Blind Source Separation over Space: an eigenanalysis approach

Bo Zhang
Department of Statistics and Finance, International Institute of Finance
School of Management, University of Science and Technology of China, Hefei, China
zhangbo890301@outlook.com

Sixing Hao and Qiwei Yao
Department of Statistics, London School of Economics, London, WC2A 2AE, UK
s.hao3@lse.ac.uk q.yao@lse.ac.uk

December 16, 2023

Abstract

We propose a new estimation method for the blind source separation model of Bachoc et al. (2020). The new estimation is based on an eigenanalysis of a positive definite matrix defined in terms of multiple normalized spatial local covariance matrices, and, therefore, can handle moderately high-dimensional random fields. The consistency of the estimated mixing matrix is established with explicit error rates even when the eigen-gap decays to zero slowly. The proposed method is illustrated via both simulation and a real data example.

Key works: Eigen-analysis; Eigen-gap; High-dimensional random field; Mixing matrix; Normalized spatial local covariance matrix.

1 Introduction

Blind source separation is an effective way to reduce the complexity in modelling p-variant spatial data (Nordhausen et al., 2015; Bachoc et al., 2020). The approach can be viewed as a version of independent component analysis (Hyvärinen et al., 2001) for multivariate spatial random fields. Though only the second moment properties are concerned, the challenge is to

decorrelate p spatial random fields at the same location as well as across different locations. Note that the standard principal component analysis does not capture spatial correlations, as it only diagonalizes the covariance matrix (at the same location). Nordhausen et al. (2015) introduced a so-called local covariance matrix (see (4) in section 2.2 below) to represent the dependence across different locations. Furthermore, it proposed to estimate the mixing matrix, defined in (1) in section 2.1 below, in the blind source separation decomposition based on a generalized eigenanalysis, which can be viewed as an extension of the principal component analysis as it diagonalizes a local covariance matrix in addition to the standard covariance matrix. To overcome the drawback of using the information from only one local covariance matrix, Bachoc et al. (2020) proposed to use multiple local covariance matrices in the estimation (see (5) in section 2.2). The method of Bachoc et al. (2020) has a clear advantage in incorporating the spatial dependence information over different ranges. It is in the spirit of JADE (joint approximate diagonalization of eigenmatrices) in non-spatial contexts. See Chapter 11 of Hyvärinen et al. (2001) and the references within. Its estimation is based on a nonlinear optimization with p^2 parameters. Hence it is compute-intensive and cannot cope with very large p.

Inspired by Bachoc et al. (2020), we propose a new method also based on multiple (normalized) local covariance matrices for estimating the mixing matrix. Different from Bachoc et al. (2020), the new method is computationally efficient as it boils down to an eigenanalysis of a positive definite matrix which is a matrix function of multiple normalized spatial local covariance matrices. Therefore it can handle the cases with the dimension of random fields in the order of a few thousands on an ordinary personal computer. While the basic idea resembles that of Chang, Guo and Yao (2018) which dealt with multiple time series, the spatial random fields concerned are sampled irregularly and non-unilaterally, and the spatial correlations spread in all directions. Furthermore, we incorporate the pre-whitening in our search for the mixing matrix. This implies estimating the covariance matrix of the process, which is assumed to be an identity matrix in Chang, Guo and Yao (2018). The normalized spatial local covariance matrix, defined in (10) below, is a modified version of the spatial local covariance matrix in Nordhausen et al. (2015), and is introduced to facilitate the effect of the pre-whitening. All these entail completely different theoretical exploration; leading to the asymptotic results under the similar setting of Bachoc et al. (2020) but allowing the dimension of the random field to diverge together with the number of the observed locations, which is assumed to be fixed in Bachoc et al. (2020).

The efficiency gain in computing of the proposed method is due to adding together the information from different normalized local covariance matrices. However different from adding covariance matrices directly such as in TDSEP (Ziehe and Müller, 1998), each term in the sum of (9) in Section 2.3 below is the product of a normalized local covariance matrix and its transpose, which, therefore, is non-negative definite matrix. This avoids the possible cancellation of the information from different normalized local covariance matrices. Note covariance matrices are not non-negative definite, and adding them together directly may leads to volatile performance due to information cancellation; see Table 1 of (Ziehe and Müller, 1998). Although the sample fourth moments occur in (9) in order to avoid the information cancellation, our goal is decorrelaton across space via diagonalizing multiple normalized local covariance matrices. Indeed the way to use the fourth moments and the purpose of using them are radically different from those of FOBI (forth-order blind identification) algorithms. See Chapter 11 of Hyvärinen et al. (2001) and the references within.

Another new contribution of the paper concerns the eigen-gap in the eigenanalysis for estimating the mixing matrix. In order to identify a consistent estimator for the mixing matrix, the standard condition is to assume that the minimum pairwise absolute difference among the eigenvalues remains positive. See Assumptions 8 and 9 of Bachoc et al. (2020). The similar conditions have been imposed in the literature in order to identify factor loading spaces in factor models (Lam and Yao, 2012). However this condition is invalid under the setting concerned in this paper when the dimension of random field p diverges to infinity, as the maximum order of the eigen-gap is p^{-1}. We show that the identification of the mixing matrix is still possible when $p \to \infty$ at the rate $p = o(n^{1/3})$. See Theorem 2 and Remark 2 in Section 3.

The rest of the paper is organised as follows. We present the spatial blind source separation model and the new estimation method in Section 2. The asymptotic properties are developed in Section 3. Numerical illustration with both simulated data and a real data set is presented in Section 4. All the technical proofs are given in the Appendix.

The R-package `BSSoverSpace`, available in the CRAN project, implements the methods proposed in this paper.

2 Setting and Methodology

2.1 Model

We adopt the spatial blind source separation model of Bachoc et al. (2020). More precisely, let $X(s) = \{X_1(s), \cdots, X_p(s)\}^\top$ be a p-variate random field defined on $s \in \mathcal{S} \subset \mathcal{R}^d$, and $X(s)$ admits the representation

$$X(s) = \Omega Z(s) \equiv \Omega\{Z_1(s), \cdots, Z_p(s)\}^\top, \tag{1}$$

where $Z_1(s), \cdots, Z_p(s)$ are p independent latent random fields, and Ω is a $p \times p$ invertible constant matrix and is called the mixing matrix. Furthermore, Bachoc et al. (2020) assumes that for any $s, u \in \mathcal{S}$,

$$EZ(s) = \mu_0, \quad \mathrm{Var}\{Z(s)\} = I_p, \quad \mathrm{Cov}\{Z(s), Z(u)\} = H(s-u), \tag{2}$$

where μ_0 is an unknown constant vector, I_p denotes the $p \times p$ identity matrix, $H(\cdot)$ is a $p \times p$ diagonal matrix

$$H(s-u) = \mathrm{diag}\{K_1(s-u), \cdots, K_p(s-u)\},$$

i.e. $\mathrm{Cov}\{Z_i(s), Z_j(u)\} = K_i(s-u)$ if $i = j$, and 0 otherwise. Let $\mu = \Omega\mu_0$. Under (1) and (2), $X(\cdot)$ is a weakly stationary process as

$$EX(s) = \mu, \quad \mathrm{Var}\{X(s)\} = \Omega\Omega^\top, \quad \mathrm{Cov}\{X(s), X(u)\} = \Omega H(s-u)\Omega^\top. \tag{3}$$

