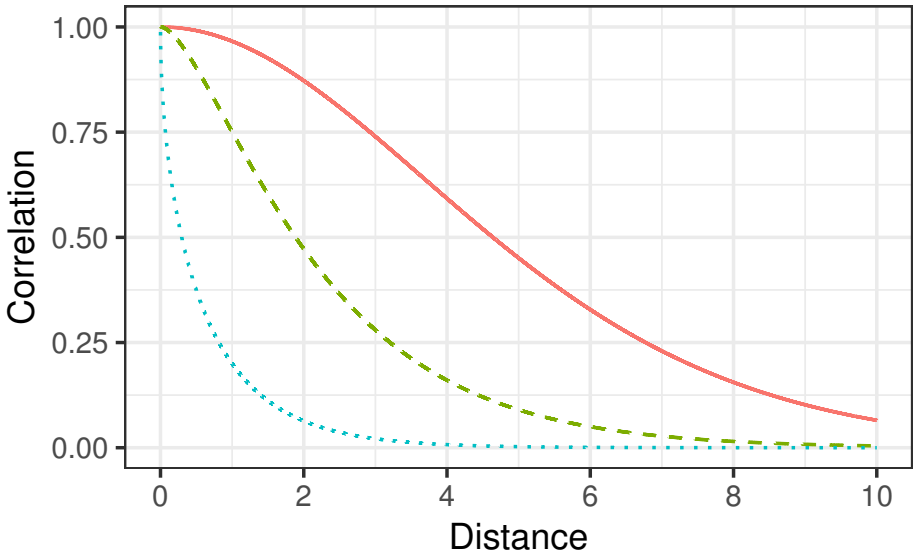
To verify the asymptotic properties of our estimators we conducted a simulation study using always 3-variate random fields and different designs for the locations mimicking an increasing domain framework and all components have Matern covariance functions with different shape and range parameters.

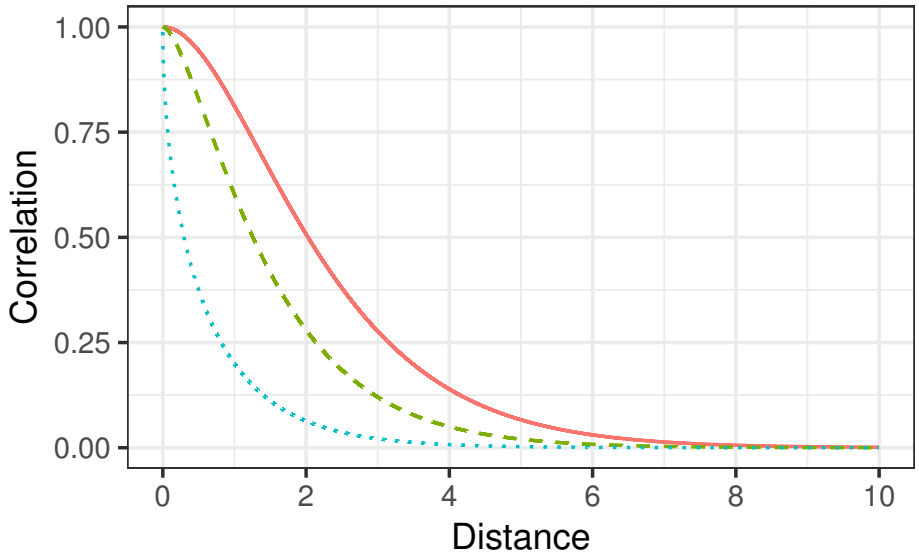
The performance measure used in the simulations is the minimum distance index

$$\mathrm{MDI}(\widehat{\mathbf{\Gamma}}) = (p-1)^{-1/2} \inf\{\|\mathbf{C}\widehat{\mathbf{\Gamma}}\mathbf{\Omega} - \mathbf{I}_p\|, \mathbf{C} \in \mathcal{C}\},$$

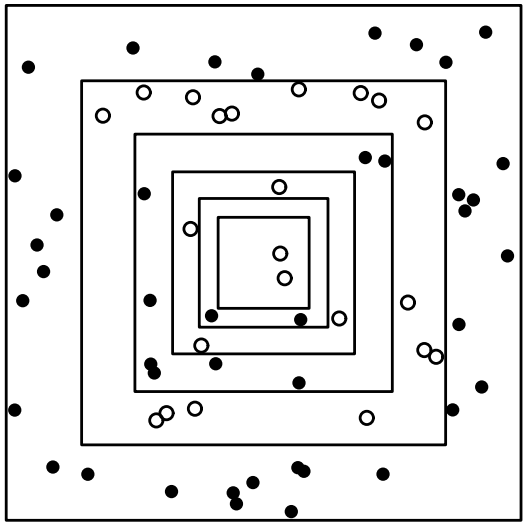
where $\mathcal{C}$ is the set of all matrices with exactly one non-zero element in each row and column and $\|\cdot\|$ is the Frobenius norm.

