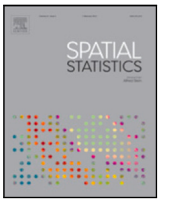
Spatial Statistics 59 (2024) 100803

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



Robust second-order stationary spatial blind source separation using generalized sign matrices

Mika Sipilä [a], Christoph Muehlmann [b], Klaus Nordhausen [a,*], Sara Taskinen [a]

[a] Department of Mathematics and Statistics, University of Jyväskylä, Finland
[b] Institute of Statistics & Mathematical, Vienna University of Technology, Austria

ARTICLE INFO

Keywords:
Affine equivariance
Bias
Multivariate spatial data
Scatter matrix
Spatial signs

ABSTRACT

Consider a spatial blind source separation model in which the observed multivariate spatial data are assumed to be a linear mixture of latent stationary spatially uncorrelated random fields. The objective is to recover an unknown mixing procedure as well as the latent random fields. Recently, spatial blind source separation methods that are based on the simultaneous diagonalization of two or more scatter matrices were proposed. In cases involving uncontaminated data, such methods can solve the blind source separation problem, however, in the presence of outlying observations, these methods perform poorly. We propose a robust blind source separation method that employs robust global and local covariance matrices based on generalized spatial signs in simultaneous diagonalization. Simulation studies are employed to illustrate the robustness and efficiency of the proposed methods in various scenarios.

1. Introduction

Many modern datasets consist of multivariate, spatially indexed data sourced from geographical locations. For a proper statistical analysis, one must account for the Tobler's (1970) first law of geography which states that 'everything is related to everything else, but near things are more related than distant things.' Hitherto, for univariate spatial data, considering the spatial dependence is quite challenging since models need to be selected for the spatial covariance function. Such models conventionally have many parameters to be estimated and to tackle the problem, it is frequently assumed that the data are stationary. Kleiber and Nychka (2012), Genton and Kleiber (2015) point out that the model selection becomes even more challenging when the data are multivariate as the number of parameters needed to model the additional cross-dependencies increases dramatically with increasing dimension.

Therefore, Nordhausen et al. (2015) suggested a blind source separation (BSS) approach for multivariate stationary spatial data where it is assumed that the observable p-variate random field is a mixture of p independent latent fields. Nordhausen et al. (2015) suggest then, based on the simultaneous diagonalization of two moment-based matrices, a spatial blind source separation (SBSS) method for estimating the p latent fields, which can subsequently be interpreted and modeled individually. The original SBSS approach has since then been extended for example in Bachoc et al. (2020), Muehlmann et al. (2022a). Muehlmann et al. (2021b) showed that the SBSS approach can be beneficial for multivariate spatial prediction.

All the SBSS methods suggested thus far are based on second moments making them preferable for Gaussian latent fields. However, these methods will most likely be inefficient for heavy-tailed latent fields and highly susceptible to outliers which are often present in spatial data, as pointed out by Harris et al. (2014), Filzmoser et al. (2014), Ernst and Haesbroeck (2017). Robust statistical methods are designed to address these issues. The methods are less sensitive to outliers and work better for other distributions than

* Corresponding author.
E-mail address: klaus.k.nordhausen@jyu.fi (K. Nordhausen).

https://doi.org/10.1016/j.spasta.2023.100803
Received 14 June 2023; Received in revised form 8 December 2023; Accepted 8 December 2023
Available online 16 December 2023

the Gaussian one. For a general overview of the robust methods, we refer to Huber and Ronchetti (2011), Hampel et al. (2011), Maronna et al. (2019), where the focus is frequently on independent and identically distributed elliptical data. In this paper we suggest robust SBSS methods that are based on the joint diagonalization of two or more matrices. We proceed as with SBSS, but we replace the non-robust matrices with matrices based on spatial signs (for details, see Oja, 2010) and their extensions, as introduced in Raymaekers and Rousseeuw (2019), by using ideas from the robust temporal BSS methods, as suggested in Ilmonen et al. (2015).

This paper is structured as follows. In Section 2 we recall the SBSS model and classical, non-robust covariance matrices that can be used to solve the SBSS problem. Thereafter, the robust SBSS methods based on generalized spatial signs are introduced. In Section 3, we compare the performances of the three robust candidates against their non-robust counterparts under different distributional and contamination conditions. Additionally, bias and maximum bias curves are provided for comparing the robustness properties. In Section 4 the methods are compared using a real data example, after which the paper is concluded in Section 5.

2. Robust SBSS

2.1. SBSS model

Let $\mathbf{x}(\mathbf{s}) = \{x_1(\mathbf{s}), \dots, x_p(\mathbf{s})\}^\top$ be a p-dimensional random field where $\mathbf{s} \in S \subseteq \mathbb{R}^d$ denotes the location of the observation $\mathbf{x}(\mathbf{s})$ and S denotes the spatial domain. We assume that $\mathbf{x}(\mathbf{s})$ follows the SBSS model, i.e. $\mathbf{x}(\mathbf{s})$ is a linear mixture of an underlying p-variate latent field $\mathbf{z}(\mathbf{s}) = \{z_1(\mathbf{s}), \dots, z_p(\mathbf{s})\}^\top$ with independent components. We formalize this below.

Definition 1. An observable random field $\mathbf{x}(\mathbf{s}) \in \mathbb{R}^p$, where $\mathbf{s} \in S$, follows the SBSS model if

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

where full-rank $p \times p$ matrix $\boldsymbol{\Omega}$ is an unknown mixing matrix and p-vector $\boldsymbol{\mu}$ is a nuisance parameter specifying the location. A latent p-variate random field $\mathbf{z}(\mathbf{s})$ satisfies the following properties

SBSS1: $\mathrm{E}(\mathbf{z}(\mathbf{s})) = \mathbf{0}$ and $\mathrm{COV}(\mathbf{z}(\mathbf{s})) = \mathbf{I}_p$, for all $\mathbf{s} \in S$

SBSS2: $z_1(\mathbf{s}), \dots, z_p(\mathbf{s})$ are mutually independent and $\mathrm{COV}\left(\mathbf{z}(\mathbf{s}), \mathbf{z}(\mathbf{s}')\right) = \boldsymbol{\Sigma}(h)$, where $\boldsymbol{\Sigma}$ is a diagonal matrix with diagonal elements depending on $h = \|\mathbf{s} - \mathbf{s}'\|$ for all $\mathbf{s}, \mathbf{s}' \in S$.

Above $\|\cdot\|$ denotes the Euclidean norm. Assumption **SBSS1** fixes the scales and locations of the latent components and is applied without loss of generality. Assumption **SBSS2** states that the components are independent and that there are no spatial cross-dependences between the latent components. Thus, we assume that the latent components are second-order stationary. Notice that the above assumptions imply that the latent components are uniquely defined only up to permutation and sign. This is sufficient for most practical applications. In the following, the sign changes will be represented by a $p \times p$ sign-change matrix $\mathbf{J} \in \mathcal{J}$, where $\mathcal{J}$ is a set of all possible $p \times p$ sign change matrices, i.e. diagonal matrices with diagonal elements being equal to ± 1. The reordering of components is expressed via a $p \times p$ permutation matrix $\mathbf{P} \in \mathcal{P}$, where $\mathcal{P}$ is a set of all $p \times p$ matrices having in each row and each column exactly one value 1 with the rest of the values being 0. Componentwise heterogeneous rescaling is expressed via a $p \times p$ diagonal matrix $\mathbf{C} \in \mathcal{C}$, where $\mathcal{C}$ is a set of all $p \times p$ diagonal matrices having strictly positive diagonal elements.

The main goal of the SBSS method is to estimate a $p \times p$ matrix (e.g. $\mathbf{W} = \boldsymbol{\Omega}^{-1}$) based on $\mathbf{x}(\mathbf{s})$ only so that it can be used to recover the latent sources in $\mathbf{z}(\mathbf{s})$. We call $\mathbf{W}$ an unmixing matrix, and it is formalized as a functional as follows (Miettinen et al., 2015).

Definition 2. For a random field, $\mathbf{x} = \mathbf{x}(\mathbf{s})$, following SBSS model specified in (1), a $p \times p$ matrix valued functional, $\mathbf{W}$ is an unmixing matrix functional if it satisfies the following:

(i) $\mathbf{W}(\mathbf{x} - \mathbf{T})$, where $\mathbf{T}$ is a location functional, recovers the latent components up to the permutation, sign change and scale, i.e. $\mathbf{W}(\mathbf{x} - \mathbf{T}) = \mathbf{PJCz}$, where $\mathbf{z} = \mathbf{z}(\mathbf{s})$ and $\mathbf{P} \in \mathcal{P}$, $\mathbf{J} \in \mathcal{J}$ and $\mathbf{C} \in \mathcal{C}$.

(ii) $\mathbf{W}$ is affine equivariant in the sense that if $\mathbf{x}$ is a random field and $\mathbf{x}^\star = \mathbf{Ax} + \mathbf{b}$ is its affine transformation, where $\mathbf{A}$ is an invertible $p \times p$ matrix and $\mathbf{b}$ is a p-vector, then it holds that if $\mathbf{W}$ and $\mathbf{W}^\star$ are computed from $\mathbf{x}$ and $\mathbf{x}^\star$, respectively, then $\mathbf{W}^\star = \mathbf{PJWA}^{-1}$, for some $\mathbf{P} \in \mathcal{P}$ and $\mathbf{J} \in \mathcal{J}$.

The above definition implies that if $\mathbf{W}$ is an unmixing matrix functional, so is $\mathbf{PJCW}$ for all transformation matrices $\mathbf{P} \in \mathcal{P}$, $\mathbf{J} \in \mathcal{J}$ and $\mathbf{C} \in \mathcal{C}$. Thus, we have a whole set of matrices equivalent to $\mathbf{W}$. The scales of unmixing matrices can be fixed in the estimation procedure, and often it is required that the scales are fixed in such a way that $\mathrm{COV}(\mathbf{PJCW}(\mathbf{x} - \mathbf{T})) = \mathbf{I}_p$, however, the order and signs of the rows of $\mathbf{W}$ still remain unidentified. Similarly, if a location functional other than the mean is used, the centering may be performed differently when some of the components are skew. For a detailed discussion on the identifiability of unmixing matrix functionals, see e.g. Ilmonen and Paindaveine (2011) and references therein.

Next, we review the local covariance matrices that are proposed in the literature to solve the SBSS problem and provide their robust extensions.

2.2. Local covariance matrix functionals

BSS problems are frequently solved by the simultaneous diagonalization of two or more scatter matrices (Nordhausen and Oja, 2018; Nordhausen and Ruiz-Gazen, 2022). In Nordhausen et al. (2015), the estimation of $\mathbf{W}$ is based on the use of local covariance matrix functionals (LCOV). Let us assume that we have n fixed spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ in S which index the random fields by $\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_n)$. For simplicity, we write $\mathbf{x}_i = \mathbf{x}(\mathbf{s}_i)$. The LCOV is defined as

$$\mathrm{LCOV}(f) = \frac{1}{n\sqrt{F_{n,f}}} \sum_{i=1}^{n} \sum_{j=1}^{n} f(\mathbf{s}_i - \mathbf{s}_j) \mathrm{E}\left(\left(\mathbf{x}_i - \mathrm{E}(\mathbf{x}_i)\right) \left(\mathbf{x}_j - \mathrm{E}(\mathbf{x}_j)\right)^\top \right), \tag{2}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a kernel function and

$$F_{n,f} = \frac{1}{n} \sum_{i,j=1}^{n} f^2(\mathbf{s}_i - \mathbf{s}_j)$$

is a normalizing function. The kernel functions used in Bachoc et al. (2020) are

Ball kernel: $f(h) = I(h \le r)$, where $r \ge 0$.

Ring kernel: $f(h) = I(r_{in} < h \le r_{out})$, where $r_{in}, r_{out} \ge 0$ and $r_{in} < r_{out}$.