2.2 The existing methods

Let $X(s_1), \cdots, X(s_n)$ be available observations. Put

$$\tilde{X}(s_i) = X(s_i) - \frac{1}{n}\sum_{j=1}^n X(s_j), \quad \tilde{Z}(s_i) = Z(s_i) - \frac{1}{n}\sum_{j=1}^n Z(s_j), \quad i = 1, \cdots, n.$$

Then the spatial local covariance matrix of Nordhausen et al. (2015) is defined as

$$\tilde{M}(f) = \frac{1}{n}\sum_{i,j=1}^n f(s_i - s_j)\tilde{X}(s_i)\tilde{X}(s_j)^\top, \tag{4}$$

where $f(\cdot)$ is a kernel function such as $f(s) = 1(h_1 \le \|s\| \le h_2)$ for some constants $0 \le h_1 < h_2 < \infty$, and $1(\cdot)$ denotes the indicator function. To recover the mixing matrix Ω, Bachoc et

al. (2020) proposed to estimate the unmixing matrix (i.e. the inverse of the mixing matrix) $\Gamma = \Omega^{-1} \equiv (\gamma_1, \cdots, \gamma_p)^\top$ by

$$\hat{\Gamma} \in \arg \max_{\Gamma \tilde{M}(f_0) \Gamma^\top = I_p} \sum_{i=1}^{k} \sum_{j=1}^{p} \{\gamma_j^\top \tilde{M}(f_i) \gamma_j\}^2, \tag{5}$$

where $f_0(s) = I(s = 0)$, and $f_1, \cdots, f_k$ are appropriately specified kernels. This is a nonlinear optimization problems with p^2 variables, which Bachoc et al. (2020) adopted the algorithm of Clarkson (1988) to solve. When $k = 1$, the objective function contains only one kernel function. Then the above optimization can be solved based on a generalized eigenanalysis; see Nordhausen et al. (2015) and Bachoc et al. (2020), though the estimation based on a single kernel requires the prior knowledge on which kernel to use for a given problem.

2.3 The new method

We now propose a new method to estimate the mixing matrix using multiple kernels but based on a single eigenanalysis. To this end, we define, for any given k kernel function $f_1(\cdot), \cdots, f_k(\cdot)$,

$$\begin{aligned} N &= E\Big[\frac{1}{k} \sum_{h=1}^{k} \big\{\frac{1}{n} \sum_{i,j=1}^{n} f_h(s_i - s_j) \tilde{Z}(s_i) \tilde{Z}(s_j)^\top\big\} \big\{\frac{1}{n} \sum_{i,j=1}^{n} f_h(s_i - s_j) \tilde{Z}(s_i) \tilde{Z}(s_j)^\top\big\}^\top\Big], \qquad (6) \\ W &= E\Big[\frac{1}{k} \sum_{h=1}^{k} \big\{\frac{1}{n} \sum_{i,j=1}^{n} f_h(s_i - s_j) \Sigma^{-1/2} \tilde{X}(s_i) \tilde{X}(s_j)^\top\big\} \Sigma^{-1} \\ &\quad \times \big\{\frac{1}{n} \sum_{i,j=1}^{n} f_h(s_i - s_j) \tilde{X}(s_i) \tilde{X}(s_j)^\top \Sigma^{-1/2}\big\}^\top\Big], \end{aligned}$$

where $\Sigma = \mathrm{Var}\{X(s)\} = \Omega\Omega^\top$. Then N and W are $p \times p$ non-negative definite matrices. Furthermore, N is a diagonal matrix, as its (i, j)-th element, for $i \neq j$, is

$$\frac{1}{n^2 k} \sum_{h=1}^{k} \sum_{\ell=1}^{p} \sum_{i_1, i_2, j_1, j_2 = 1}^{n} f_h(s_{i_1} - s_{j_1}) f_h(s_{i_2} - s_{j_2}) E\{\tilde{Z}_i(s_{i_1}) \tilde{Z}_\ell(s_{j_1}) \tilde{Z}_j(s_{i_2}) \tilde{Z}_\ell(s_{j_2})\} = 0,$$

which is guaranteed by the fact that the components of $Z(\cdot)$ are the p independent random fields. Since Ω is a $p \times p$ full rank matrix, we can rewrite $\Omega = V_\Omega \Lambda_\Omega U_\Omega$, where V_Ω and U_Ω are two $p \times p$ orthogonal matrices, and Λ_Ω is a diagonal matrix. Then $\Sigma^{-1/2} = V_\Omega \Lambda_\Omega^{-1} V_\Omega^\top$. Combining this and (1), we have

$$W = V_\Omega U_\Omega N U_\Omega^\top V_\Omega^\top, \tag{7}$$

i.e. the columns of $U_W \equiv V_\Omega U_\Omega$ are the p orthonormal eigenvectors of matrix W with the diagonal elements of N as the corresponding eigenvalues. As $\Sigma^{1/2}U_W = V_\Omega \Lambda_\Omega V_\Omega^\top V_\Omega U_\Omega = \Omega$, this paves the way to identifying mixing matrix Ω. We summarize the finding in the proposition below.

Proposition 1. *Under the condition (2), the mixing matrix Ω defined in (1) is of the form $\Sigma^{1/2}U_W$, where the columns of U_W are the p orthonormal eigenvectors of matrix W. Moreover, those p eigenvectors are identifiable, upto the sign changes, if the p diagonal elements of N are distinct from each other.*

Note that the sign changes of any columns of U_W will not change the independence of the components of $Z(\cdot)$ in (1), as $Z(s) = U_W^\top \Sigma^{-1/2} X(s)$. By Proposition 1, we define an estimator for the mixing matrix as

$$\hat{\Omega} = \hat{\Sigma}^{1/2}\hat{U}_W, \tag{8}$$

where $\hat{\Sigma} = n^{-1}\sum_{1\le j\le n}\tilde{X}(s_j)\tilde{X}(s_j)^\top$, and the columns of $\hat{U}_W$ are the p orthonormal eigenvectors of matrix

$$\hat{W} = \frac{1}{k}\sum_{h=1}^{k}\hat{M}(f_h)\hat{M}(f_h)^\top. \tag{9}$$

In the above expression, $\hat{M}(f_h)$ is a normalized local covariance matrix defined as

$$\hat{M}(f) = \frac{1}{n}\sum_{i,j=1}^{n} f(s_i - s_j)\hat{\Sigma}^{-1/2}\tilde{X}(s_i)\tilde{X}(s_j)^\top\hat{\Sigma}^{-1/2}. \tag{10}$$

This estimation procedure is implemented in Algorithm 1 below. In comparison to the local covariance matrix (4), we replace $X(\cdot)$ by its standardized version $\hat{\Sigma}^{-1/2}\tilde{X}(\cdot)$. This effectively pre-whitens the data in our search for the mixing matrix.