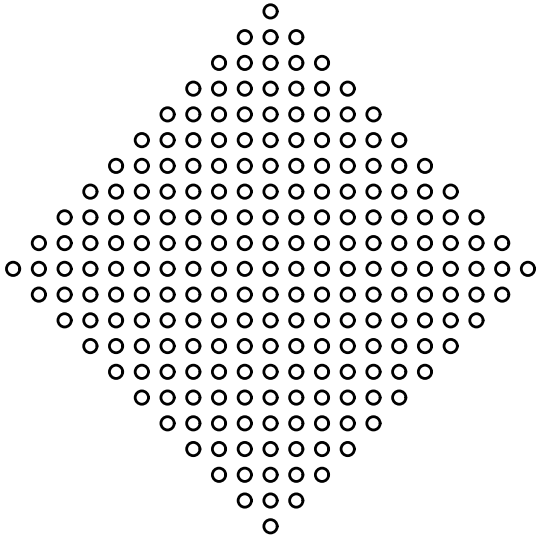
Used Matern functions

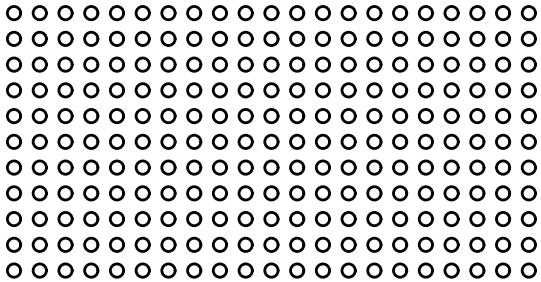




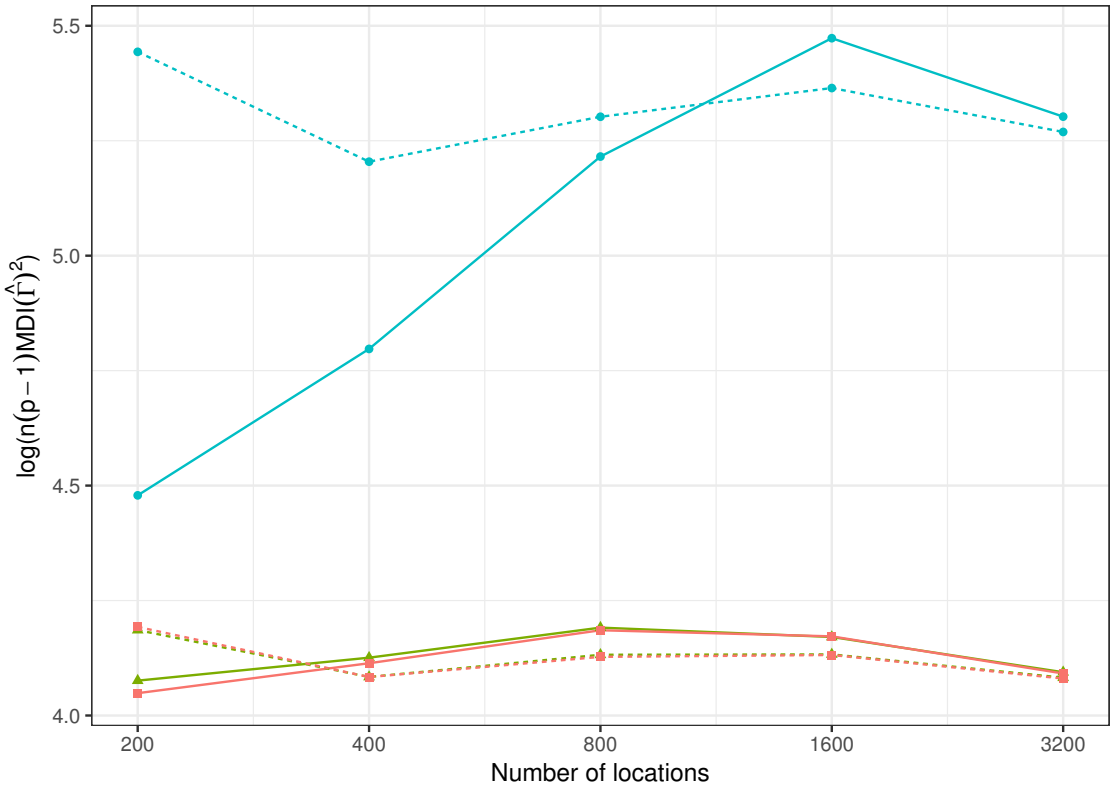
Increasing domain settings







Results I



The solid lines give the mean values of $n(p-1)\,\mathrm{MDI}(\widehat{\Gamma})^2$ in the first simulation and the dashed lines correspond to the asymptotic approximations of the same quantities. The three used local covariance matrices are $B(1)$ (blue line), $R(1,2)$ (green line) and $\{B(1), R(1,2)\}$ (red line).