Gauss kernel: $f(h) = \exp(-0.5(\Phi^{-1}(0.95)h/r)^2)$, where $\Phi^{-1}(0.95)$ is the 95% quantile of a standard normal distribution and $r > 0$.

Above $I(\cdot)$ denotes the indicator function. The use of a ball kernel implies that only those locations that are separated by a maximum distance of r receive a positive weight. The ring kernel assigns positive weights only to those locations that are separated by a minimum of r_{in} and a maximum of r_{out}. The Gauss kernel can be considered as a smooth version of the ball kernel. In our simulation studies presented in Section 3 we use only ball and ring kernels. Notice that a special case of $\mathrm{LCOV}(f)$ is obtained using the ball kernel with $r = 0$. We denote such kernel as f_0. As $F_{n,f_0} = 1$, we have

$$\mathrm{LCOV}(f_0) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left(\left(\mathbf{x}_i - \mathrm{E}(\mathbf{x}_i)\right) \left(\mathbf{x}_i - \mathrm{E}(\mathbf{x}_i)\right)^\top \right),$$

and thus, $\mathrm{LCOV}(f_0)$ corresponds to a regular covariance matrix.

2.3. Location vector and covariance matrix functionals based on signs

In Nordhausen et al. (2015), an unmixing matrix functional $\mathbf{W}$ is found by simultaneously diagonalizing $\mathrm{LCOV}(f_0)$ and $\mathrm{LCOV}(f)$. Naturally, the separation performance depends heavily on the choice of f and, as is well known from the time series context, the use of several covariance matrices in the estimation of the unmixing matrix improves the performance of a BSS method (Miettinen et al., 2016). Bachoc et al. (2020) extended the approach based on two local covariance matrices so that $\mathbf{W}$ is found by jointly diagonalizing $\mathrm{LCOV}(f_0)$ and $\mathrm{LCOV}(f_1), \ldots, \mathrm{LCOV}(f_K)$, where $f_1, \ldots, f_K$ are K different kernels. It is also argued that the most natural choice for kernels would be to choose K non-overlapping ring kernels. We return to simultaneous and joint diagonalization later in this section. As neither the mean vector used for centering nor the $\mathrm{LCOV}(f)$ are robust, the methods based on these quantities are highly sensitive to outliers and inefficient when random fields originate from heavy-tailed distributions. Therefore, we propose robust counterparts for the methods presented in Nordhausen et al. (2015), Bachoc et al. (2020), Muehlmann et al. (2024) in the following.

When selecting robust location vectors and covariance matrix functionals that will be used to define robust unmixing matrices $\mathbf{W}$, we reference the time series context by Ilmonen et al. (2015). First, we replace $\mathrm{LCOV}(f_0)$ and the mean vector that is used for centering with the Hettmansperger-Randles (HR) shape matrix and location vector (Hettmansperger and Randles, 2002), which are known to be affine equivariant. Recall that for the p-variate random fields $\mathbf{x}_i = \mathbf{x}(\mathbf{s}_i)$, $i = 1, \ldots, n$, the HR functionals for the location vector and shape matrix, denoted by $\mathbf{T}$ and $\mathbf{V}$, solve

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left(\mathbf{U}(\mathbf{y}_i)\right) = \mathbf{0} \quad \text{and} \quad \frac{p}{n} \sum_{i=1}^{n} \mathrm{E}\left(\mathbf{U}(\mathbf{y}_i)\mathbf{U}(\mathbf{y}_i)^\top\right) = \mathbf{I}_p, \tag{3}$$

where $\mathbf{U}(\cdot)$ is a spatial sign score defined as $\mathbf{U}(\mathbf{y}) = \mathbf{y}/\|\mathbf{y}\|$, for $\mathbf{y} \neq \mathbf{0}$, and $\mathbf{U}(\mathbf{0}) = \mathbf{0}$ (Möttönen and Oja, 1995), $\mathbf{y}_i = \mathbf{V}^{-\frac{1}{2}}(\mathbf{x}_i - \mathbf{T})$, $i = 1, \ldots, n$, and $\mathbf{V}$ is standardized so that $\mathrm{tr}(\mathbf{V}) = p$ (for example). The resulting location vector $\mathbf{T}$ is known as the transformation–retransformation (TR) spatial median (Chakraborty et al., 1998), and matrix $\mathbf{V}$ is Tyler's shape matrix (Tyler, 1987) with respect to the TR spatial median. For the robustness and efficiency properties of HR estimates under the elliptical model, see Hettmansperger and Randles (2002).

LCOVs based on a kernel f are replaced by the generalized local spatial sign covariance matrices (gLSSCMs) which build upon the generalized spatial sign covariance matrix (gSSCM) proposed by Raymaekers and Rousseeuw (2019) and are defined as follows.

Definition 3. Assume n locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ in $\mathcal{S}$ and write $\mathbf{x}_i = \mathbf{x}(\mathbf{s}_i)$, $i = 1, \ldots, n$, for a p-dimensional multivariate random field. The gLSSCM is then defined as

$$\begin{aligned} &\text{gLSSCM}(f, w) \\ &= \frac{1}{n\sqrt{F_{n,f}}} \sum_{i=1}^{n} \sum_{j=1}^{n} f(\mathbf{s}_i - \mathbf{s}_j) \mathrm{E}\left(w(r_i) w(r_j) \left(\mathbf{x}_i - \mathbf{T}\right) \left(\mathbf{x}_j - \mathbf{T}\right)^{\top} \right), \end{aligned} \tag{4}$$

where $\mathbf{T}$ is a location vector functional, $r_i = \|\mathbf{x}_i - \mathbf{T}\|$ and $w : \mathbb{R} \to \mathbb{R}$ is a radial function.

When f is chosen as the ball kernel with $r = 0$, gLSSCM reduces to gSSCM as defined in Raymaekers and Rousseeuw (2019). Notice that their choice for a location vector functional $\mathbf{T}$ is a k-step least trimmed squares (LTS) estimator, however, we prefer to use the HR location vector. Raymaekers and Rousseeuw (2019) studied the robustness properties of gSSCM under elliptical distributions using influence functions and asymptotic breakdown points. Several suggestions for radial functions (w) are also listed. In our simulation studies presented in Section 3 we use the following functions:

Spatial sign:

$$w(r) = 1/r.$$

Winsor:

$$w(r) = \begin{cases} 1 & \text{if } r \leq Q, \\ Q/r & \text{if } Q \leq r. \end{cases}$$

Quadratic Winsor:

$$w(r) = \begin{cases} 1 & \text{if } r \leq Q, \\ Q^2/r^2 & \text{if } Q \leq r. \end{cases}$$

The cutoff Q is defined as $Q = r_{(h)}$, where $r_{(h)}$ is the hth-order statistic of $\{r_1, \ldots, r_n\}$ and $h = (n + p + 1)/2$.

Before introducing robust unmixing matrix functionals, we state in the following proposition that gLSSCM(f, w) is an orthogonal equivariant functional. This property is required to obtain a genuine unmixing matrix functional (Oja et al., 2006).

Proposition 1. *Let* $\mathbf{T}$ *be any affine (or orthogonal) equivariant location vector functional. Thus,* gLSSCM(f, w) *is an orthogonal equivariant functional such that if* $\mathbf{x}^\star = \mathbf{U}\mathbf{x} + \mathbf{b}$, *where* $\mathbf{x}$ *is a* p*-variate random field,* $\mathbf{U}$ *is an orthogonal* $p \times p$ *matrix and* $\mathbf{b}$ *is a* p*-vector, then for* gLSSCM$^\star(f, w)$ *and* gLSSCM(f, w) *based on* $\mathbf{x}^\star$ *and* $\mathbf{x}$, *respectively, it holds that* gLSSCM$^\star(f, w) = \mathbf{U}$gLSSCM$(f, w)\mathbf{U}^\top$.

The proof is straightforward as kernel and radial functions are invariant with respect to the orthogonal transformation.

2.4. Robust unmixing matrices

In Tong et al. (1990) and Belouchrani et al. (1997), BSS problems are solved in the time series context using the simultaneous and joint diagonalizations of two or more covariance matrices, respectively. Later, similar concepts have been applied in various BSS settings. For recent reviews, see for example Pan et al. (2022), Nordhausen and Ruiz-Gazen (2022). Next, we propose robust unmixing matrix functionals based on the location vector and covariance matrix functionals introduced in the previous section. For convenience, hereinafter, we denote $\mathbf{V}_0$ as the HR shape matrix and $\mathbf{V}_k$ as the gLSSCM(f_k, w), $k = 1, \ldots, K$. As before, $\mathbf{T}$ is the HR location vector functional.

We start by defining a robust unmixing matrix functional that uses the simultaneous diagonalization of two matrices. For that we require an additional assumption to ensure that the resulting unmixing matrix is identifiable.

SBSS3: $z_1(\mathbf{s}), \ldots, z_p(\mathbf{s})$ follow symmetric probability distributions.

Notice that under the SBSS model (1) with assumptions **SBSS1**-**SBSS3**, $\mathbf{T} = \boldsymbol{\mu}$, $\mathbf{V}_0 = \boldsymbol{\Omega}\boldsymbol{\Lambda}_0\boldsymbol{\Omega}^\top$ and $\mathbf{V}_k = \boldsymbol{\Omega}\boldsymbol{\Lambda}_k\boldsymbol{\Omega}^\top$, where $\boldsymbol{\Lambda}_0$ and $\boldsymbol{\Lambda}_k$ are diagonal matrices. Thus, all matrices involved possess the so-called independence property (Nordhausen and Tyler, 2015), i.e. they are diagonal matrices for random fields with symmetrically distributed, independent components. Therefore we suggest the following new functional.

Definition 4. Consider a random field, $\mathbf{x}(\mathbf{s})$, following the SBSS model (1) and satisfying assumption **SBSS3**. The $p \times p$ matrix-variate functional $\mathbf{W}_{2\mathbf{V}}$ is defined as the matrix which simultaneously diagonalizes $\mathbf{V}_0$ and $\mathbf{V}_1$. Thus, $\mathbf{W}_{2\mathbf{V}}$ solves

$$\mathbf{W}_{2\mathbf{V}}\mathbf{V}_0\mathbf{W}_{2\mathbf{V}}^\top = \mathbf{I}_p \quad \text{and} \quad \mathbf{W}_{2\mathbf{V}}\mathbf{V}_1\mathbf{W}_{2\mathbf{V}}^\top = \mathbf{D},$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements in decreasing order.

The resulting latent fields are standardized with respect to $\mathbf{V}_0$ and uncorrelated with respect to $\mathbf{V}_1$. In practice, $\mathbf{W}_{2\mathbf{V}}$ and $\mathbf{D}$ can be found as an eigenvalue-eigenvector solution for $\mathbf{V}_0^{-1}\mathbf{V}_1$.

Proposition 2. *The $p \times p$ matrix $\mathbf{W}_{2\mathbf{V}}$ defined in Definition 4 is a genuine unmixing matrix functional in the sense of Definition 2 if and only if the diagonal elements of $\mathbf{\Lambda}_1$ are distinct.*

As $\mathbf{V}_0$ is affine equivariant and $\mathbf{V}_1$ is orthogonal equivariant and both functionals possess the independence property, the proof follows from Theorem 2 in Oja et al. (2006).

Remark 1. We consider the unmixing matrix $\mathbf{W}_{2\mathbf{V}}$ more robust than the original one introduced in Nordhausen et al. (2015) as all individual parts of $\mathbf{W}_{2\mathbf{V}}$ are of a robust nature. This will be evaluated in more detail in a simulations study later.

As the performance of $\mathbf{W}_{2\mathbf{V}}$ naturally depends on the chosen kernel (f_1), it is often better to perform the joint diagonalization of several local covariance matrices. This is formalized in the following.