Algorithm 1: Eigenanalysis approach for BSS over space

Input: $X(s_1), \cdots, X(s_n)$ and $f_1(\cdot), \cdots, f_k(\cdot)$.

(i) Compute $\tilde{X}(s_i) = X(s_i) - \frac{1}{n}\sum_{j=1}^{n}X(s_j)$ and $\hat{\Sigma} = n^{-1}\sum_{1\le j\le n}\tilde{X}(s_j)\tilde{X}(s_j)^\top$.

(ii) Compute $\hat{W}$ in (9).

(iii) Compute eigenvalues $\hat{\Lambda}_W$ and eigenvectors $\hat{U}_W$ of matrix $\hat{W}$.

(iv) Compute $\hat{\Omega}^{-1} = \hat{U}_W^\top\hat{\Sigma}^{-1/2}$.

Output: $\hat{Z}(s_i) = \hat{\Omega}^{-1}X(s_i)$, $i = 1, \cdots, n$.

Remark 1. The proposed new method makes use of the normalized 4th moments of the observations while the methods of Bachoc et al. (2020) and Nordhausen et al. (2015) only

depend on the 2nd moments. However the 4th moments occur only in the matrix products $\hat{M}(f_h)\hat{M}(f_h)^\top$ in defining $\hat{W}$ in (9), and each of those products is a non-negative definite matrix. We add together those non-negative definite matrices, instead of $\hat{M}(f_h)$ (as suggested in Ziehe and Müller (1998)), to avoid the information cancellation from different $\hat{M}(f_h)$. See also Chang, Guo and Yao (2018). Note that both our way of using the fourth moments and our purpose of using them are radically different from those of FOBI (Hyvärinen et al., 2001, Chapter 11). For example W in (6) is a $p \times p$ matrix with the (l, m)-th element

$$E\Big[\frac{1}{n^2 k}\sum_{h=1}^{k}\sum_{v=1}^{p}\sum_{i,j,c,d=1}^{n} f_h(s_i - s_j)f_h(s_c - s_d)\tilde{Z}_l(s_i)\tilde{Z}_m(s_c)\tilde{Z}_v(s_j)\tilde{Z}_v(s_d)\Big],$$

while a FOBI algorithm would use instead a $p^2 \times p^2$ quadricovariance matrix with the elements being the fourth order cumulants (Ferreol, 2005). Our goal is to avoid information cancellation while diagonalizing different local covariance matrices. FOBI is to diagonalize a quadricovariance matrix.

3 Asymptotic properties

We consider the asymptotic behaviour of the estimator $\hat{\Omega}$ when $n \to \infty$ and p either remaining fixed or $p = o(n)$. Since $\hat{\Omega}^{-1}X(s) = \hat{\Omega}^{-1}\Omega Z(s)$, we will focus on $\hat{\Gamma}_\Omega = \hat{\Omega}^{-1}\Omega$. We introduce some regularity conditions first.

A1. In model (1), $Z_1(\cdot), \cdots, Z_p(\cdot)$ are p independent and strictly stationary random fields on R^d, and condition (2) holds. Furthermore, $Z(\cdot)$ is sub-Gaussian in the sense that there exists a constant $C_0 > 0$ independent of p for which

$$\sup_{\beta \geq 1, 1 \leq i \leq p} \beta^{-1/2}\{E|Z_i(s)|^\beta\}^{1/\beta} \leq C_0. \tag{11}$$

Moreover, for any unit vector $(a_1, \cdots, a_n)^\top \in R^n$ and $1 \leq \ell \leq p$, $\sum_{i=1}^n a_i Z_\ell(s_i)$ is sub-Gaussian.

A2. There exist positive constants Δ, α and A (independent of n and p) such that for any $1 \leq i \neq j \leq n$ and $n \geq 2$, $\|s_i - s_j\| \geq \Delta$, and for $s, u \in R^d$, $1 \leq \ell \leq p$ and $1 \leq h \leq k$ (k is fixed),

$$|\mathrm{Cov}\{Z_\ell(s+u), Z_\ell(s)\}| \leq A/(1 + \|u\|^{d+\alpha}), \tag{12}$$

$$|f_h(s)| \leq A/(1 + \|s\|^{d+\alpha}). \tag{13}$$

A3. Let $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ be the diagonal elements of matrix N defined in (6), arranged in the descending order. There exist integers $0 = p_0 < p_1 < \cdots < p_m = p$ for which

$$\limsup_{n\to\infty} \max_{1\leq i\leq m} |\lambda_{p_{i-1}+1} - \lambda_{p_i}| = 0, \qquad \text{and} \tag{14}$$

$$\liminf_{n\to\infty} \min_{1\leq i<m} |\lambda_{p_i} - \lambda_{p_i+1}| = C_1 > 0, \tag{15}$$

where $m \geq 2$ is a fixed integer, and C_1 is a constant independent of p.

Conditions A1 and A2 are essentially the same as Assumptions 1-7 of Bachoc et al. (2020), though we impose only the sub-Gaussianality instead of requiring $Z(\cdot)$ to be normally distributed. In addition, our setting allows p to diverge together with n. Condition A3 is required for distinguishing the columns of the mixing matrix Ω from each other. Those p columns are completely identifiable when p is fixed and $m = p$. Then condition (14) vanishes, and (15) ensures that the p diagonal elements of matrix N are distinct from each other (see Proposition 1). The similar conditions (i.e. with p fixed) were imposed in Bachoc et al. (2020): see Assumptions 8 and 9 therein. Note that condition (15) cannot hold when $m = p \to \infty$. When $p \to \infty$ together with n, (14) and (15) ensure that the estimated mixing matrix $\hat{\Omega}$ transforms $X(\cdot)$ into m independent subvectors; see Theorem 1 below. Recalling the definition of N in (6), we can see that the choice of kernels should satisfy Condition A3. This is the same for Bachoc et al. (2020).

Without the loss of generality, we assume that the p components of $Z(\cdot)$ are arranged in the order such that the diagonal elements of matrix N in (6) are in the descending order. This simplifies the presentation of Theorem 1 substantially. Write $\hat{W} = \hat{U}_W \hat{\Lambda}_W \hat{U}_W^\top$ as its spectral decomposition, i.e.

$$\hat{\Lambda}_W = \mathrm{diag}(\hat{\lambda}_{W,1}, \cdots, \hat{\lambda}_{W,p}),$$

where $\hat{\lambda}_{W,1} \geq \cdots \geq \hat{\lambda}_{W,p} \geq 0$ are the eigenvalues of $\hat{W}$, and the columns of the orthogonal matrix $\hat{U}_W$ are the corresponding eigenvectors. Consequently,

$$\hat{\Gamma}_\Omega = \hat{\Omega}^{-1}\Omega = \hat{U}_W^\top \hat{\Sigma}^{-1/2}\Omega. \tag{16}$$

Corollary 1 below shows that $\hat{\Omega}^{-1}\Omega = \hat{\Gamma}_\Omega \xrightarrow{P} I_p$ when p is finite and $m = p$ in Condition A3. To state a more general result first, put $q_i = p_i - p_{i-1}$ for $i = 1, \cdots, m$ (see Condition A3), and

$$\hat{\Omega}^{-1}\Omega = \hat{\Gamma}_\Omega = \begin{pmatrix} \hat{\Gamma}_{\Omega,11} & \cdots & \hat{\Gamma}_{\Omega,1m} \\ \cdots & \cdots & \cdots \\ \hat{\Gamma}_{\Omega,m1} & \cdots & \hat{\Gamma}_{\Omega,mm} \end{pmatrix}, \tag{17}$$

where submatrix $\hat{\Gamma}_{\Omega,ij}$ is of the size $q_i \times q_j$.