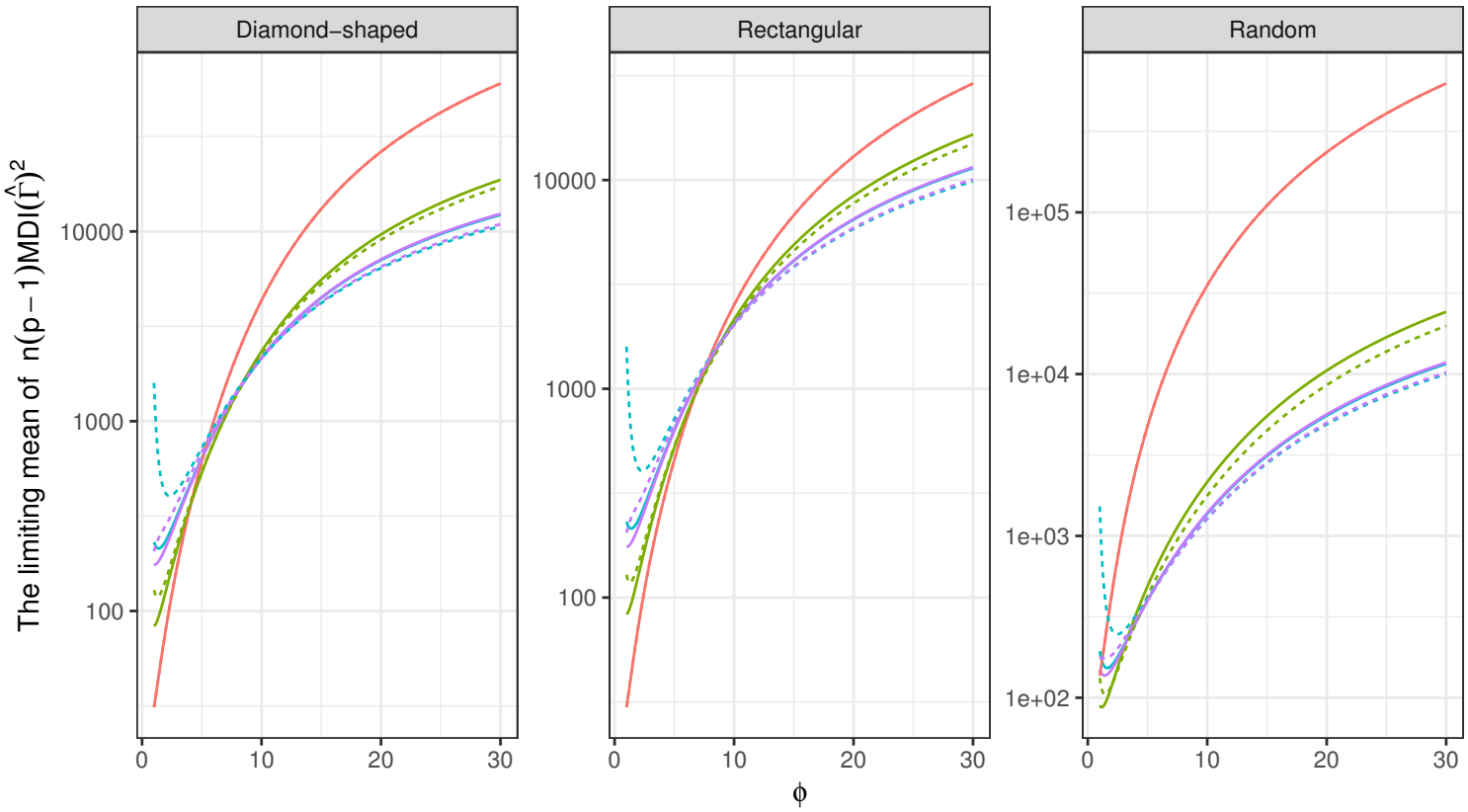
Simulation study II

The second simulations study explores the effect of the range of the latent fields and compares the choice of the different local covariances.

We consider a total of eight different local covariance matrices, $B(r), R(r-1,r)$ for $r = 1, 3, 5$, and the joint diagonalisations of the previous sets: $\{B(1), B(3), B(5)\}$ and $\{R(0,1), R(2,3), R(4,5)\}$.

Comparisons are based on the asymptotic approximations to the distribution of $n(p-1)\,\mathrm{MDI}(\widehat{\mathbf{\Gamma}})^2$.