Definition 5. Consider a random field, $\mathbf{x}(\mathbf{s})$, following the SBSS model (1) and satisfying assumption **SBSS3**. The $p \times p$ matrix variate functional $\mathbf{W}_{K\mathbf{V}}$ is the $p \times p$ matrix that maximizes

$$\sum_{k=1}^{K} \| \mathrm{diag}\left(\mathbf{W}_{K\mathbf{V}} \mathbf{V}_k \mathbf{W}_{K\mathbf{V}}^{\top}\right) \|_F^2$$

under the constraint, $\mathbf{W}_{K\mathbf{V}} \mathbf{V}_0 \mathbf{W}_{K\mathbf{V}}^{\top} = \mathbf{I}_p$. Here $\| \cdot \|_F$ denotes the Frobenius norm and $diag(\mathbf{A})$ is a diagonal matrix with the diagonal elements as in $\mathbf{A}$.

Proposition 3. *The $p \times p$ matrix $\mathbf{W}_{K\mathbf{V}}$ defined in Definition 5 is a genuine unmixing matrix functional in the sense of Definition 2 if and only if for each $i, j = 1, \ldots, p$, there exists a $k \in \{1, \ldots, K\}$ such that $[\mathbf{\Lambda}_k]_{ii} \neq [\mathbf{\Lambda}_k]_{jj}$.*

The proof of this proposition follows the same outline given in the time series context in Matilainen et al. (2015) and in the spatial case in Bachoc et al. (2020), together with the additional requirement of symmetry required for the independence property. To solve the maximization problem given in Definition 5, many algorithms exist, as discussed in Illner et al. (2015). We will use the algorithm based on Givens rotations, as outlined in Clarkson (1988).

Remark 2. We consider the unmixing matrix $\mathbf{W}_{K\mathbf{V}}$ more robust than the ones introduced in Bachoc et al. (2020) as all individual parts of $\mathbf{W}_{K\mathbf{V}}$ are of a robust nature. This will again be evaluated in more detail in a simulations study later on.

Finally, notice that we denote the unmixing matrix estimate corresponding to $\mathbf{W}_{2\mathbf{V}}$ and $\mathbf{W}_{K\mathbf{V}}$ as $\hat{\mathbf{W}}_{2\mathbf{V}}$ and $\hat{\mathbf{W}}_{K\mathbf{V}}$, respectively. The estimates are obtained by using the sample counterparts, $\hat{\mathbf{T}}$ and $\hat{\mathbf{V}}_0, \ldots, \hat{\mathbf{V}}_K$, in the unmixing matrix estimation. In the following section, we discuss the efficiency and robustness properties of the two new proposals based on simulation studies.

3. Simulation studies

Simulation studies are performed to compare the proposed robust SBSS methods with the non-robust SBSS methods of Nordhausen et al. (2015), Bachoc et al. (2020) and to verify the performance of the robust SBSS method in the presence of outliers and when the latent fields are non-Gaussian. All SBSS methods are implemented in the R package SpatialBSS (Muehlmann et al., 2022c), and the simulation study can be reproduced using R 4.2.1 (R Core Team, 2022) together with the RandomFields (Schlather et al., 2022), JADE (Miettinen et al., 2017), sp (Pebesma and Bivand, 2005), rgdal (Bivand et al., 2022) and ggplot2 (Wickham, 2016) packages. Note that the package RandomFields is no longer in CRAN but can be used in R 4.2.1.

3.1. Generation of the latent random fields

In the simulation studies the sample locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ are uniformly generated points on the map of Finland, which is presented in Fig. 1 together with the sampled points, where $n = 1000$. The latent random fields are generated using the Matérn correlation function defined by

$$K(h; \nu, \phi) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{h}{\phi}\right)^{\nu} K_{\nu}\left(\frac{h}{\phi}\right), \tag{5}$$

where $h = \|\mathbf{s}_i - \mathbf{s}_j\|$ is the Euclidean distance between two locations $\mathbf{s}_i$ and $\mathbf{s}_j$, $\nu > 0$ is a scale parameter, $\phi > 0$ is a range parameter, Γ is the gamma function, and K_{ν} is the modified Bessel function of the second kind with shape parameter ν.

To compare the methods in the presence of outliers, the Gaussian random fields are contaminated to have either global or local outliers. For stationary random fields, global outliers are extreme values in magnitude, whereas local outliers are not extreme in magnitude but do not follow the local spatial correlation structure. Two types of global outliers are examined, uniformly occurring outliers and aggregated outliers. The uniform globally contaminated Gaussian random fields $\mathbf{x}^{\alpha l}(\mathbf{s}_i)$ are generated as

$$\mathbf{x}^{\alpha l}(\mathbf{s}_i) = (1 - \gamma_i^{\alpha})\mathbf{x}(\mathbf{s}_i) + \gamma_i^{\alpha}(\mathbf{x}(\mathbf{s}_i) + l\mathbf{1}_p), \tag{6}$$

where $\mathbf{x}(\mathbf{s}_i), i = 1, \ldots, n$, is a p-variate Gaussian random field, γ_i^{α} are independently and identically distributed Bernoulli variables with $P(\gamma_i^{\alpha} = 1) = \alpha$, $l \in \mathbb{R}$, and $\mathbf{1}_p$ is a p-vector of ones. The contamination rate, i.e. $\alpha \in [0, 1/2)$, defines the proportion of observations to be contaminated.

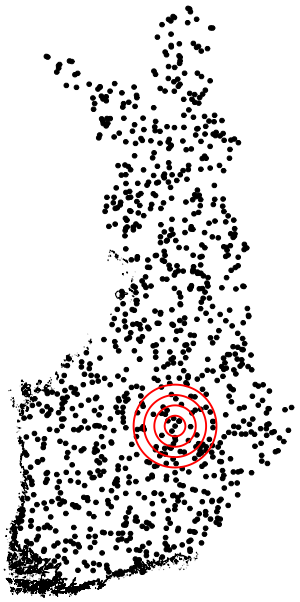


Fig. 1. Map of Finland with $n = 1000$ uniformly sampled points and ring kernels for radii 20, 40, 60 and 80 km.

The global aggregated outliers are generated to m clusters using the ideas of Kerry and Oliver (2007). The aggregated outliers are generated by randomly selecting m sample locations as cluster center points. The observations are sorted based on the distance to any of the center points. The proportion α of the observations are selected to be contaminated based on the lowest distance. The aggregated globally contaminated Gaussian random field $\tilde{\mathbf{x}}^{al}$ is then generated as

$$\tilde{\mathbf{x}}^{al}(\mathbf{s}_i) = (1 - \delta_i)\mathbf{x}(\mathbf{s}_i) + \delta_i(\mathbf{x}(\mathbf{s}_i) + l\mathbf{1}_p), \tag{7}$$

where $\delta_i = 1$ if the ith observation is selected to be contaminated, otherwise $\delta_i = 0$.

For generating local outliers, we need to define a neighborhood $\mathcal{N}_i$ for each sample location $\mathbf{s}_i$, $i = 1, \ldots, n$. One approach is to consider a neighborhood that contains a fixed number of k observations. Thus, $\mathcal{N}_i$ contains $k - 1$ nearest points $\mathbf{s}_i$. Another approach is to construct the neighborhood so that it contains all points within a range r. Thus, $\mathcal{N}_i$ contains all points $\mathbf{s}_j$ for which $\|\mathbf{s}_i - \mathbf{s}_j\| < r$ (Ernst and Haesbroeck, 2017). In the following simulations the latter approach is used with $r = 10$ km. The outliers are generated following the procedure of Ernst and Haesbroeck (2017) and reviewed next.

1. Sort the observations $\mathbf{x}_i = \mathbf{x}_i(\mathbf{s}_i)$, $i = 1, \ldots, n$, from the highest to the lowest by the first principal component score (PC-1) according to Harris et al. (2014).
2. Set k as $\alpha n/2$ rounded to the nearest integer, and select the set of local outlier points S_{out} by finding k observations with the highest PC-1 values and k observations with the lowest PC-1 values on the condition that for all $\mathbf{s}_i, \mathbf{s}_j \in S_{out}$ it holds that $\mathcal{N}_i \neq \mathcal{N}_j$.
3. Generate the local outliers by swapping observations $\mathbf{x}_i$ and $\mathbf{x}_{2k+1-i}$, $i = 1, \ldots, k$, where the observations $\mathbf{x}_1, \ldots, \mathbf{x}_{2k}$ are sorted based on their PC-1 values from the smallest to the largest.

Fig. 2 illustrates a three-variate Gaussian field with and without local and global outliers with a contamination rate of $\alpha = 0.05$.

Finally, to compare the performance of the methods when the distribution of the latent fields is heavy-tailed, we generate Student's t_υ-distributed random fields, where υ is the degree of freedom. We denote the univariate t_υ-distributed random fields by $x^{t_\upsilon}(\mathbf{s}_i)$ and compute them as in Muehlmann et al. (2021b) by generating $\upsilon + 1$ zero-centered Gaussian random fields $x_j(\mathbf{s}_i)$ with unit variance and applying

$$x^{t_\upsilon}(\mathbf{s}_i) = \frac{\sqrt{\upsilon}\; x_1(\mathbf{s}_i)}{\sqrt{\sum_{j=2}^{\upsilon+1} x_j(\mathbf{s}_i)^2}},$$

for $i = 1, \ldots, n$.

3.2. Finite sample efficiencies

In the simulations, we use $n = 1000$ and $n = 2000$ as the number of locations. For both setups the locations are generated once and kept constant throughout the simulations. The p-variate latent random fields, with dimension $p = 3$, are freshly sampled in each simulation iteration.

Fig. 2. Three-variate Gaussian field before contamination (first column), after adding 5% of local outliers (second column), after adding 5% of global outliers (third column) and after adding two clusters of aggregated outliers (fourth column). The local outliers are marked with black borders.

Two different Matérn covariance models are considered. Parameters ν_i and ϕ_i for latent field i defining the Matérn correlation structure are selected according to Bachoc et al. (2020). The range parameters ϕ_i are scaled to match the dimensions of the map of Finland. Therefore, model 1 uses $(\nu_1, \phi_1) = (2, 20)$, $(\nu_2, \phi_2) = (1, 20)$ and $(\nu_3, \phi_3) = (0.25, 20)$, and model 2 uses $(\nu_1, \phi_1) = (6, 24)$, $(\nu_2, \phi_2) = (1, 30)$ and $(\nu_3, \phi_3) = (0.25, 20)$. The chosen Matérn correlation functions are presented in Fig. 3.

In every simulation setting the latent fields are generated independently 2000 times for each covariance model. The unmixing matrix estimates $\hat{\mathbf{W}}$ to be compared are the SBSS estimates based on local covariance matrices (LCOV(f)), as defined in Nordhausen et al. (2015), Bachoc et al. (2020), and the robust estimates utilizing spatial sign (RSBSS-S), Winsor (RSBSS-W) and quadratic Winsor (RSBSS-Q) as the radial functions. For each method, the effect of the number of kernels K is examined by comparing the performances for $K = 1, \ldots, 4$. The parameters of the ring kernels are given in Table 1, where $K = 1$ corresponds to a ball kernel with radii $r = 20$. The ring radius are presented in Fig. 1, which has the map of Finland, and in Fig. 3, which has the Matérn correlation functions. As all unmixing matrix estimates are affine equivariant, we choose $\boldsymbol{\Omega} = \mathbf{I}_p$ as the true mixing matrix.

Table 1
Parameters for the ring kernels with varying number of rings $K = 1, \ldots, 4$. The parameter vector $\mathbf{r}^{(i)} = (r_{in}^{(i)}, r_{out}^{(i)})$ defines the parameters for the ith ring.