Theorem 1. *Let Conditions A1-A3 hold. As $n \to \infty$ and $p = o(n)$, it holds that*

$$\|\hat{\Gamma}_{\Omega,ii}\| = 1 + O_p\{n^{-1/2}p^{1/2}\}, \quad \|\hat{\Gamma}_{\Omega,ii}\|_{min} = 1 + O_p\{n^{-1/2}p^{1/2}\} \quad 1 \le i \le m, \tag{18}$$

$$\|\hat{\Gamma}_{\Omega,ij}\| = O_p\{n^{-1/2}p^{1/2}\}, \quad 1 \le i \ne j \le m, \quad \text{and} \tag{19}$$

$$\|\hat{\Lambda}_W - \Lambda\| = O_p(n^{-1/2}p^{1/2}), \tag{20}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_p)$, and λ_i are specified in Condition A3.

Theorem 1 implies that $\hat{\Gamma}_{\Omega,ij} \xrightarrow{P} 0$ for any $i \ne j$. Hence the transformed process $\hat{\Omega}^{-1}X(\cdot) = \hat{\Gamma}_{\Omega}Z(\cdot)$ can only be divided into the m asymptotically independent random fields of dimensions $q_1, \cdots, q_m$ respectively. This is due to the lack of separation of the corresponding eigenvalues within each of those m groups; see (14). On the other hand, Theorem 1 still holds, under some additional conditions, if the components of $Z(\cdot)$ within each of those m groups are not independent with each other. Then this is in the spirit of the so-called multidimensional independent component analysis of Cardoso (1998). In practice, one needs to identify the m latent groups among the p components of $\hat{\Omega}^{-1}X(\cdot)$, which can be carried out by adapting the procedures in Section 2.2 of Chang, Guo and Yao (2018). By (20), $\hat{\Lambda}_W$ will indicate how those eigenvalues are different from each other; see Condition A3.

Note that Theorem 1 holds when either p is fixed and finite, or $p/n \to 0$ as $n \to \infty$. When p is fixed and $m = p$ in Condition A3, all $\hat{\Gamma}_{\Omega,ij}$ reduces to a scale and $q_i = 1$. Then Corollary 1 below follows from Theorem 1 immediately.

Corollary 1. *Let Conditions A1-A3 hold with $m = p$, and p be a fixed integer. Then as $n \to \infty$, $\|I_p - \hat{\Omega}^{-1}\Omega\| = O_p(n^{-1/2})$.*

A key condition in Corollary 1 for identifying all the columns of the mixing matrix is that the eigengap defined as

$$v_{\mathrm{gap}} = \min_{1 \le i \ne j \le p} |\lambda_i - \lambda_j| \tag{21}$$

remains bounded away from 0, which is implied by (15) when $p = m$ is fixed. This condition cannot be fulfilled when p diverges (together with n). To appreciate the performance of the proposed procedure when p is large in relation to n, we present Theorem 2 below which indicates that the mixing matrix can still be estimated consistently but at much slower rates when the eigengap v_{gap} decays to 0 provided p diverges to ∞ not too fast; see Remark 2 below.

A4. $\limsup_{n\to\infty} v_{\text{gap}}^{-1} n^{-1/2} p^{1/2} = 0$.

Theorem 2. *Let conditions A1, A2 and A4 hold. Denote by $\hat{\gamma}_{\Omega,ij}$ the (i,j)-th entry of matrix $\hat{\Gamma}_\Omega$. Then as $n, p \to \infty$, it holds that*

$$\hat{\gamma}_{\Omega,ij} = O_p(n^{-1/2} p^{1/2} v_{\text{gap}}^{-1} |j-i|^{-1}) \quad \text{for } 1 \le i \ne j \le p, \quad \text{and} \tag{22}$$

$$\hat{\gamma}_{\Omega,ii} = 1 + O_p(n^{-1} p v_{\text{gap}}^{-2} + n^{-1/2} p^{1/2}) \quad \text{for } i = 1, \cdots, p. \tag{23}$$

Moreover, (20) still holds.

Remark 2. Note that $\lambda_1 - \lambda_p \ge (p-1) v_{\text{gap}}$, and, therefore, $v_{\text{gap}} = O(p^{-1})$. Thus it follows from condition A4 that $p = o(n^{1/3})$, i.e. in order to fully identify the mixing matrix, p cannot be too large in the sense that $p/n^{1/3} \to 0$.

4 Numerical illustration

4.1 Simulation

We illustrate the finite sample properties of the proposed method by simulation. We set the dimension of random fields at p =3 and 50, and the sample size n (i.e. the number of locations) between 100 to 2000. The coordinates of those n locations are drawn independently from $U(0, 50)^2$. Both Gaussian and non-Gaussian random fields are used. Also included in the simulation is the method of Bachoc et al. (2020). For each setting, we replicate the simulation 1000 times.

The p-variate random fields $X(\cdot)$ are generated according to (1) in which $Z_1(\cdot), \cdots, Z_p(\cdot)$ are p independent random fields with either $N(0,1)$ or t_5 marginal distributions, and the Matern correlation function

$$\rho(s) = 2^{1-\kappa} \Gamma(\kappa)^{-1} (s/\phi)^\kappa B_\kappa(s/\phi),$$

where $\kappa > 0$ is the shape parameter, $\phi > 0$ is the range parameter, $\Gamma(\cdot)$ is the Gamma function, and B_κ is the modified Bessel function of the second kind of order κ. We set different values of (κ, ϕ) for different Z_j. More precisely κ's are drawn independently from $U(0,6)$, and ϕ's are drawn independently from $U(0,2)$. The mixing matrix Ω in (1) is set to be the $p \times p$ identity matrix.

To measure the accuracy of the estimation for Ω, we define

$$D(\Omega, \hat{\Omega}) = \frac{1}{2p(\sqrt{p}-1)} \sum_{j=1}^{p} \Big\{ \frac{(\sum_{1\le i\le p} d_{ij}^2)^{1/2}}{\max_{1\le i\le p} |d_{ij}|} + \frac{(\sum_{1\le i\le p} d_{ji}^2)^{1/2}}{\max_{1\le i\le p} |d_{ji}|} - 2 \Big\},$$

where d_{ij} is the (i,j)-th element of matrix $\Omega^{-1}\hat{\Omega}$. As

$$p^{-1/2} \le \max_{1\le i\le p} |d_{ij}| \Big/ \big(\sum_{1\le i\le p} d_{ij}^2 \big)^{1/2} \le 1.$$

it holds that $D(\Omega, \hat{\Omega}) \in [0,1]$, and $D(\Omega, \hat{\Omega}) = 0$ if $\hat{\Omega}$ is a column permutation and/or column sign changes of Ω.