Results II



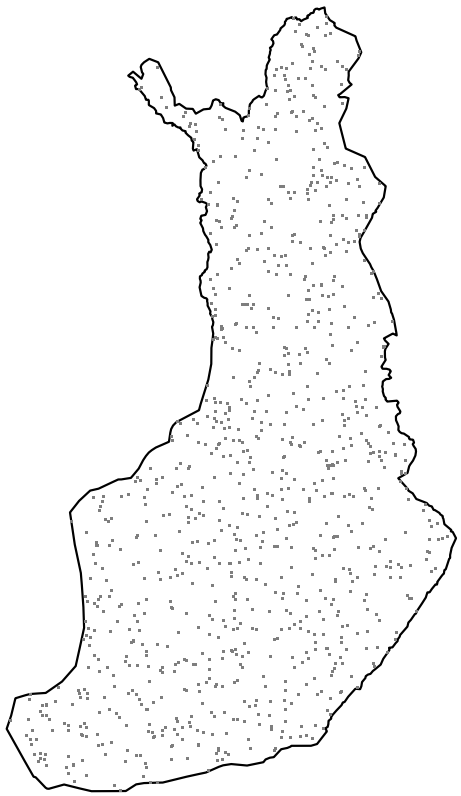
The solid and the dashed lines correspond, respectively, to the ball and ring kernels and the value of the parameter r is indicated by the color of the line as follows: 1 (red), 3 (green), 5 (blue), J (purple). The y-axis has a logarithmic scale.

Simulation study III

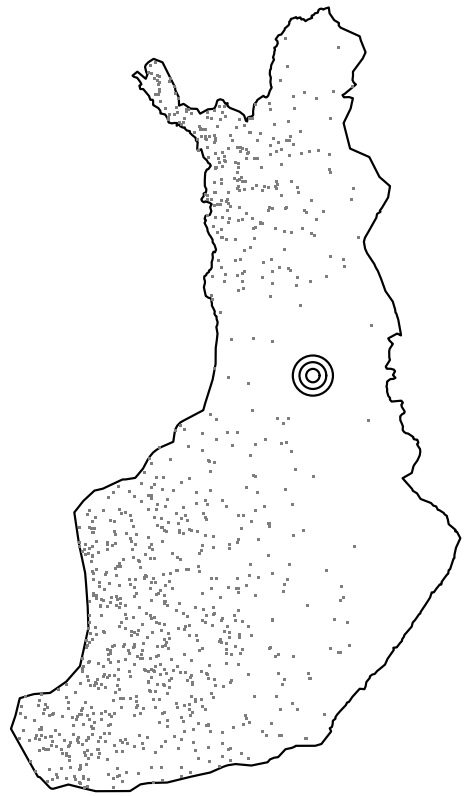
For efficiency comparisons for many different local covariance matrices we considered a two different fixed locations settings based on the map of Finland using 1000 locations.
Estimates were obtained with the local covariance matrix kernels $B(r), R(r-10, r), G(r)$, where $r = 10, 20, 30, 100$, and the joint diagonalisation of each of the three quadruplets $\{B(10), B(20), B(30), B(100\}$, $\{R(10), R(20), R(30), R(100\}$ and $\{G(10), G(20), G(30), G(100\}$ adding up to a total of 15 estimators.
For 2000 Repetitions the MDI indices were computed.