K	$\mathbf{r}^{(1)}$	$\mathbf{r}^{(2)}$	$\mathbf{r}^{(3)}$	$\mathbf{r}^{(4)}$
1	(0, 20)			
2	(0, 20)	(20, 40)		
3	(0, 20)	(20, 40)	(40, 60)	
4	(0, 20)	(20, 40)	(40, 60)	(60, 80)

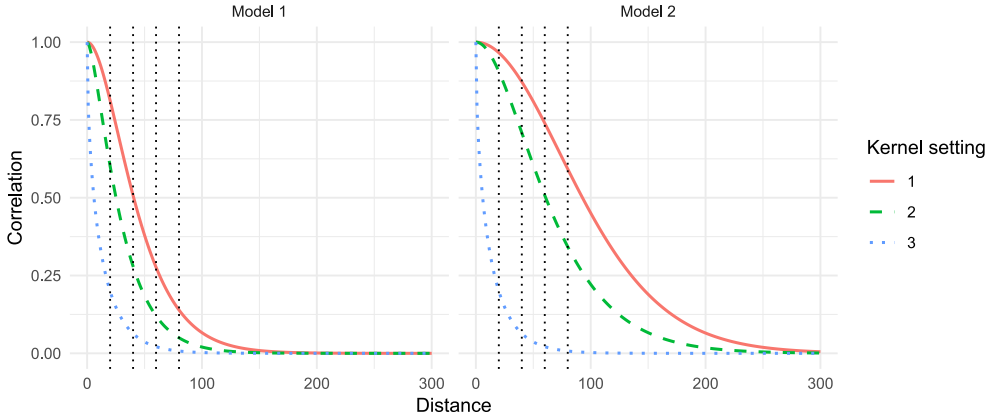


Fig. 3. Matérn covariance functions for the first latent field (solid red line), the second latent field (dashed green line) and the third latent field (dotted blue line) used in the simulation studies. Parameters (ν, ϕ) are $(2, 20)$, $(1, 20)$, $(0.25, 20)$ in model 1 and $(6, 24)$, $(1, 30)$ and $(0.25, 20)$ in model 2. The dotted vertical lines at distances $20, 40, 60$ and 80 km illustrate the boundaries of ring kernels.

There are many performance measures to compare different BSS approaches in simulation studies as for example reviewed in Nordhausen et al. (2011). Some of them require that all algorithms scale the latent components using the same scale measure which is not the case in our comparison. A measure not affected by the scaling is the minimum distance (MD) index (Ilmonen et al., 2010; Lietzen et al., 2020), which is a function of the gain matrix $\hat{\mathbf{G}} = \boldsymbol{\Omega}\hat{\mathbf{W}}$.

The MD index is calculated as

$$\mathrm{MD}(\hat{\mathbf{G}}) = \frac{1}{\sqrt{p-1}} \inf_{\mathbf{C} \in C} \|\mathbf{C}\hat{\mathbf{G}} - \mathbf{I}_p\|, \tag{8}$$

where C is a set of all the matrices of the form $\mathbf{C} = \mathbf{PJS}$, where $\mathbf{P}$, $\mathbf{J}$ and $\mathbf{S}$ are transformation matrices as defined in Section 2. The MD index takes values in the range of 0 to 1, where 0 corresponds to the optimal value.

To simplify the interpretation of the MD indices in our simulation studies easier, the average MD index of a random guess in the case of $p = 3$ is calculated. Based on 10000 random matrices generated by filling their elements independently from a standard normal distribution, an average random MD index of 0.78 is obtained. Therefore, if $\mathrm{MD}(\hat{\mathbf{G}}) > 0.78$, the result is no better than a random guess.

In the first simulation study, the performances of the methods are compared when the latent fields are Gaussian and there are no outliers present. The results presented in Fig. 4 imply that the robust methods perform very similarly to SBSS. For model 1 and when $n = 1000$, the best performing methods use two rings, whereas when $n = 2000$ the methods that use the ball kernel perform slightly better than the others. For model 2, the methods that use multiple rings display similar performances. The methods based on two or three rings deliver the best results. Notice that although the data here are uncontaminated Gaussian, the classical unmixing matrices do not necessarily perform better than the new proposals. The performance depends on how well the eigenvalues of the local scatter functionals are separated. This will in general differ here as the matrices considered are not necessarily proportional to each other in a similar fashion as the traditional scatter matrices are in an iid framework. In fact, the RSBSS-S method slightly outperforms the classical SBSS method, when the number of kernels is not the optimal one. This implies that even in the uncontaminated case, the RSBSS-S method is less sensitive to nonoptimal choice of kernels than SBSS. Similar behavior is present for RSBSS-W and RSBSS-Q methods as well under model 1, when the number of kernels is 4. For other kernel settings and under model 2, the differences between RSBSS-W and RSBSS-Q are negligible as compared to SBSS.

The second simulation study compares the methods in presence of uniformly generated global outliers. The uniform globally contaminated Gaussian random fields are generated using the contamination rate of $\alpha = 0.05$. The results presented in Fig. 5 show that the performance of SBSS in the presence of uniform global outliers is as bad as a random guess. When compared with the uncontaminated case, the performances of the robust methods drop only marginally when more than one ring kernel is used, but more noticeably when the ball kernel is used. RSBSS-Q performs slightly better than RSBSS-S and RSBSS-W. The best results are obtained when at least two rings are used.

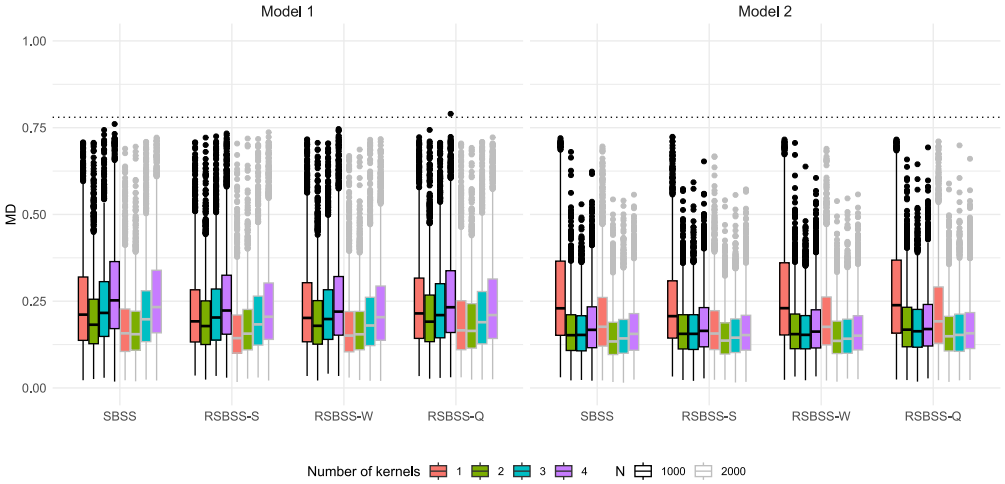


Fig. 4. MD indices of 2000 trials for SBSS, RSBSS-S, RSBSS-W and RSBSS-Q for the Gaussian random fields under model 1 (left) and under model 2 (right) when $n = 1000$ (black borders) and $n = 2000$ (gray borders). The dotted line represents the average MD index of a randomly generated gain matrix.

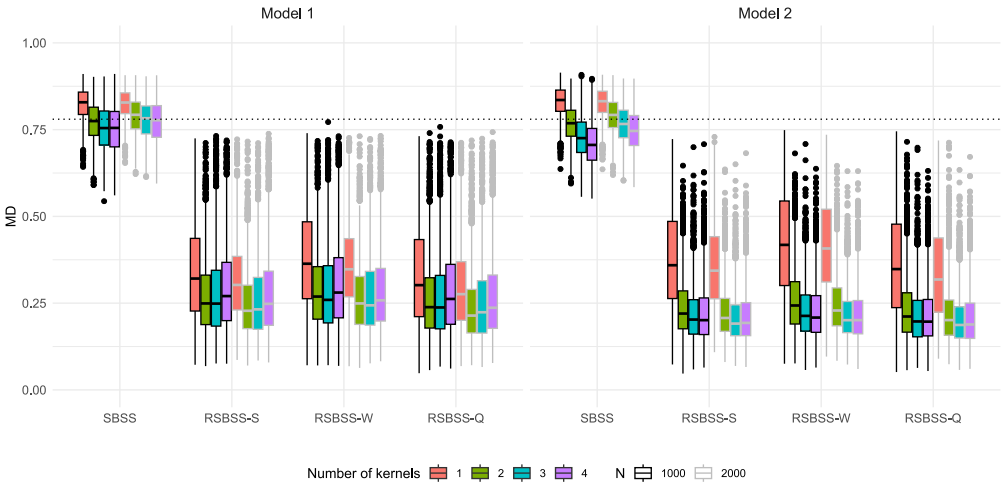


Fig. 5. MD indices of 2000 trials for SBSS, RSBSS-S, RSBSS-W and RSBSS-Q for Gaussian random fields with 5% uniformly generated global outliers under model 1 (left) and under model 2 (right) when $n = 1000$ (black borders) and $n = 2000$ (gray borders). The dotted line represents the average MD index of a randomly generated gain matrix.

In the third simulation study, the methods are compared in the presence of aggregated global outliers. The aggregated globally contaminated Gaussian random fields are generated using the number of aggregated clusters $m = 2$ and the contamination rate of $\alpha = 0.05$. The results presented in Fig. 6. In the presence of aggregated global outliers, the performance of SBSS is nearly as bad as in the presence of uniform global outliers. The performances of the robust methods drop only marginally overall as compared to the uncontaminated setup. For RSBSS-Q, the performance drops slightly more when the ball kernel is used. For RSBSS-Q, the best results are obtained with two ring kernels under model 1 and with four ring kernels under model 2. The results of RSBSS-S and RSBSS-W are very similar to the results in uncontaminated case, meaning that the ball kernel is the best option under model 1 when $n = 1000$ and two ring kernels is the best option in other scenarios. For uniform global outliers, the performance drop of the robust methods is noticeable when the ball kernel is used, whereas for aggregated global outliers, the robust methods are not as sensitive to the choice of the number of kernels.

In the fourth simulation study, locally contaminated Gaussian fields are generated using the contamination rate of $\alpha = 0.05$. The results are presented in Fig. 7. When introducing local outliers to Gaussian fields, the differences between performances of the non-robust and robust methods are not as significant as those in the case with global outliers, although SBSS still performs worse than the robust methods. In this setup, the RSBSS-S and RSBSS-Q methods using three or four rings perform the best overall, although the difference is not significant when compared with the performance of RSBSS-W.

In the fifth simulation study, t_3-distributed random fields are generated to compare the performances of the methods when the latent fields are heavy-tailed. The results in Figs. 8 show that the performance of SBSS decreases, particularly when multiple rings

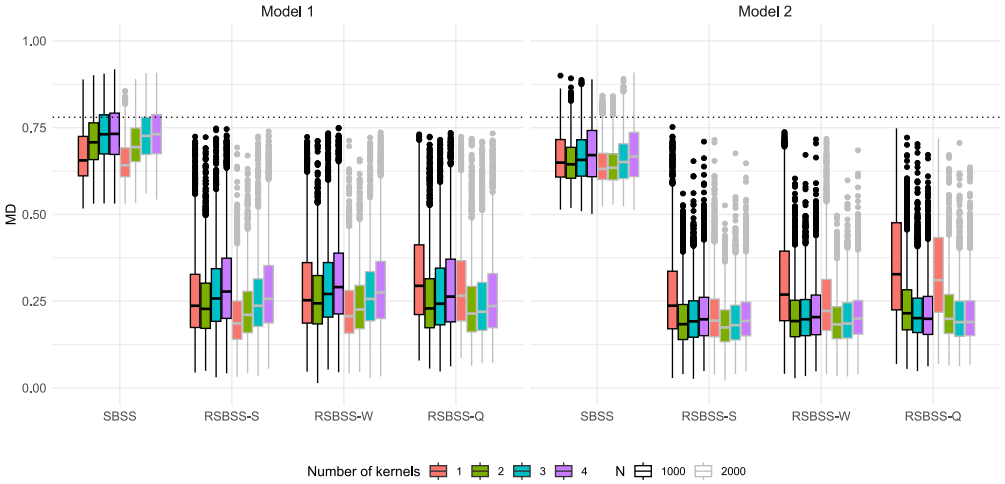


Fig. 6. MD indices of 2000 trials for SBSS, RSBSS-S, RSBSS-W and RSBSS-Q for Gaussian random fields with 5% aggregated global outliers under model 1 (left) and under model 2 (right) when $n = 1000$ (black borders) and $n = 2000$ (gray borders). The dotted line represents the average MD index of a randomly generated gain matrix.