We set $k = 10$ in (9), and

$$f_h(s) = 1(c_{h-1} < \|s\| \le c_h), \qquad h = 1, \cdots, 10, \tag{24}$$

where $0 = c_0 < c_1 < \cdots < c_{10} = \infty$ are specified such that for each $h = 1, \cdots, 10$, $\{(s_i, s_j) : 1 \le i < j \le n,\ c_{h-1} < \|s_i - s_j\| \le c_h\}$ contains the 10% of the total pairs (s_i, s_j), $1 \le i < j \le n$.

The boxplots of $D(\Omega, \hat{\Omega})$ obtained in the 1000 replications are presented in Figures 1–4. Estimations by the method of Bachoc et al. (2020) are computed using the R-function `sbss`, provided in R-package SpatialBSS. In addition to the multiple kernel estimation, we also compute the estimates with a single kernel, using each of the 10 kernels in (24), For computing the multiple kernel method of Bachoc et al. (2020), we set the maximum number of iterations at 2000. By using a single kernel, the method of Bachoc et al. (2020) leads to almost identical estimates as those obtained by the proposed method (with the same single kernel). Therefore we omit the detailed results.

Figures 1 – 4 and Tables 1 – 4 indicate clearly that both the methods with multiple kernels outperform most of those with a single kernel, and the proposed method outperforms the multiple kernel method of Bachoc et al. (2020) especially when p is large (i.e. $p = 50$). The proposed method with multiple kernels performs about the same as that with the best single kernel (i.e. Kernel 1 $f_1(\cdot)$). The accuracy of estimation improves with the increase in the number of observations n, which can be seen as a decrease in $D(\Omega, \hat{\Omega})$ in Figures 1–4. Among all single kernel methods, those using kernel f_1 perform the best, as those estimations include the 10% nearest locations. Indeed the Matern correlation is the strongest at the smallest distance. On the other hand, the performances for the Gaussian and the non-Gaussian random fields are about the same. See Figures 1 & 2, and Figures 3 & 4.

The iterative algorithm for implementing the multiple kernel method of Bachoc et al. (2020) is to solve a nonlinear optimization problem with p^2 parameters. When $p = 50$, it failed to converge within the 2000 iterations in some of the 1000 simulation replications. The numbers of failures with n =100, 500, 1000 and 2000 are, respectively, $3, 1, 2$ and 1 for the Gaussian random fields, and $6, 3, 3$ and 1 for the non-Gaussian random fields. We only include the results from the converged replications in the figures.

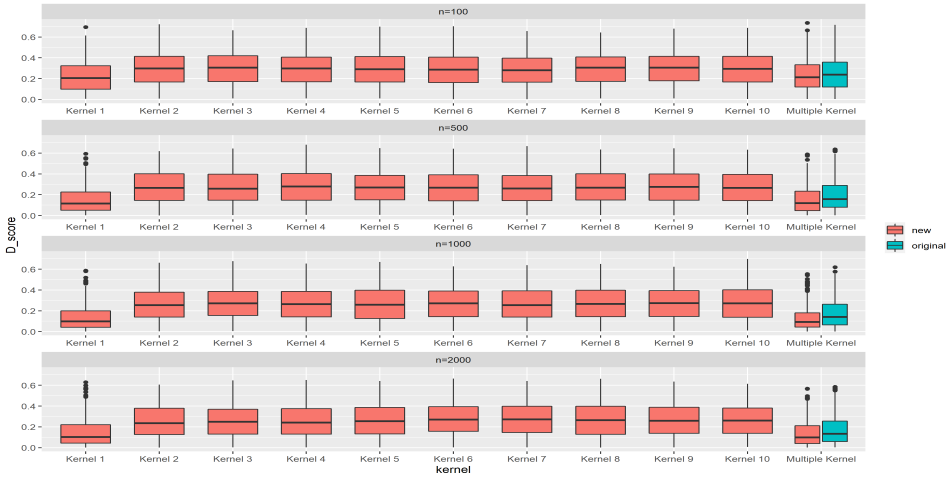

Figure 1: Boxplots of $D(\Omega, \hat{\Omega})$ for the proposed method using the 10 kernels (new) in (24), or each of those 10 kernels (Kernel 1, $\cdots$, Kernel 10), and the method of Bachoc et al. (2020) using the 10 kernels (original) in a simulation with 1000 replications for the Gaussian random fields. The number of observations n is 100, 500, 1000 or 2000 (from top to bottom), and the dimension of random fields is $p = 3$.

The estimated eigengaps for the proposed method for the Gaussian random fields are presented in Figures 5 and 6. As n increases, the eigengap also increases. Under low-dimensional setting $p = 3$, the estimates based on single kernel f_1 entail the largest eigengaps and the smallest estimation errors $D(\Omega, \hat{\Omega})$ (see also Theorem 2). However when $p = 50$, using the multiple kernels leads to the largest eigengaps and the smallest estimation errors. The patterns with the non-Gaussian random fields are similar and not reported here to save space.

Kernel	1	2	3	4	5	6	7	8	9	10	Multiple(new)	Multiple(original)
n=100	0.0814	0.2284	0.2707	0.2584	0.2594	0.2617	0.2542	0.2688	0.2619	0.2517	0.0933	0.1298
n=500	0.0248	0.1437	0.2019	0.2051	0.2042	0.1873	0.1830	0.1926	0.2076	0.2071	0.0327	0.0444
n=1000	0.0189	0.1124	0.1992	0.1782	0.1800	0.1746	0.1862	0.1803	0.1823	0.1887	0.0233	0.0324
n=2000	0.0164	0.1194	0.1870	0.1631	0.1686	0.1746	0.1533	0.1761	0.1701	0.1845	0.0204	0.0260

Table 1: Median of $D(\Omega, \hat{\Omega})$ from the proposed method using the 10 single kernels, or multiple kernel(including all 10 ring kernels), and the method of Bachol et el. using the multiple kernel (multiple original) in a simulation with 1000 replications for the Gaussian random fields. The number of observations n is 100, 500, 1000 or 2000 , and the dimension of random fields is $p = 3$.

Kernel	1	2	3	4	5	6	7	8	9	10	Multiple(new)	Multiple(original)
n=100	0.0837	0.2377	0.2656	0.2579	0.2676	0.2664	0.2478	0.2503	0.2440	0.2531	0.0915	0.1194
n=500	0.0244	0.1526	0.2028	0.2001	0.2052	0.1998	0.1923	0.1964	0.2028	0.2107	0.0284	0.0424
n=1000	0.0178	0.1096	0.1767	0.1943	0.1868	0.1812	0.1741	0.1627	0.1885	0.1911	0.0215	0.0326
n=2000	0.0165	0.1230	0.1907	0.1765	0.1676	0.1663	0.1652	0.1625	0.1742	0.1818	0.0194	0.0293

Table 2: Median of $D(\Omega, \hat{\Omega})$ from the proposed method using the 10 single kernels, or multiple kernel(including all 10 ring kernels), and the method of Bachol et el. using the multiple kernel (multiple original) in a simulation with 1000 replications for the non-Gaussian random fields. The number of observations n is 100, 500, 1000 or 2000 , and the dimension of random fields is $p = 3$.