Designs for efficiency simulations

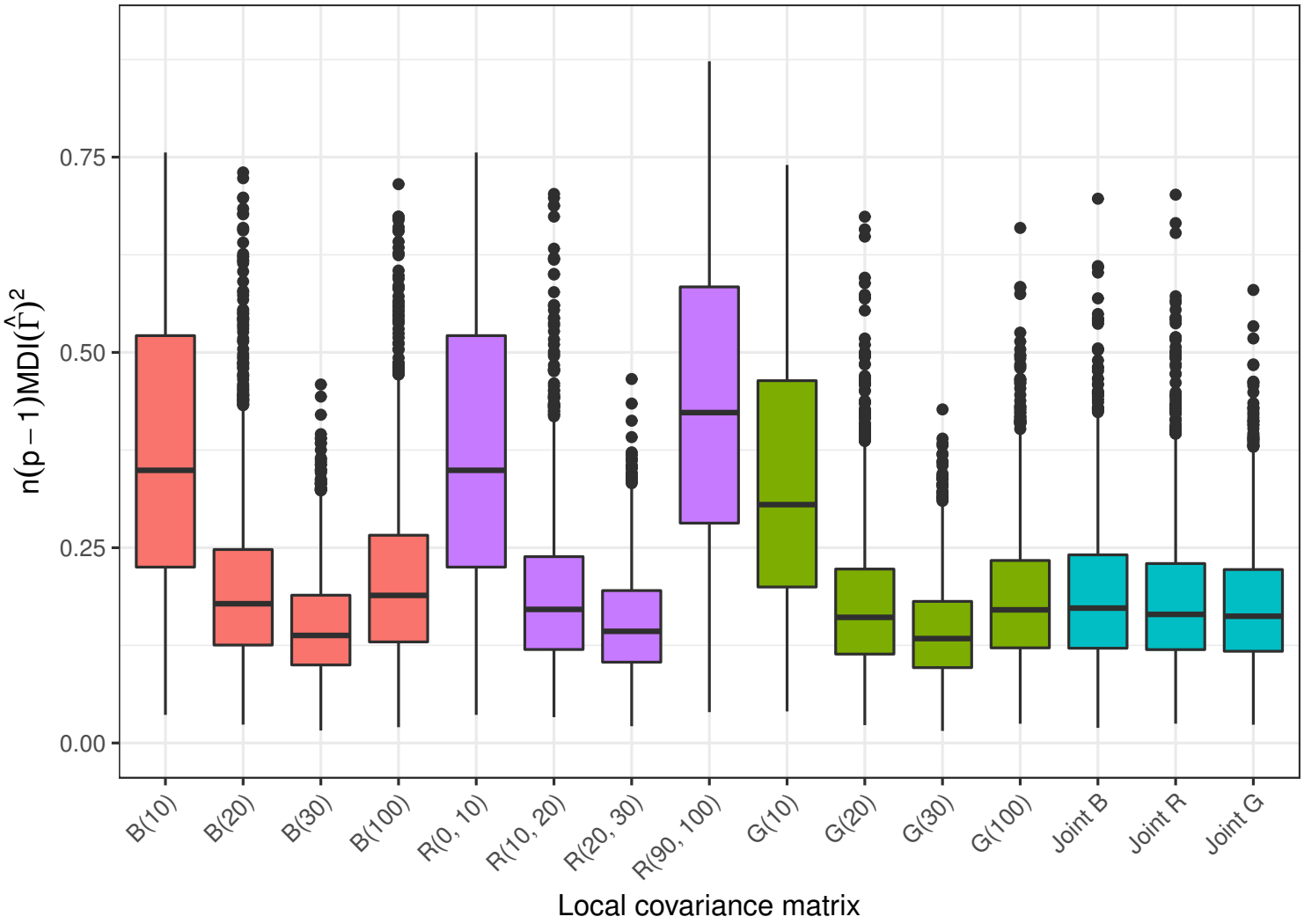
Uniform



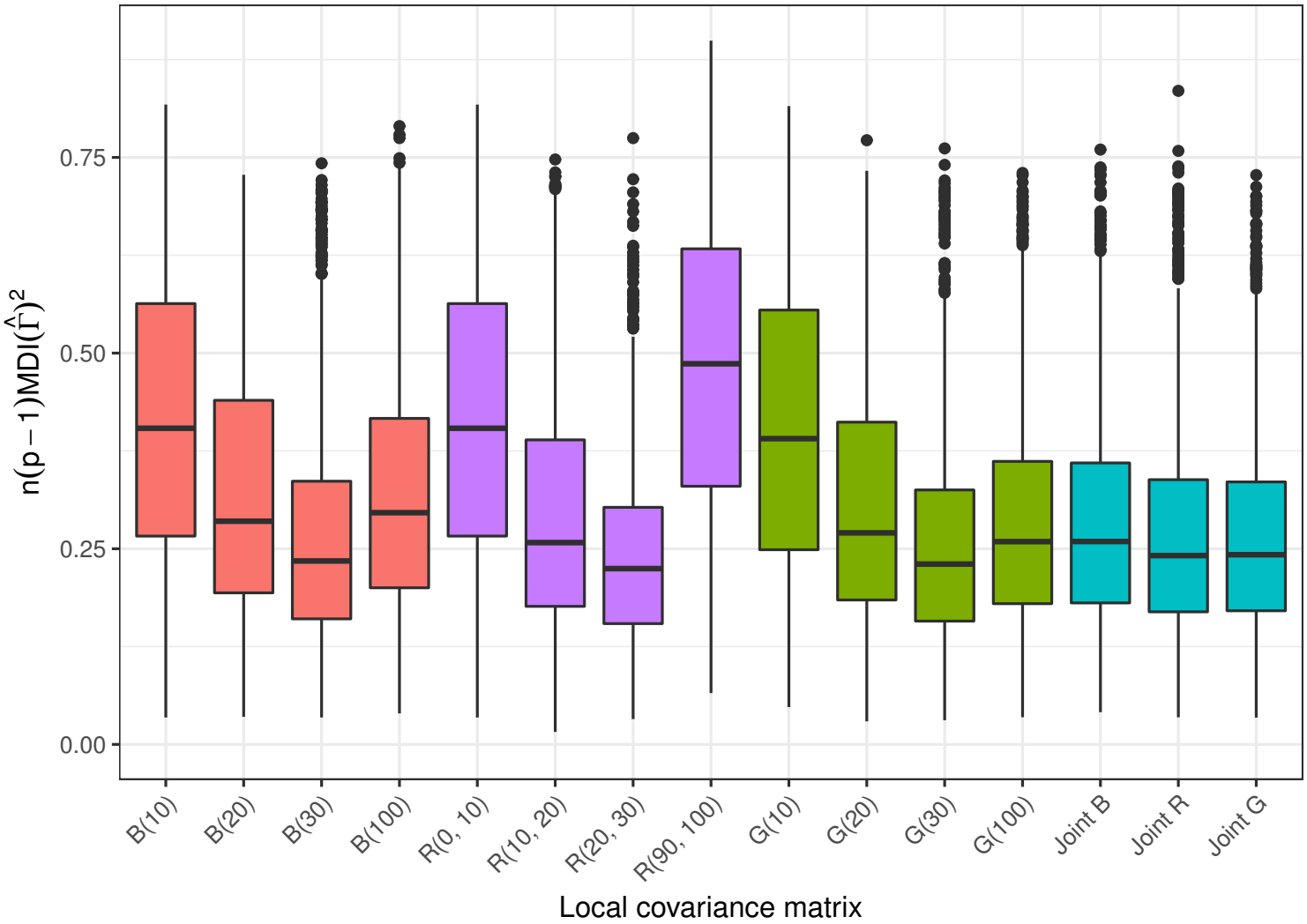
Skew



Results III (Finland map uniform)



Results III (Finland map skew)



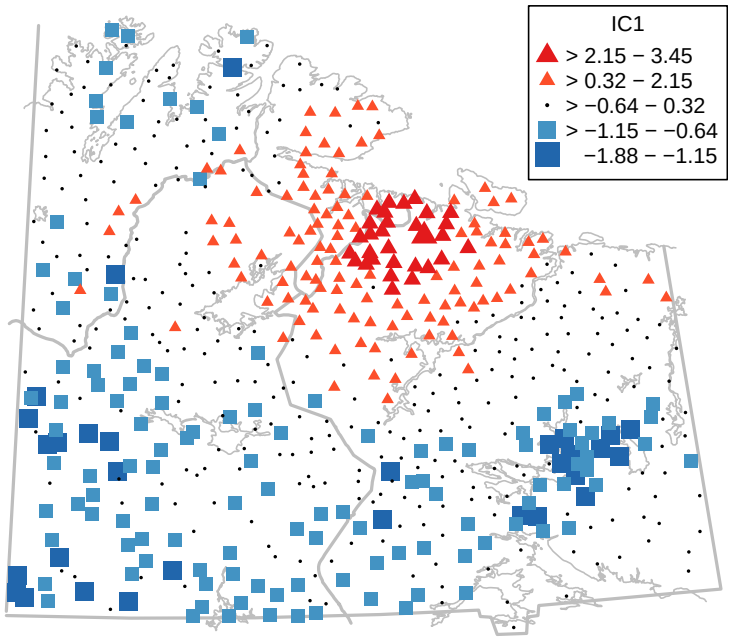
SBSS for the Kola data I

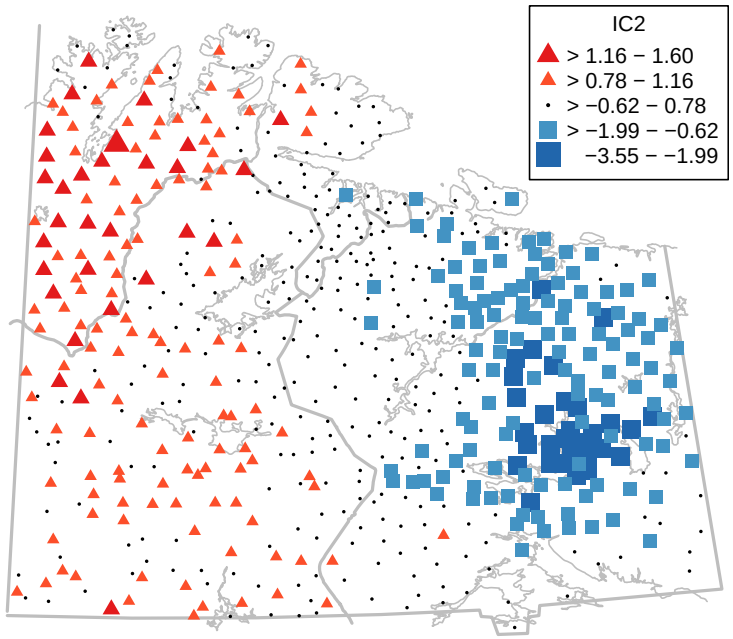
We applied SBSS to the Kola data (ilr transformation).
Together with an subject expert, when using only one local covariance matrix, a good option seemed to be a ball kernel with radius 50km. This yielded six meaningful components.

However using four ring kernels R(0,25), R(25,50), R(50,75) and R(75,100) yielded more or less the same six components of interest.