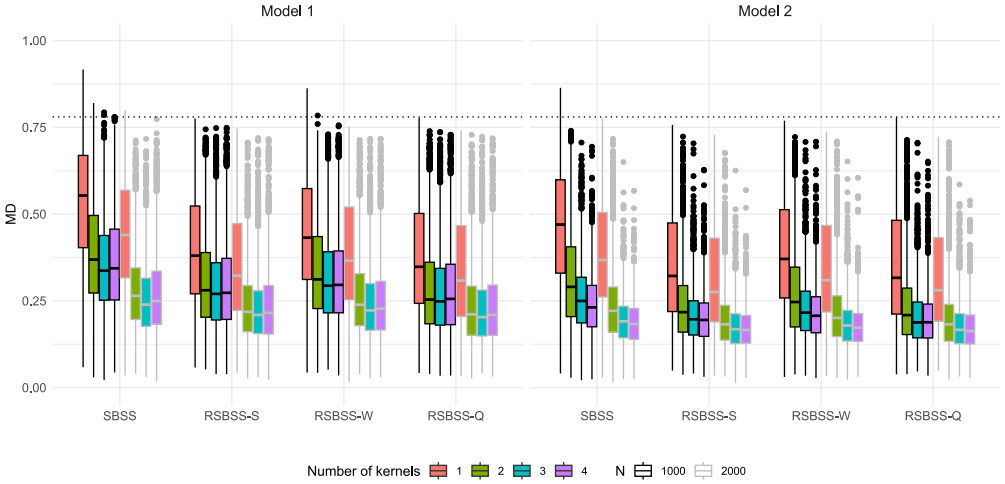


Fig. 7. MD indices of 2000 trials for SBSS, RSBSS-S, RSBSS-W and RSBSS-Q for the Gaussian random fields with 5% local outliers under model 1 (left) and under model 2 (right) when $n = 1000$ (black borders) and $n = 2000$ (gray borders). The dotted line represents the average MD index of a randomly generated gain matrix.

are used, whereas the performances of the robust methods remain stable. For the robust methods, the best performing methods use the ball kernel for model 1 and two rings for model 2. RSBSS-S and RSBSS-W slightly outperform RSBSS-Q in this setup.

In general, all robust SBSS methods display similar performances. They perform as well as SBSS when the random fields are Gaussian and outperform SBSS in other simulation setups, particularly in presence of uniform or aggregated global outliers. The robust methods become less sensitive to uniform global outliers and local outliers when multiple rings are used instead of the ball kernel. This is because to perform well, the eigenvalues of $\hat{\mathbf{V}}_0^{-1}\hat{\mathbf{V}}_1$, where $\hat{\mathbf{V}}_0$ and $\hat{\mathbf{V}}_1$ are sample counterparts of matrices given in Definition 4, have to be distinct. With finite samples, this may not be true, thus the use of several matrices $\hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_k$ (that is, several ring kernels) often gives better performance. In presence of aggregated global outliers, the results are more similar to the uncontaminated setup for all number of kernels. The simulations were repeated by sampling the locations independently in each trial, and the results remained the same. In conclusion, robust methods are generally safer options than SBSS; in addition to their performance being similar to that of SBSS in the Gaussian setting, they still perform consistently better in presence of outliers or non-Gaussian data. To guarantee high efficiency under contaminated data, robust methods should be used with multiple ring kernels rather than with the ball kernel to avoid a decrease in performance. The RSBSS-S and RSBSS-Q methods appear to tolerate uniform global outliers and local outliers slightly better than RSBSS-W, which might be because of the fact that the spatial sign and the quadratic Winsor radial functions downweight the outliers more heavily than the Winsor radial function. However, for aggregated

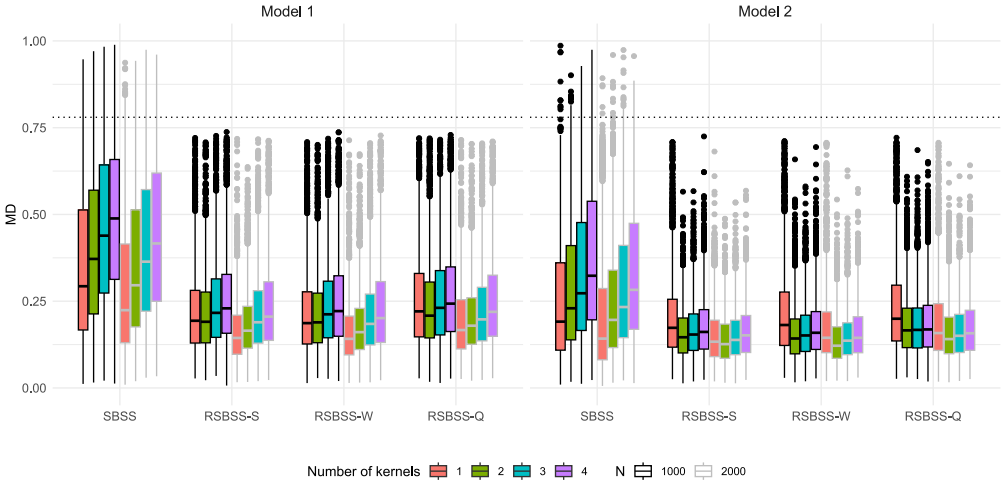


Fig. 8. MD indices of 2000 trials for SBSS, RSBSS-S, RSBSS-W and RSBSS-Q for t_3-distributed random fields under model 1 (left) and under model 2 (right) when $n = 1000$ (black borders) and $n = 2000$ (gray borders). The dotted line represents the average MD index of a randomly generated gain matrix.

global outliers and t_3-distributed data, a heavy downweighting might not be beneficial as RSBSS-Q is slightly outperformed by RSBSS-S and RSBSS-W, although the differences are small.

3.3. Bias and maximum bias curves

In an iid setting, the robustness properties are often evaluated using influence functions and bias curves, however, such concepts are difficult to extend even to a time series framework as pointed out for example by Maronna et al. (2019, Chapter 8.11). This is then even more difficult in a spatial data framework where there is no natural order in the observations and where there are two competing asymptotic frameworks (in-fill and increasing domain asymptotics) available.

To study the robustness properties of the different methods discussed in this paper, we use ideas of Rousseeuw (2006), Muler et al. (2009) to get a notion of bias and maximum bias curves of the estimated unmixing matrices in the case of contaminated random fields. We use as the bias of an unmixing matrix estimate $\hat{\mathbf{W}}$ in case of contaminated Gaussian random fields $\mathbf{x}^{\alpha l}(\mathbf{s}_i)$, as defined in (6) measured at a specific set of n locations $\mathbf{S}$, the MD index as follows

$$\text{Bias}(\hat{\mathbf{W}}, \boldsymbol{\Omega}, \alpha, \mathbf{S}) = \text{MD}(\hat{\mathbf{G}}), \tag{9}$$

and the maximum bias is measured by

$$\text{MB}(\hat{\mathbf{W}}, \boldsymbol{\Omega}, \alpha, \mathbf{S}) = \sup_{l \in L} \text{MD}(\hat{\mathbf{G}}). \tag{10}$$

Thus, bias and maximum bias are conditional on a specific map with a specific sample size and maximum bias might change for different configurations of the map or sample sizes. Due to finite sample, the bias is not zero even in the uncontaminated case. Nevertheless, we think this approach gives a meaningful comparison and ordering of the SBSS methods under consideration in the presence of uniform global outliers.

In our concrete case we sampled $n = 1000$ locations and then repeatedly generated a three-variate uniform globally contaminated Gaussian random field under model 1 using a contamination rate of $\alpha = 0.05$. Based on our empirical investigations, the bias curves reach their maximum either when $l \in (0, 100]$, or when $l \to \infty$. Therefore, the maximum bias for varying α is in practice approximated using $l = \{1, 2, \ldots, 100, 1000, 2000\}$. The bias curves presented in Fig. 9 show that the bias of SBSS increases significantly, when $l > 1$. The bias increase is fastest for the ball kernel, but decelerates when more kernels are used. The bias of the robust methods varies before $l = 3$, and remains stable when $l > 3$. For all robust methods, the biases are higher when the ball kernel is used compared with the biases obtained when multiple ring kernels are used. RSBSS-S and RSBSS-W have very similar biases, and the lowest bias is obtained when RSBSS-Q is used with multiple ring kernels.

The maximum bias curves presented in Fig. 10 show that SBSS breaks down instantly when at least one large outlier is present in the data. For the robust methods, the maximum bias increases almost linearly when the fraction of outliers increases from 5% to 25%. When the fraction of outliers is more than 22%, all maximum bias curves stay above the limit of a random guess. When the contamination rate $\alpha < 0.17$, RSBSS-Q has the lowest maximum bias, whereas RSBSS-W has the highest one. The maximum bias of RSBSS-S is in between these two. When $\alpha > 0.17$, all of the robust methods exhibit similar performances. For all the robust methods, the maximum bias is larger when the ball kernel is used, and the results are very similar when multiple ring kernels are exploited.

Based on the results above we consider RSBSS-Q with several kernels to be the best performing method. The use of several kernels appears, as already indicated in many other BSS frameworks, to be a safe choice. All three robust methods appear quite

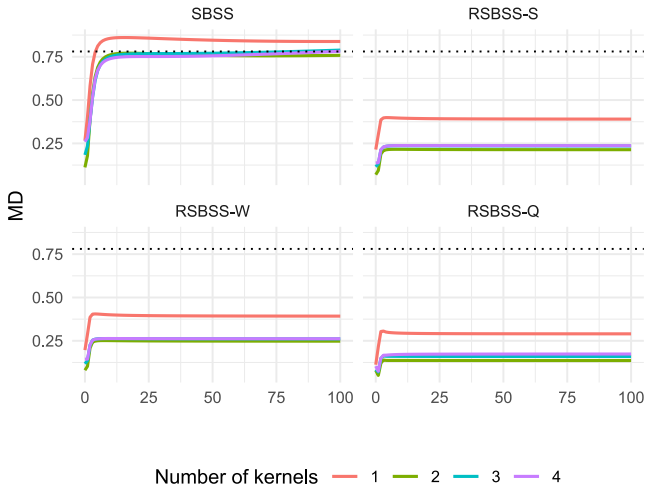


Fig. 9. Bias curves of SBSS, RSBSS-S, RSBSS-W and RSBSS-Q for the globally contaminated Gaussian field $\mathbf{x}^{\alpha l}(\mathbf{s}_i) = (1-\gamma_i^{\alpha})\mathbf{x}(\mathbf{s}_i) + \gamma_i^{\alpha}(\mathbf{x}(s_i) + l\mathbf{1}_p)$, where the contamination rate is $\alpha = 0.05$ and l varies in the x-axis. The dotted black line represents the average MD index of the randomly generated mixing matrix.