4.2 A real data example

We apply the proposed method to the moss data from the Kola project in the R package `StatDa` (See Filzmoser (2015)). The data consists of chemical elements discovered in terrestrial moss at the 594 locations in northern Europe; see the map in Fig.D.1 of Bachoc et al. (2020). More information on the data is presented in Reimann et al. (2008). Following the lead of Nordhausen et al. (2015) and Bachoc et al. (2020), we apply the so-called isometric-log-ratio transformation to the 31 compositional chemical elements in the data. The transformed data are used in our analysis with $n = 594$ and $p = 30$. We standardize the data first such that the sample mean is 0 and the sample variance is I_{30}.

We apply the proposed estimation method with 10 kernels specified as in (24). The scores of the first six independent components (IC), corresponding to the six largest eigenvalues of $\hat{W}$ (see table 5), are plotted in Figure 7; showing some interesting spatial patterns. For example,

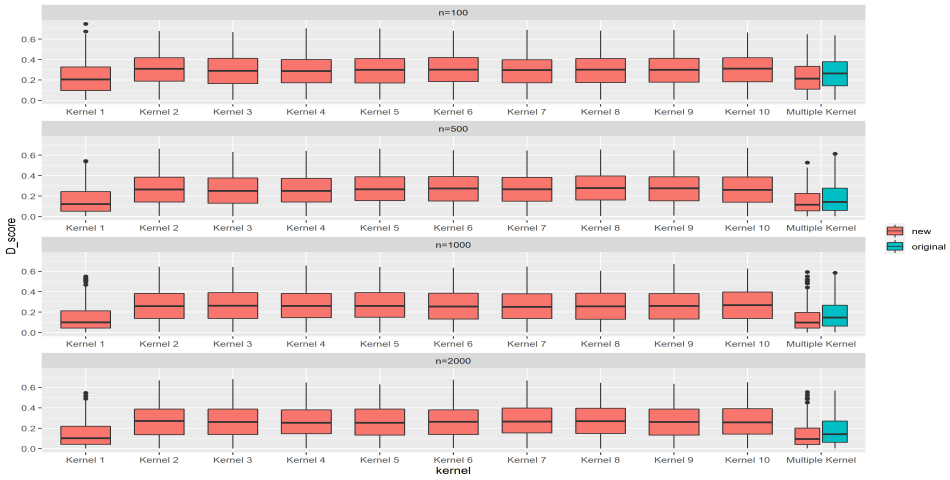

Figure 2: Boxplots of $D(\Omega, \hat{\Omega})$ for the proposed method using the 10 kernels (new) in (24), or each of those 10 kernels (Kernel 1, $\cdots$, Kernel 10), and the method of Bachoc et al. (2020) using the 10 kernels (original) in a simulation with 1000 replications for the non-Gaussian random fields. The number of observations n is 100, 500, 1000 or 2000 (from top to bottom), and the dimension of random fields is $p = 3$.

the 1st IC can be viewed as a contrast between the locations in the west and those in the east, and the 2nd IC is that between the north and the south. Figure 8 displays the absolute correlation coefficients between the first twelve ICs and those obtained in Nordhausen et al. (2015) which was referred as 'gold standard' by Bachoc et al. (2020). While the ICs derived from the two methods differ from each other, the two sets of ICs correlate with each other significantly. For example the correlation between the 1st IC derived from our new method and the 2nd IC obtained in Nordhausen et al. (2015) is 0.92. Note that the 'gold standard' estimation was obtained using the kernel specified with the relevant subject knowledge. In contrast our estimation is based on the multiple kernels defined generically in (24).

The six largest eigenvalues of $\hat{W}$ are listed in Table 5. The eigengaps $\Delta_i = \hat{\lambda}_{i-1} - \hat{\lambda}_i$ for $i = 7, \cdots, 30$ are plotted in Figure 9. It is clear that the eigengaps among the 13 largest eigenvalues are large. Based on Theorem 1, we have

$$\hat{\Omega}^{-1}\Omega = \hat{\Gamma}_\Omega = \begin{pmatrix} \hat{\Gamma}_{\Omega,aa} & \hat{\Gamma}_{\Omega,ab} \\ \hat{\Gamma}_{\Omega,ba} & \hat{\Gamma}_{\Omega,bb} \end{pmatrix}, \tag{25}$$

where $\hat{\Gamma}_{\Omega,aa}$ is a 12×12 matrix satisfying $\|\hat{\Gamma}_{\Omega,aa} - I_{12}\| = O_p(n^{-1/2}p^{1/2})$. Theorem 1 also shows

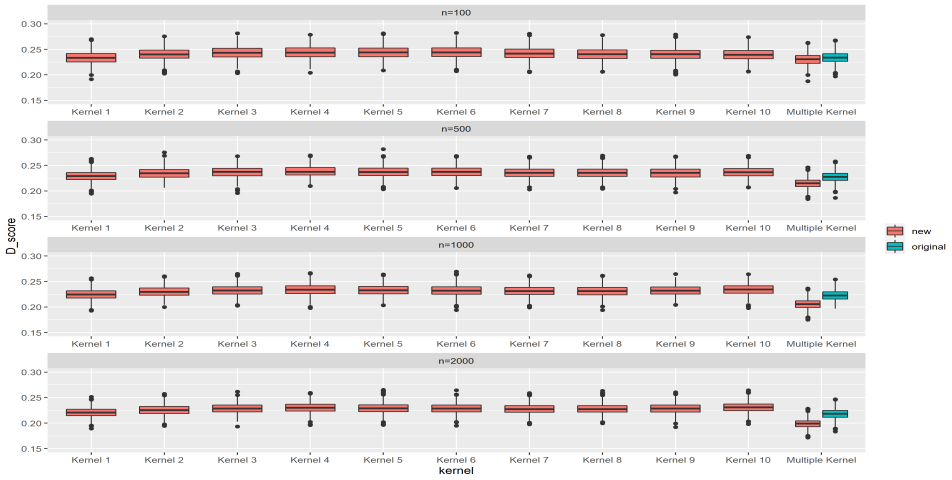

Figure 3: Boxplots of $D(\Omega, \hat{\Omega})$ for the proposed method using the 10 kernels (new) in (24), or each of those 10 kernels (Kernel 1, $\cdots$, Kernel 10), and the method of Bachoc et al. (2020) using the 10 kernels (original) in a simulation with 1000 replications for the Gaussian random fields. The number of observations n is 100, 500, 1000 or 2000 (from top to bottom), and the dimension of random fields is $p = 50$.

that $\|\hat{\Gamma}_{\Omega,ab}\| = O_p(n^{-1/2}p^{1/2})$, $\|\hat{\Gamma}_{\Omega,ba}\| = O_p(n^{-1/2}p^{1/2})$ and $\|\hat{\Gamma}_{\Omega,bb}\| = 1+O_p(n^{-1/2}p^{1/2})$. Thus, we are reasonably confident that the estimated first 12 ICs are reliable. Moreover, we rewrite $\hat{\Omega}^\top\hat{\Omega}$ as

$$\hat{U}_W^\top\hat{\Sigma}\hat{U}_W = \hat{\Omega}^\top\hat{\Omega} = \begin{pmatrix} \hat{\Omega}_{aa} & \hat{\Omega}_{ab} \\ \hat{\Omega}_{ba} & \hat{\Omega}_{bb} \end{pmatrix}, \tag{26}$$

where $\hat{\Omega}_{aa}$ is a 12×12 matrix. We gain $tr(\hat{\Omega}_{aa}) = 6.62$ and $tr(\hat{\Omega}^\top\hat{\Omega}) = 8.89$ by calculating. Thus, the major variation of the 30 variables are largely reflected by the 12 largest ICs.