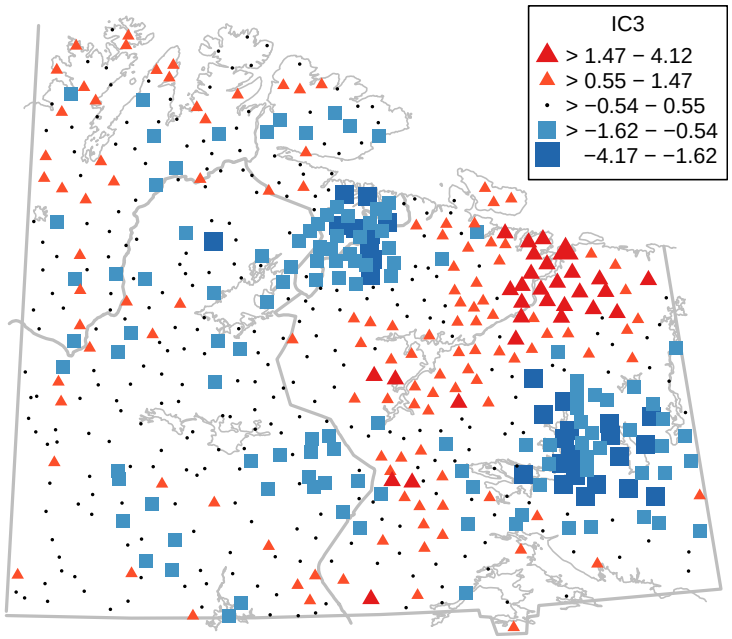
The six components of interest based on the joint diagonalization are shown on the following slides.