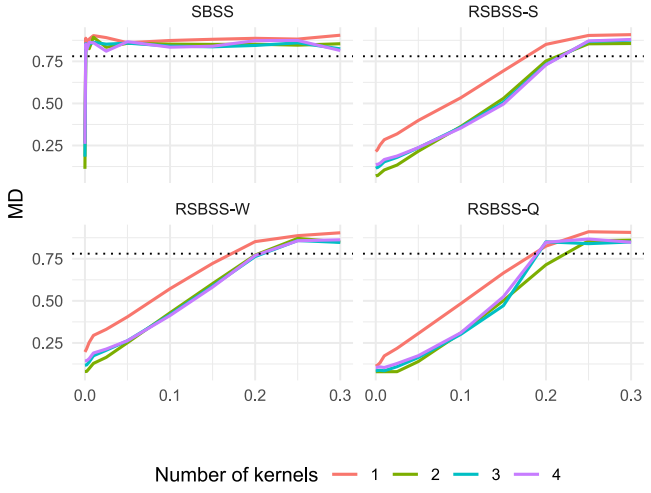


Fig. 10. Maximum bias curves of RSBSS-S, RSBSS-W and RSBSS-Q for the globally contaminated Gaussian field $\mathbf{x}^{\alpha l}(\mathbf{s}_i) = (1-\gamma_i^{\alpha})\mathbf{x}(\mathbf{s}_i) + \gamma_i^{\alpha}(\mathbf{x}(s_i) + l\mathbf{1}_p)$, where α varies in the x-axis. The dotted black line represents the average MD index of the randomly generated mixing matrix.

similar in terms of performance. They do not lose much in performance in the Gaussian framework, however, they are clearly better in the other settings considered. Out of the robust methods, we prefer RSBSS-Q as it appears, performance-wise, to be in between RSBSS-S and RSBSS-W. However, compared with RSBSS-S, it has the advantage of producing weights for observations, which can be exploited for (global) outlier identification. Based on the simulations and the bias curves, RSBSS-Q also appears to have the best tolerance against the global outliers. In the presence of local outliers the robust methods perform only slightly better than their non-robust counterparts, which should be investigated in future research.

4. A real data example

To illustrate the introduced methods we consider a geochemical dataset that has already been considered in the context of outlier detection by Filzmoser et al. (2005). Specifically, we use a dataset derived from the Kola geochemical mapping project (Reimann et al., 2008) that comprises 617 soil samples from the organic layer (O-horizon) taken at different locations alongside the Kola peninsula covering Norway, Finland and Russia. This dataset is available in the R package StatDA (Filzmoser, 2020). The left panel of Fig. 11 shows all considered sample locations. For each sample, the concentrations of the chemical elements As, Cd, Co, Cu, Mg, Pb and Zn are measured. Co, Cu, As, Cd and Pb can be considered as contaminants originating from smelters whereas Mg and Zn are not emitted by smelters; rather, Mg is influenced by marine aerosols from the Arctic coast (Reimann et al., 2000; Filzmoser et al., 2005). Thus, these effects might lead to unusual observations that can be treated appropriately by the introduced robust methods.

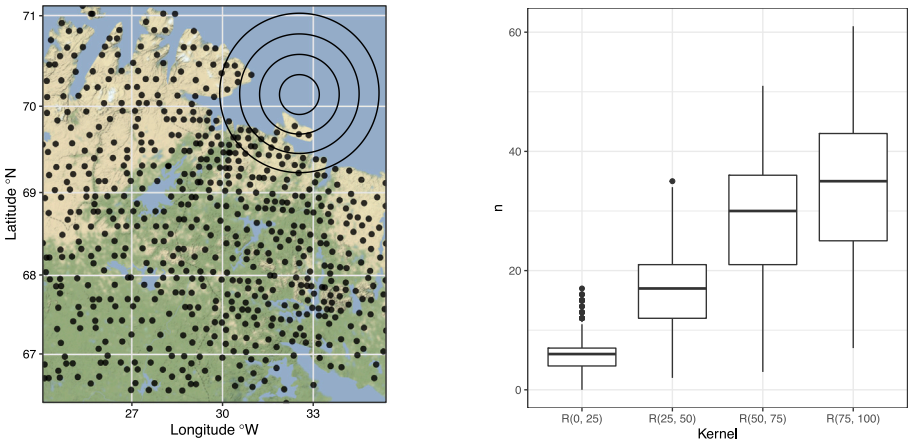


Fig. 11. Map of sample locations and rings with radii of 25, 50, 75 and 100 km (in UTM zone 35) depicting the kernel choices (left panel). Boxplots of neighboring sample locations captured by the kernels for each sample location (right panel).

Table 2
Pseudoeigenvalues of the (robust) local covariance matrices (normalized to unit sum) for the four used spatial BSS methods and all six latent components.

	IC1	IC2	IC3	IC4	IC5	IC6
SBSS	0.5499	0.4004	0.0388	0.0078	0.0030	0.0001
RSBSS-S	0.6530	0.3017	0.0350	0.0078	0.0019	0.0007
RSBSS-W	0.6144	0.3385	0.0356	0.0089	0.0020	0.0005
RSBSS-Q	0.6478	0.3108	0.0296	0.0090	0.0015	0.0014

As the dataset consists of relative concentration measurements we treat it as compositional data and use the log-ratio methodology to account for the relativeness. Specifically, we transform the data into centered log-ratio coordinates (clr). With clr coordinates, the interpretations of the loadings are straightforward; however, they are collinear, which prevents the application of BSS methods. To solve this issue, one conventionally transforms the data further into isometric log-ratio coordinates (ilr). Interpreting ilr coordinates is more involved but they are of full rank. The transformation from clr into ilr coordinates is achieved by an orthogonal transformation $\mathbf{U}$ (denoted as the contrast matrix), which reduces the dimension of the clr coordinates by one. With ilr coordinates, BSS methods can be applied unhindered to estimate an unmixing matrix $\hat{\mathbf{W}}$; moreover, due to the affine equivariance property of the introduced methods, this orthogonal transformation does not change the estimated latent components. To use the simple interpretation in terms of clr coordinates, we combine the contrast and estimated unmixing matrix with the loadings matrix by $\mathbf{U}\hat{\mathbf{W}}$. For the present dataset, the dimension in ilr coordinates equals $p = 6$ and the contrast matrix is formed by using pivot ilr coordinates. Details on the employed coordinate systems and the use of the log-ratio methodology in geochemistry can be found in Filzmoser et al. (2018), the use of the log-ratio methodology in the context of (S)BSS is discussed in Muehlmann et al. (2021a), Nordhausen et al. (2021, 2015) in detail.

To apply the introduced SBSS methods, suitable kernel functions need to be selected. Thus far, no concrete theoretical guidelines have been reported; hence, the kernel functions are best chosen by domain experts. However, the unavailability of domain knowledge can result in vague practical guidelines. From a theoretical viewpoint, every kernel choice should deliver the same latent components if the underlying statistical model holds and the spatial covariance functions are different for the chosen kernels as illustrated in Fig. 3 with different boundaries for the ring kernel functions. Based on these two considerations, a meaningful kernel setting might be found by applying the methods with many different kernel settings and choosing the ones with consistently stable results. For the present dataset, a characteristic spatial variation length of 50 km may be employed as this was found to be a suitable in the context of SBSS on a similar dataset derived from the Kola project by Nordhausen et al. (2015). Moreover, the simulations above and the outline in Bachoc et al. (2020) hint that using more kernels is a safe choice. Thus, next we compare the results of SBSS with those of the three introduced RSBSS methods, all with four ring kernels using radii of 25, 50, 75 and 100 km (in UTM zone 35). The circles in the left panel of Fig. 11 depict these choices, and the boxplots on the right panel show the number of neighboring sample locations captured by the kernel choices for each sample location.

For the RSBSS-W and the RSBSS-Q methods the radial function values can be depicted on a map (Fig. 12 upper panels). In both maps two clusters of low values in the heavy-metal-industry regions can be observed and low radial function values are assigned to the Norway coast region. These low values might originate from the contamination from the smelters and the marine effect. The boxplot (Fig. 12 lower panel) reflects the nature of the radial functions: higher values are observed in the RSBSS-Q method due to the square and for both methods roughly half of the observations are downweighted. The most downweighted observations are mainly the observations with high occurrences of As, Cu and Co, or the observations with a high occurrence of Pb. The observations with high As, Cu and Co are present in clusters in the heavy metal industry regions as aggregated outliers, and are not outliers

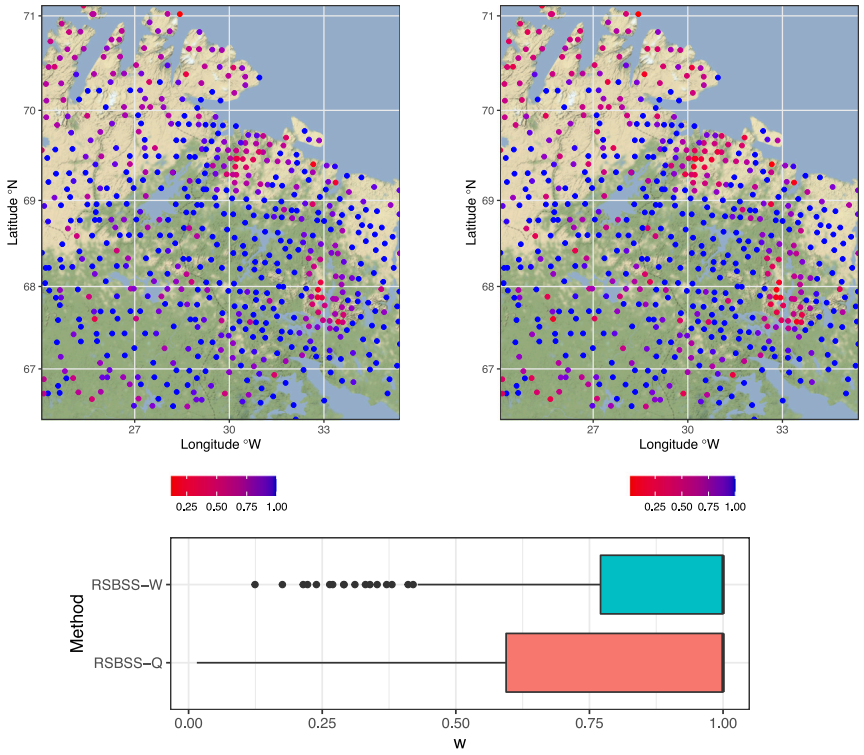


Fig. 12. Maps of the radial function values for the Winsor function (RSBSS-W) (upper left) and the quadratic Winsor function (RSBSS-Q) (upper right). Boxplots of the values depicted on the maps (bottom).