Supplementary material

We provide the proofs of Theorems 1-2 in Supplementary material. We also provide an additional example for simulations in Supplementary material.

Kernel	1	2	3	4	5	6	7	8	9	10	Multiple(new)	Multiple(original)
n=100	0.2337	0.2404	0.2433	0.2438	0.2442	0.2442	0.2418	0.2408	0.2405	0.2394	0.2308	0.2339
n=500	0.2295	0.2348	0.2375	0.2378	0.2373	0.2377	0.2356	0.2356	0.2355	0.2369	0.2153	0.2276
n=1000	0.2247	0.2300	0.2326	0.2343	0.2331	0.2323	0.2313	0.2313	0.2321	0.2346	0.2059	0.2228
n=2000	0.2207	0.2254	0.2285	0.2303	0.2293	0.2288	0.2275	0.2275	0.2286	0.2310	0.1993	0.2184

Table 3: Median of $D(\Omega, \hat{\Omega})$ from the proposed method using the 10 single kernels, or multiple kernel(including all 10 ring kernels), and the method of Bachol et el. using the multiple kernel (multiple original) in a simulation with 1000 replications for the Gaussian random fields. The number of observations n is 100, 500, 1000 or 2000 , and the dimension of random fields is $p = 50$.

Kernel	1	2	3	4	5	6	7	8	9	10	Multiple(new)	Multiple(original)
n=100	0.2332	0.2391	0.2425	0.2429	0.2425	0.2425	0.2417	0.2393	0.2390	0.2383	0.2295	0.2336
n=500	0.2292	0.2331	0.2363	0.2374	0.2369	0.2369	0.2359	0.2348	0.2352	0.2372	0.2143	0.2278
n=1000	0.2250	0.2305	0.2324	0.2338	0.2338	0.2327	0.2317	0.2312	0.2328	0.2341	0.2059	0.2228
n=2000	0.2203	0.2249	0.2281	0.2296	0.2292	0.2281	0.2269	0.2277	0.2288	0.2303	0.1990	0.2172

Table 4: Median of $D(\Omega, \hat{\Omega})$ from the proposed method using the 10 single kernels, or multiple kernel(including all 10 ring kernels), and the method of Bachol et el. using the multiple kernel (multiple original) in a simulation with 1000 replications for the non-Gaussian random fields. The number of observations n is 100, 500, 1000 or 2000 , and the dimension of random fields is $p = 50$.

References

Bachoc, F. (2014). Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of multivariate analysis*, **125**, 1-35.

Bachoc, F., Genton, M.G., Nordhausen, K., Ruiz-Gazen, A. and Virta, J. (2020). Spatial blind source separation. *Biometrika*, **107**, 627-646.

Cardoso, J. (1998). Multidimensional independent component analysis. In *Proceedings of the 1998 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, **4**, 1941-1944.

Table 5: The six largest eigenvalues of $\hat{W}$ (with $k = 10$) for the real data example.

i	1	2	3	4	5	6
$\hat{\lambda}_i$	1136.50	877.59	444.21	161.34	126.16	81.13

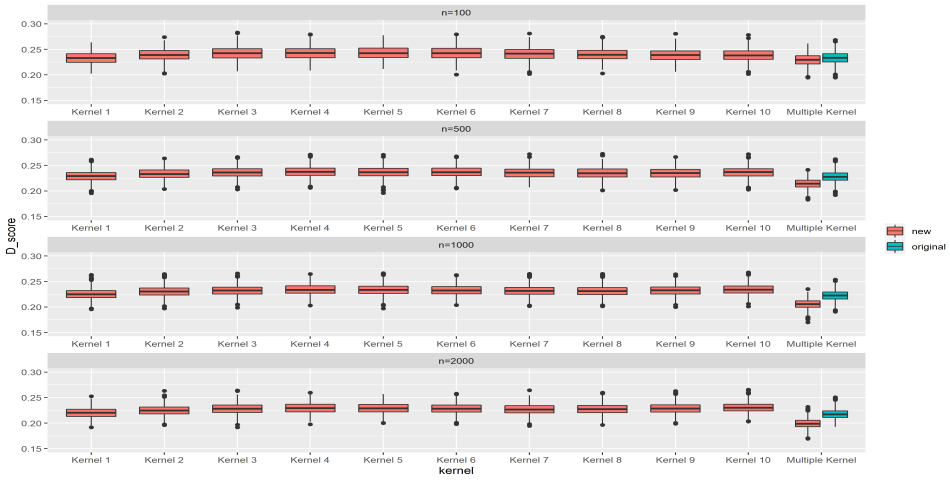

Figure 4: Boxplots of $D(\Omega, \hat{\Omega})$ for the proposed method using the 10 kernels (new) in (24), or each of those 10 kernels (Kernel 1, $\cdots$, Kernel 10), and the method of Bachoc et al. (2020) using the 10 kernels (original) in a simulation with 1000 replications for the non-Gaussian random fields. The number of observations n is 100, 500, 1000 or 2000 (from top to bottom), and the dimension of random fields is $p = 50$.

Chang, J., Guo, B. and Yao, Q. (2018). Principal component analysis for second-order stationary vector time series. *The Annals of Statistics*, **46**, 2094-2124.

Clarkson, D.B. (1988). Remark AS R71: A remark on algorithm AS 211. The F-G diagonalization algorithm. *Applied Statistics*, **37**, 147-151.

Ferreol, A., Albera, L. and Chevalier, P. (2005). Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM). *IEEE Transactions on Signa Processing*. **53**, 1640-1653.

Filzmoser, P. (2015). `StatDA`: Statistical Analysis for Environmental Data. R package version 1.6.9.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*. Wiley, New York.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, **40**, 694-726.