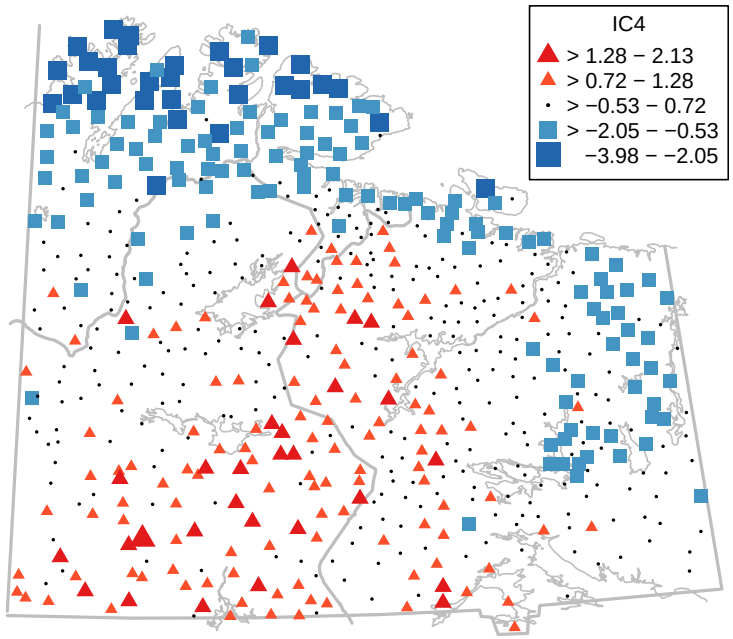
SBSS for the Kola data II



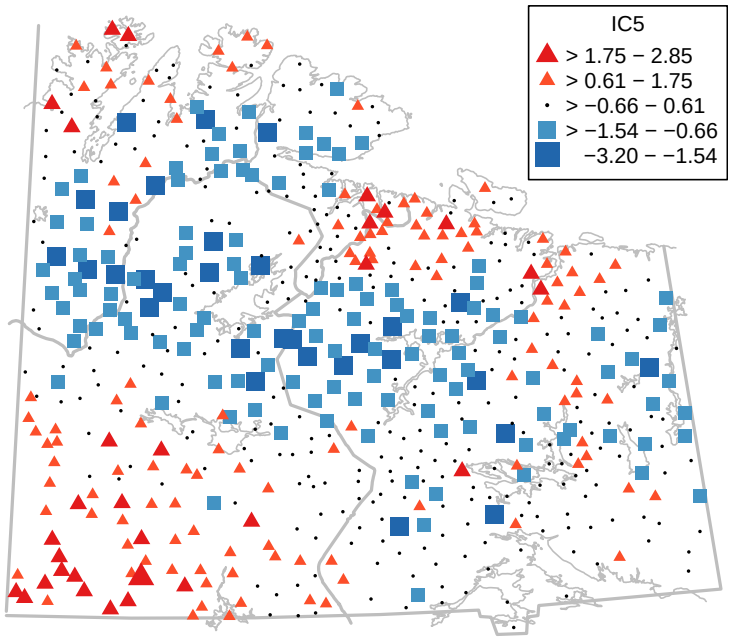


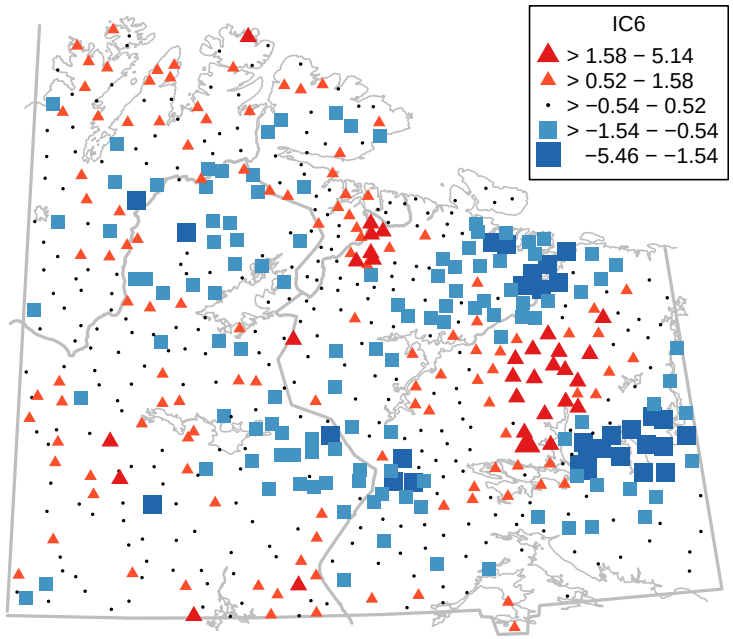
SBSS for the Kola data III





SBSS for the Kola data IV





Comparison of spatial prediction methods

Univariate spatial prediction is usually called **kriging** and multivariate spatial prediction **cokriging** and there many variants of (co-) kriging.

Common to all of them is however that spatial covariance functions need to be specified.

In the following simulation study the goal is to compare cokringing against univariate kriging based on SBSS. As a reference we made also prediction using neural networks.

Simulation setting for predictions study

Setting 1: 3-variate SBSS model with 3 Gaussian fields having $C_1(Z_1) = C_{sp}(h; 1, 2)$, $C_2(Z_2) = C_m(h; 1, 0.5, 2)$ and $C_3(Z_3) = C_m(h; 1, 1, 2)$ as covariance functions.

Setting 2: 3-variate SBSS model with 3 t_5 fields having $C_1(Z_1) = C_{sp}(h; 1, 2)$, $C_2(Z_2) = C_m(h; 1, 0.5, 2)$ and $C_3(Z_3) = C_m(h; 1, 1, 2)$ as covariance functions.

Setting 3: 3-variate parsimonious multivariate Matérn model (PMat)

In all settings the mixing matrices entries of the $\mathbf{\Omega}$'s were drawn from $N(0, 1)$ for all simulations.

Simulation setting for predictions study

To select the spatial locations we consider two variants:

uniform x and y coordinates are independently sampled from $\mathcal{U}(0, 1)$.

skew x and y coordinates are independently sampled from $\mathcal{B}(2, 5)$.

Then all coordinates are multiplied by 35 to obtain the sampling domain $\mathcal{S} = [0, 35] \times [0, 35]$.
We sample for each variant $n_{\mathcal{S}} = 1225$ sites.
In additionally define a grid $\mathcal{S}^* = \{(x + 0.5, y + 0.5) : x, y \in \mathbb{Z} \cap [0, 35]\}$ which again has $n_{\mathcal{S}^*} = 1225$ sites.
For every iteration of the simulation we simulate the field for each coordinate variant on $\mathcal{S}$ and $\mathcal{S}^*$ at once and use for each coordinate variant the values of the field on $\mathcal{S}$ to predict the values on $\mathcal{S}^*$.
Mean squared error (MSE) between the predictions and simulated (true) values on $\mathcal{S}^*$ are used as the performance measure of the spatial prediction.

Prediction performance

Field Setting	Variant	BN	SNN	LMC + Cokriging	SBSS + Spheric	SBSS + Matérn
SBSS Normal	uniform	0.66(0.03)	0.63(0.03)	0.57(0.03)	0.49(0.02)	**0.47(0.02)**
	skew	1.20(0.14)	1.16(0.13)	0.96(0.08)	0.94(0.10)	**0.90(0.08)**
SBSS t5	uniform	1.40(0.15)	1.32(0.15)	1.26(0.15)	1.10(0.13)	**1.07(0.13)**
	skew	2.22(0.31)	2.13(0.29)	1.82(0.21)	1.77(0.24)	**1.72(0.21)**
PMat	uniform	0.34(0.02)	0.33(0.02)	**0.29(0.01)**	0.30(0.02)	0.31(0.02)
	skew	0.56(0.07)	0.54(0.07)	0.45(0.04)	**0.44(0.04)**	**0.44(0.04)**

Key References

Nordhausen, K., Oja, H., Filzmoser, P. and Reimann, C. (2015): Blind Source Separation for Spatially Correlated Compositional Data. Mathematical Geosciences, 47, 753–770.

Bachoc, F., Genton M.G., Nordhausen, K.,Ruiz-Gazen A. and Virta, J. (2020): Spatial Blind Source Separation. Biometrika, 107, 627–646.

Mühlmann, C., Nordhausen, K. and Yi, M.: On CoKriging, Neural Networks, and Spatial Blind Source Separation for Multivariate Spatial Prediction. To appear in IEEE Geoscience and Remote Sensing Letters.

Mühlmann, C., Nordhausen, K. and Virta, J..: SpatialBSS: Blind Source Separation for Multivariate Spatial Data. `https://CRAN.R-project.org/package=SpatialBSS`.

Thank you for your attention!

klaus.nordhausen@tuwien.ac.at