Table 3
Loadings matrices from clr to ICs for SBSS, RSBSS-S, RSBSS-W and RSBSS-Q (in order from top to bottom).

	clr(As)	clr(Cd)	clr(Co)	clr(Cu)	clr(Mg)	clr(Pb)	clr(Zn)
IC1	1.02	−0.11	0.01	−1.32	0.28	0.16	−0.04
IC2	−0.18	0.29	0.52	−0.26	−1.58	0.86	0.35
IC3	2.75	−0.86	−0.48	−0.17	−0.08	−0.91	−0.25
IC4	−0.24	0.60	−1.53	1.15	0.36	1.09	−1.43
IC5	0.04	−1.65	−0.50	0.38	−0.42	0.19	1.94
IC6	0.08	−2.58	1.10	−0.44	0.50	1.78	−0.44
IC1	0.02	−0.10	0.11	−0.51	0.04	0.39	0.05
IC2	0.34	−0.13	−0.18	−0.13	0.61	−0.49	−0.01
IC3	1.29	−0.27	−0.13	−0.15	−0.23	−0.61	0.10
IC4	−0.03	0.03	−0.66	0.49	0.28	0.55	−0.67
IC5	0.06	0.80	0.05	−0.07	0.15	−0.13	−0.85
IC6	0.16	−1.11	0.45	−0.18	0.27	0.75	−0.34
IC1	−0.04	0.10	−0.05	0.51	−0.18	−0.30	−0.04
IC2	0.26	−0.06	−0.21	−0.00	0.60	−0.54	−0.04
IC3	1.31	−0.27	−0.12	−0.19	−0.19	−0.64	0.10
IC4	0.00	0.07	−0.67	0.50	0.27	0.52	−0.69
IC5	0.03	0.77	0.08	−0.09	0.15	−0.11	−0.84
IC6	0.16	−1.13	0.43	−0.17	0.28	0.76	−0.33
IC1	0.05	−0.14	0.12	−0.50	−0.01	0.44	0.04
IC2	0.39	−0.12	−0.15	−0.21	0.59	−0.52	0.01
IC3	1.26	−0.21	−0.17	−0.10	−0.25	−0.63	0.10
IC4	−0.05	0.03	−0.67	0.50	0.32	0.54	−0.67
IC5	−0.03	1.14	−0.10	−0.01	0.04	−0.35	−0.69
IC6	0.23	−0.78	0.42	−0.18	0.30	0.62	−0.61

when they are compared with adjacent observations. In contrast, the observations with high Pb appear to occur more randomly and are clear outliers when they are compared with adjacent observations. The three most extreme Pb values are 1110.0, 388.0 and 284.0 in the original scale, when 97.5% of the data are in the range of [4.07, 47.4]. For these observations, the assigned the weights are 0.239, 0.175, 0.124 by RSBSS-W and 0.015, 0.031 and 0.057 by RSBSS-Q, from the highest observation to the lowest

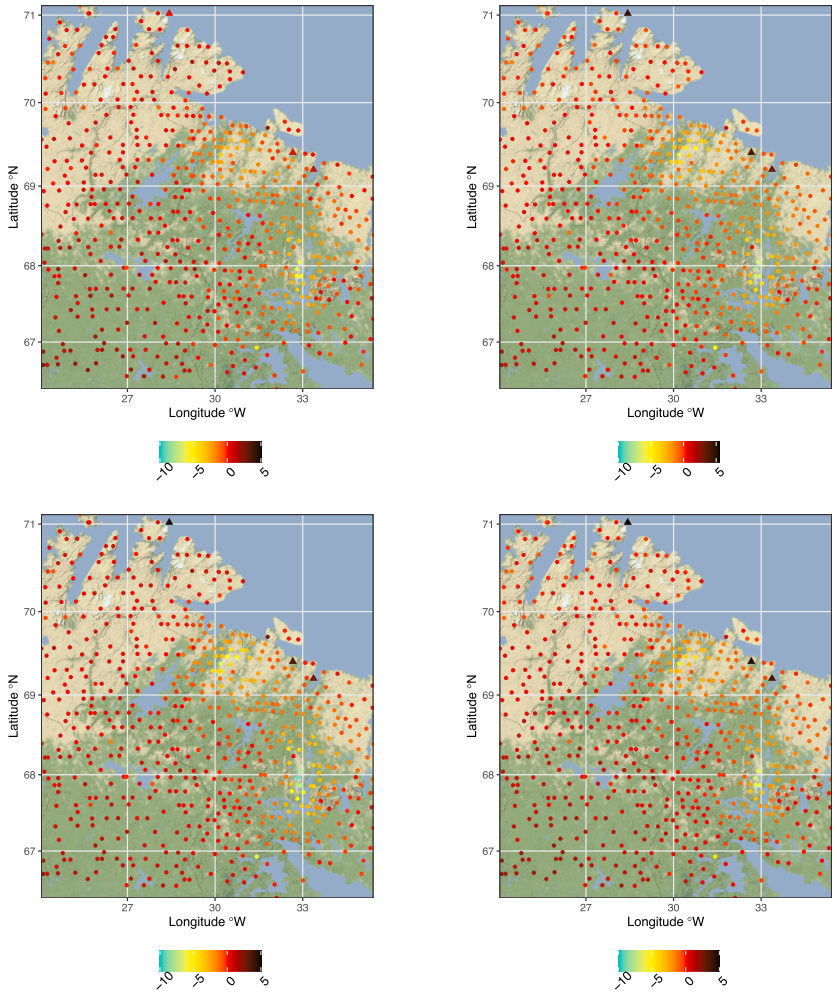


Fig. 13. Maps for the first latent field for the SBSS (upper left), RSBSS-S (upper right), RSBSS-W (lower left) and RSBSS-Q (lower right) methods. The fields are scaled to have zero median and mean absolute deviation equal to one. The outliers with unusually high occurrence in Pb are marked by triangles.

one. For both methods, these weights are within five smallest weights. These observations are marked as triangles in Figs. 13 and 14, where the northernmost triangle shows the extreme Pb value. The weights provided by RSBSS-W and RSBSS-Q appear to be useful in outlier detection, as the most downweighted 10% of the observations match well with the outlier detection results of Filzmoser et al. (2005).

After applying all four methods, several values/tools can be used to elaborate on the results. In the first step, the pseudoeigenvalues of the (robust) local covariance matrices highlight the importance of the separated latent random fields. According to Table 2, all four methods emphasize the first two latent fields. The robust methods emphasize the first latent field more than SBSS, whereas SBSS has a stronger weight in the second latent field.

On inspecting the loading matrices (from clr to ICs) for all four methods in Table 3, it becomes evident that the first component is formed with high loadings for Cu and As for the classical SBSS. In contrast, by downweighting the outlying observations, robust SBSS methods emphasize Cu and Pb. In addition, the RSBSS-W method has a small loading of Mg, which is not present in the other two robust methods. The maps of the first latent component illustrated by Fig. 13, all show clear clusters around the heavy metal industry regions. The fields resulting from the robust methods have considerably high values in the locations with the three largest Pb values, which are marked as triangles. For the second latent component, the classical SBSS considers Mg and Pb and, to a lesser extent, Co and Zn to be important. The robust methods favor the Mg, Pb and As, but assign a very low loading value to Zn (in contrast to the classical SBSS). Moreover, Cu is set to zero in the RSBSS-W method but not in the other two robust ones. The maps for the second latent field (Fig. 14) show a clear separation between the coastal region and the inland. The most noticeable difference between the results of SBSS and the robust methods is in loading of the Pb in the first latent field. A plausible explanation is that to capture behaviors similar to that obtained using the robust methods, the classical SBSS method has to ignore the entire element with such large outliers. Overall, the first latent field might separate the industrial contamination effect from the background and

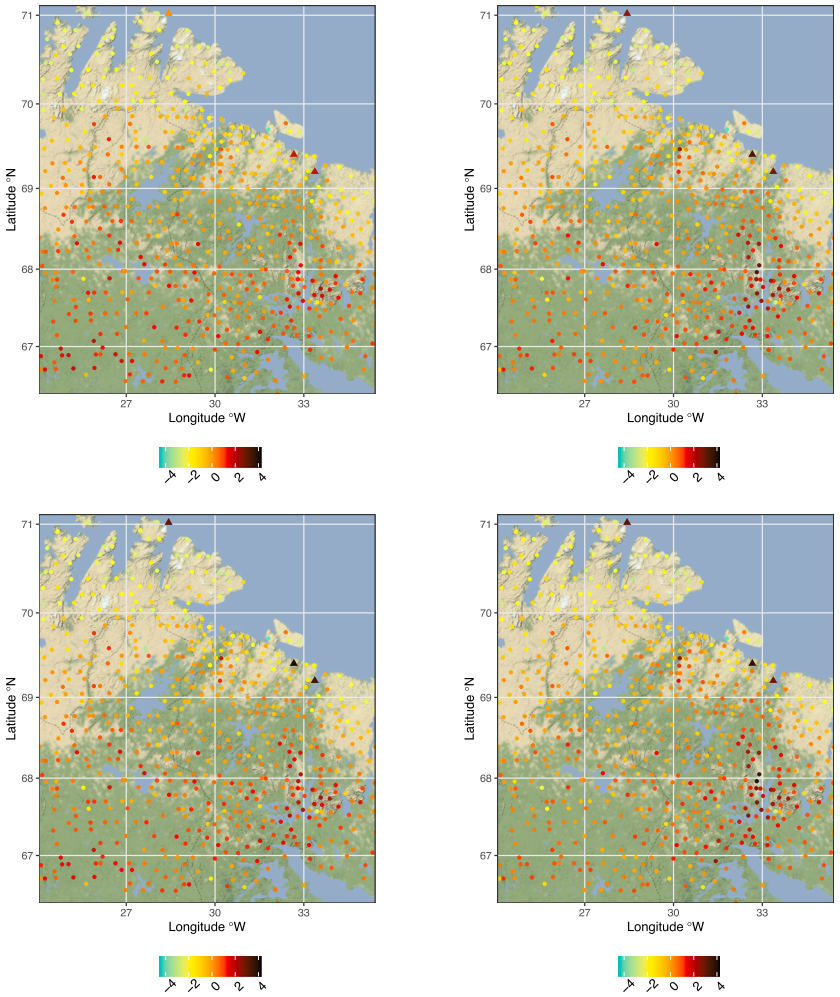


Fig. 14. Maps for the second latent field for the SBSS (upper left), RSBSS-S (upper right), RSBSS-W (lower left) and RSBSS-Q (lower right) methods. The fields are scaled to have zero median and mean absolute deviation equal to one. The outliers with unusually high occurrence in Pb are marked by triangles.

Table 4
Pairwise MD indices between different loadings matrices (from ilr to ICs).

	SBSS	RSBSS-S	RSBSS-W	RSBSS-Q
SBSS	0.00	0.31	0.25	0.45
RSBSS-S	0.00	0.00	0.15	0.23
RSBSS-W	0.00	0.00	0.00	0.32
RSBSS-Q	0.00	0.00	0.00	0.00

the second one the marine effect from the background. A more in-depth interpretation of the different results can be provided by domain experts.

Lastly, the differences in the loading matrices for the four different spatial BSS methods can be quantified by calculating the MD index amongst all pairwise combinations, as shown in Table 4. As already hinted by the interpretation of the loadings matrices, the results of SBSS, when compared with those of its robust counterparts, differ considerably. The most significant differences were observed when SBSS was compared with RSBSS-Q. RSBSS-S and RSBSS-W delivered similar results, whereas RSBSS-Q, having heavier downweighting for the outlying observations, displayed slightly different results also when compared with RSBSS-S and RSBSS-W.

5. Discussion and conclusion

We suggested a novel robust BSS method for stationary spatial data which, based on our simulation studies, is more efficient than the classical method for heavy-tailed fields and robust against global and local outliers. Notably, the robustness of our method against local outliers is only minimally better than that of the classical method. Overall, we noticed that the effect of local outliers

on the estimates appeared less severe than that of global outliers. All the methods discussed herein require a data analyst to decide which kernels and how many of them to use for the local covariance matrices. Piccolotto et al. (2022) suggested a visual analytic approach to guide the choice for classical methods, and we believe that their approach is also applicable to the framework suggested herein.

The robust methods proposed in this paper utilize (generalized) spatial signs. In future works, we will explore other approaches for increasing the robustness of the SBSS method, particularly considering local outliers. SBSS was recently extended to non-stationary spatial data (Muehlmann et al., 2022a), and another objective of ours is to develop robust methods for this setting. Furthermore, we intend to develop the large sample properties of the suggested methods in the future studies. The challenge here is that for spatial methods there are two asymptotic frameworks available, infill asymptotics and increasing domain asymptotics. These will be explored in a separate study. As shown using the example in Section 4, several components appear to have very small pseudoeigenvalues, indicating that these components may be white noise. Muehlmann et al. (2024) provided a test for the number of white noise components in the classical SBSS context and we believe that a similar test can also be provided for the robust methods.

Acknowledgments

We thank the two anonymous reviewers for helping us improve and clarify this manuscript. We are grateful to Marie Ernst and Gentiane Haesbroeck for giving us access to their code for producing local spatial outliers. We also wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. The work of CM and KN was supported by the Austrian Science Fund P31881-N32. The work of KN and ST was supported by the COST Action HiTEc (CA21163). The work of MS was supported by the Vilho, Yrjö and Kalle Väisälä Fund, Finland.

References

Bachoc, F., Genton, M.G., Nordhausen, K., Ruiz-Gazen, A., Virta, J., 2020. Spatial blind source separation. Biometrika 107, 627–646. http://dx.doi.org/10.1093/biomet/asz079.

Belouchrani, A., Abed Meraim, K., Cardoso, J.-F., Moulines, E., 1997. A blind source separation technique based on second order statistics. IEEE Trans. Signal Process. 45, 434–444. http://dx.doi.org/10.1109/78.554307.

Bivand, R., Keitt, T., Rowlingson, B., 2022. rgdal: Bindings for the 'Geospatial' data abstraction library. URL: https://CRAN.R-project.org/package=rgdal, R package version 1.5-32.

Chakraborty, B., Chaudhuri, P., Oja, H., 1998. Operating transformation retransformation on spatial median and angle test. Statist. Sinica 8, 767–784.

Clarkson, D., 1988. A least squares version of algorithm AS 211: The F-G diagonalization algorithm. Appl. Stat. 37, 317–321. http://dx.doi.org/10.2307/2347375.

Ernst, M., Haesbroeck, G., 2017. Comparison of local outlier detection techniques in spatial multivariate data. Data Min. Knowl. Discov. 31, 371–399. http://dx.doi.org/10.1007/s10618-016-0471-0.

Filzmoser, P., 2020. StatDA: Statistical analysis for environmental data. URL: https://CRAN.R-project.org/package=StatDA, R package version 1.7.4.

Filzmoser, P., Garrett, R.G., Reimann, C., 2005. Multivariate outlier detection in exploration geochemistry. Comput. Geosci. 31 (5), 579–587. http://dx.doi.org/10.1016/j.cageo.2004.11.013.

Filzmoser, P., Hron, K., Templ, M., 2018. Applied Compositional Data Analysis: With Worked Examples in R. Springer, Cham, http://dx.doi.org/10.1007/978-3-319-96422-5.

Filzmoser, P., Ruiz-Gazen, A., Thomas-Agnan, C., 2014. Identification of local multivariate outliers. Statist. Papers 55, 29–47. http://dx.doi.org/10.1007/s00362-013-0524-z.

Genton, M.G., Kleiber, W., 2015. Cross-covariance functions for multivariate geostatistics. Statist. Sci. 30 (2), 147–163. http://dx.doi.org/10.1214/14-STS487.

Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W., 2011. Robust Statistics: The Approach Based on Influence Functions. Wiley, New York.

Harris, P., Brunsdon, C., Charlton, M., Juggins, S., Clarke, A., 2014. Multivariate spatial outlier detection using robust geographically weighted methods. Math. Geosci. 46, 1–31. http://dx.doi.org/10.1007/s11004-013-9491-0.

Hettmansperger, T.P., Randles, R.H., 2002. A practical affine equivariant multivariate median. Biometrika 89, 851–860. http://dx.doi.org/10.1093/biomet/89.4.851.

Huber, P., Ronchetti, E., 2011. Robust Statistics. Wiley, Hoboken.

Illner, K., Miettinen, J., Fuchs, C., Taskinen, S., Nordhausen, K., Oja, H., Theis, F.J., 2015. Model selection using limiting distributions of second-order blind source separation algorithms. Signal Process. 113, 95–103. http://dx.doi.org/10.1016/j.sigpro.2015.01.017.

Ilmonen, P., Nordhausen, K., Oja, H., Ollila, E., 2010. A new performance index for ICA: properties, computation and asymptotic analysis. In: International Conference on Latent Variable Analysis and Signal Separation. Springer, pp. 229–236. http://dx.doi.org/10.1007/978-3-642-15995-4_29.

Ilmonen, P., Nordhausen, K., Oja, H., Theis, F., 2015. An affine equivariant robust second-order BSS method. In: Vincent, E., Yeredor, A., Koldovský, Z., Tichavský, P. (Eds.), Latent Variable Analysis and Signal Separation: 12th International Conference. LVA/ICA 2015, Springer International Publishing, pp. 328–335.

Ilmonen, P., Paindaveine, D., 2011. Semiparametrically efficient inference based on signed ranks in symmetric independent component models. Ann. Statist. 39, 2448–2476, URL: https://api.semanticscholar.org/CorpusID:18399840.

Kerry, R., Oliver, M., 2007. Determining the effect of asymmetric data on the variogram. II. outliers. Comput. Geosci. 33 (10), 1233–1260.

Kleiber, W., Nychka, D., 2012. Nonstationary modeling for multivariate spatial processes. J. Multivariate Anal. 112, 76–91. http://dx.doi.org/10.1016/j.jmva.2012.05.011.

Lietzen, N., Virta, J., Nordhausen, K., Ilmonen, P., 2020. Minimum distance index for BSS, generalization, interpretation and asymptotics. Austrian J. Stat. 49 (4), 57–68. http://dx.doi.org/10.17713/ajs.v49i4.1130.

Maronna, R.A., Martin, R.D., Yohai, V.J., Salibián-Barrera, M., 2019. Robust Statistics: Theory and Methods (with R). John Wiley & Sons, New York.

Matilainen, M., Nordhausen, K., Oja, H., 2015. New independent component analysis tools for time series. Statist. Probab. Lett. 105, 80–87. http://dx.doi.org/10.1016/j.spl.2015.04.033.

Miettinen, J., Illner, K., Nordhausen, K., Oja, H., Taskinen, S., Theis, F., 2016. Separation of uncorrelated stationary time series using autocovariance matrices. J. Time Series Anal. 37, 337–354. http://dx.doi.org/10.1111/jtsa.12159.

Miettinen, J., Nordhausen, K., Taskinen, S., 2017. Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp. J. Stat. Softw. 76 (2), 1–31. http://dx.doi.org/10.18637/jss.v076.i02.

Miettinen, J., Taskinen, S., Nordhausen, K., Oja, H., 2015. Fourth moments and independent component analysis. Statist. Sci. 30 (3), 372–390. http://dx.doi.org/10.1214/15-STS520.

Möttönen, J., Oja, H., 1995. Multivariate spatial sign and rank methods. J. Nonparametr. Stat. 5, 201–213. http://dx.doi.org/10.1080/10485259508832643.
Muehlmann, C., Bachoc, F., Nordhausen, K., 2022a. Blind source separation for non-stationary random fields. Spatial Stat. 47, 100574. http://dx.doi.org/10.1016/j.spasta.2021.100574.
Muehlmann, C., Bachoc, F., Nordhausen, K., Yi, M., 2024. Test of the latent dimension of a spatial blind source separation model. Stat. Sinica http://dx.doi.org/10.5705/ss.202021.0326, (in press).
Muehlmann, C., Fačevicová, K., Gardlo, A., Janečková, H., Nordhausen, K., 2021a. Independent component analysis for compositional data. In: Daouia, A., Ruiz-Gazen, A. (Eds.), Advances in Contemporary Statistics and Econometrics: Festschrift in Honor of Christine Thomas-Agnan. Springer, Cham, pp. 525–545. http://dx.doi.org/10.1007/978-3-030-73249-3_27.
Muehlmann, C., Nordhausen, K., Yi, M., 2021b. On cokriging, neural networks, and spatial blind source separation for multivariate spatial prediction. IEEE Geosci. Remote Sens. Lett. 18 (11), 1931–1935. http://dx.doi.org/10.1109/LGRS.2020.3011549.
Muehlmann, C., Sipilä, M., Nordhausen, K., Taskinen, S., Virta, J., 2022c. SpatialBSS: Blind source separation for multivariate spatial data. URL: https://CRAN.R-project.org/package=SpatialBSS, R package version 0.13-0.
Muler, N., Pena, D., Yohai, V.J., 2009. Robust estimation for ARMA models. Ann. Statist. 37 (2), 816–840. http://dx.doi.org/10.1214/07-AOS570.
Nordhausen, K., Fischer, G., Filzmoser, P., 2021. Blind source separation for compositional time series. Math. Geosci. 53, 905–924. http://dx.doi.org/10.1007/s11004-020-09869-y.
Nordhausen, K., Oja, H., 2018. Independent component analysis: A statistical perspective. WIREs: Comput. Stat. 10, e1440. http://dx.doi.org/10.1002/wics.1440.
Nordhausen, K., Oja, H., Filzmoser, P., Reimann, C., 2015. Blind source separation for spatial compositional data. Math. Geosci. 47 (7), 753–770. http://dx.doi.org/10.1007/s11004-014-9559-5.
Nordhausen, K., Ollila, E., Oja, H., 2011. On the performance indices of ICA and blind source separation. In: 2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications. pp. 486–490. http://dx.doi.org/10.1109/SPAWC.2011.5990458.
Nordhausen, K., Ruiz-Gazen, A., 2022. On the usage of joint diagonalization in multivariate statistics. J. Multivariate Anal. 188, 104844. http://dx.doi.org/10.1016/j.jmva.2021.104844.
Nordhausen, K., Tyler, D.E., 2015. A cautionary note on robust covariance plug-in methods. Biometrika 102 (3), 573–588. http://dx.doi.org/10.1093/biomet/asv022.
Oja, H., 2010. Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks. Springer, New York.
Oja, H., Sirkiä, S., Eriksson, J., 2006. Scatter matrices and independent component analysis. Austrian J. Stat. 35, 175–189.
Pan, Y., Matilainen, M., Taskinen, S., Nordhausen, K., 2022. A review of second-order blind identification methods. WIREs Comput. Stat. 14, e1550. http://dx.doi.org/10.1002/wics.1550.
Pebesma, E.J., Bivand, R.S., 2005. Classes and methods for spatial data in R. R News 5 (2), 9–13, URL: https://CRAN.R-project.org/doc/Rnews/.
Piccolotto, N., Bögl, M., Muehlmann, C., Nordhausen, K., Filzmoser, P., Miksch, S., 2022. Visual parameter selection for spatial blind source separation. Comput. Graph. Forum 41 (3), 157–168. http://dx.doi.org/10.1111/cgf.14530.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL: https://www.R-project.org/.
Raymaekers, J., Rousseeuw, P., 2019. A generalized spatial sign covariance matrix. J. Multivariate Anal. 171, 94–111. http://dx.doi.org/10.1016/j.jmva.2018.11.010.
Reimann, C., Banks, D., Kashulina, G., 2000. Processes influencing the chemical composition of the O-horizon of podzols along a 500-km north–south profile from the coast of the Barents sea to the Arctic circle. Geoderma 95 (1), 113–139. http://dx.doi.org/10.1016/S0016-7061(99)00088-9.
Reimann, C., Filzmoser, P., Garrett, R., Dutter, R., 2008. Statistical Data Analysis Explained: Applied Environmental Statistics with R. Wiley, Chichester, http://dx.doi.org/10.1002/9780470987605.
Rousseeuw, P.J., 2006. Maxbias curve. In: Encyclopedia of Statistical Sciences. John Wiley & Sons, http://dx.doi.org/10.1002/0471667196.ess1094.pub2.
Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Auel, J., Menck, P.J., Gross, S., Ober, U., Ribeiro, P., Ripley, B.D., Singleton, R., Pfaff, B., R Core Team, 2022. RandomFields: Simulation and analysis of random fields. R package version 3.3.14.
Tobler, W., 1970. A computer movie simulating urban growth in the detroit region. Econ. Geogr. 46, 234–240. http://dx.doi.org/10.2307/143141.
Tong, L., Soon, V., Huang, Y., Liu, R., 1990. AMUSE: A new blind identification algorithm. In: Proceedings of IEEE International Symposium on Circuits and Systems. IEEE, pp. 1784–1787. http://dx.doi.org/10.1109/ISCAS.1990.111981.
Tyler, D.E., 1987. A distribution-free M-estimator of multivariate scatter. Ann. Statist. 15, 234–251. http://dx.doi.org/10.1214/aos/1176350263.
Wickham, H., 2016. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, URL: https://ggplot2.tidyverse.org.