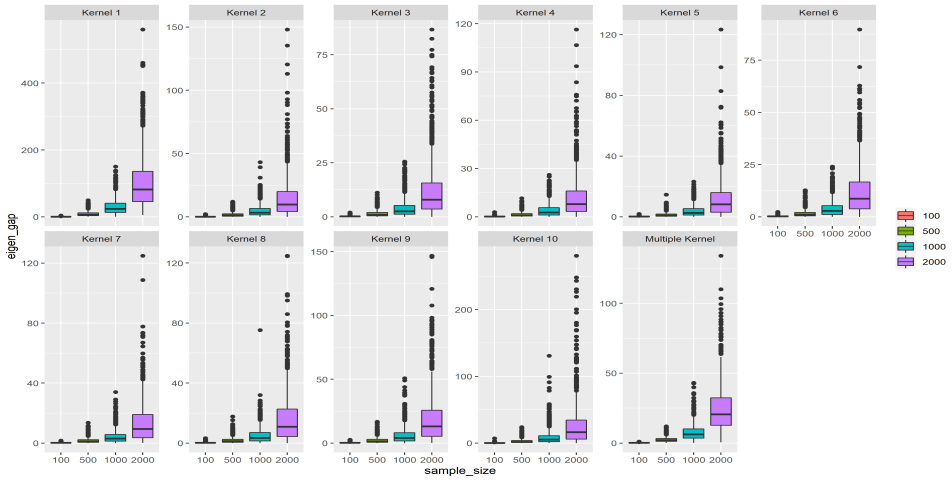

Figure 5: Boxplots of the estimated eigengaps of the proposed method using the 10 kernels (Multiple kernels) in (24), or each of those 10 kernels (Kernel 1, $\cdots$, Kernel 10) for the Gaussian random fields. Number of observations n is set at 100, 500, 1000 and 2000, the dimension of random fields is $p = 3$.

Nordhausen, K., Oja, H., Filzmose, P. and Reiman, C. (2015). Blind source separation for spatial compositional data. *Mathematical Geosciences*, **47**, 753-770.

Reimann, C., Filzmoser, P., Garrett, R. and Dutter, R. (2008). *Statistical Data Analysis Explained. Applied Environmental Statistics with R.* Wiley, Chicester.

Ziehe, A. and Müller, KR. (1998). TDSEP – an efficient algorithm for blind separation using time structure. In *ICANN 98: Perspectives in Neural Computing*. L. Niklasson, M. Boden and T. Ziemke (eds). Springer, London, pp.675-680.

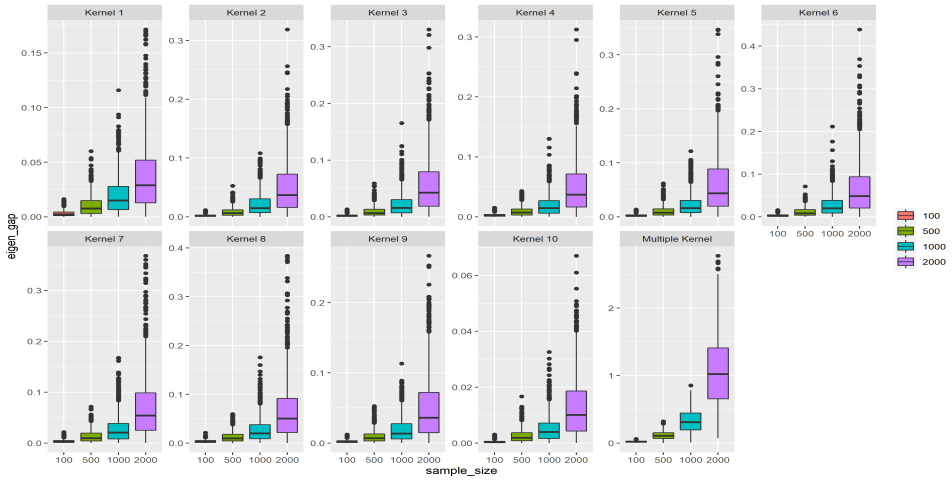

Figure 6: Boxplots of the estimated eigengaps of the proposed method using the 10 kernels (Multiple kernels) in (24), or each of those 10 kernels (Kernel 1, $\cdots$, Kernel 10) for the Gaussian random fields. Number of observations n is set at 100, 500, 1000 and 2000, the dimension of random fields is $p = 50$.

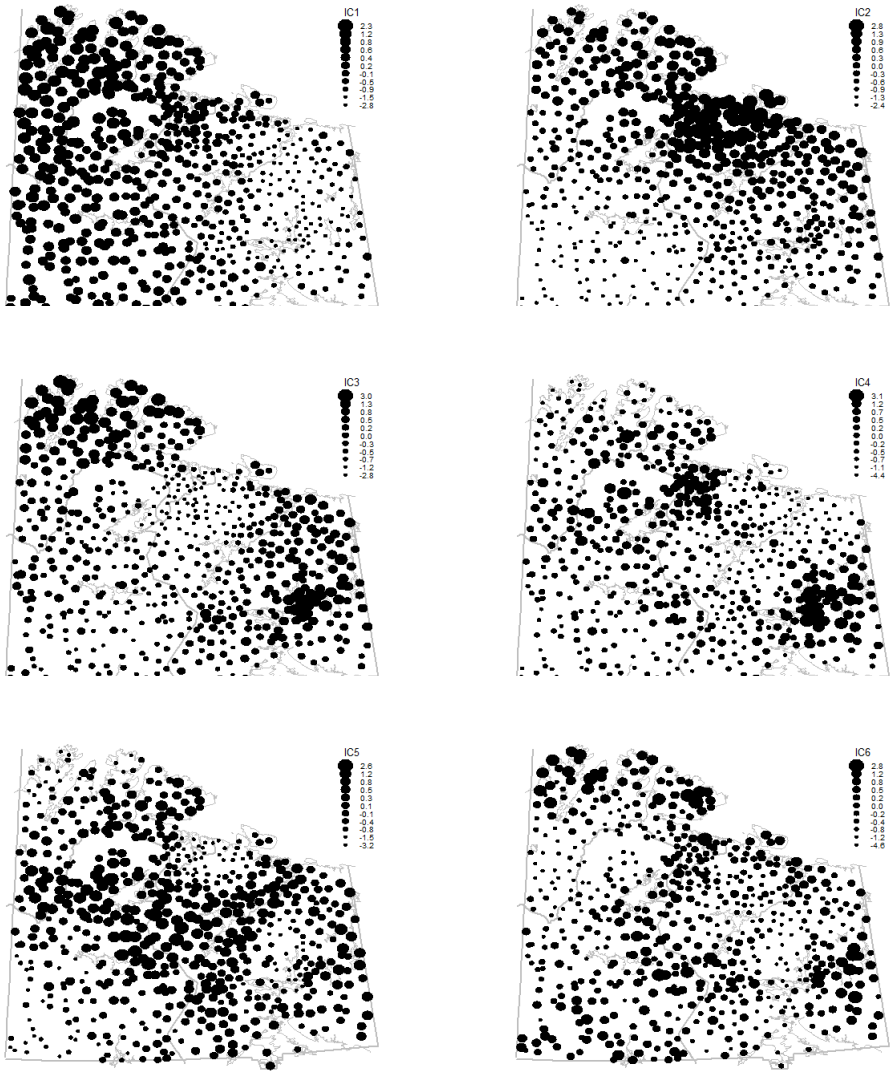

Figure 7: The scores of the first six independent components over the 594 observation locations.

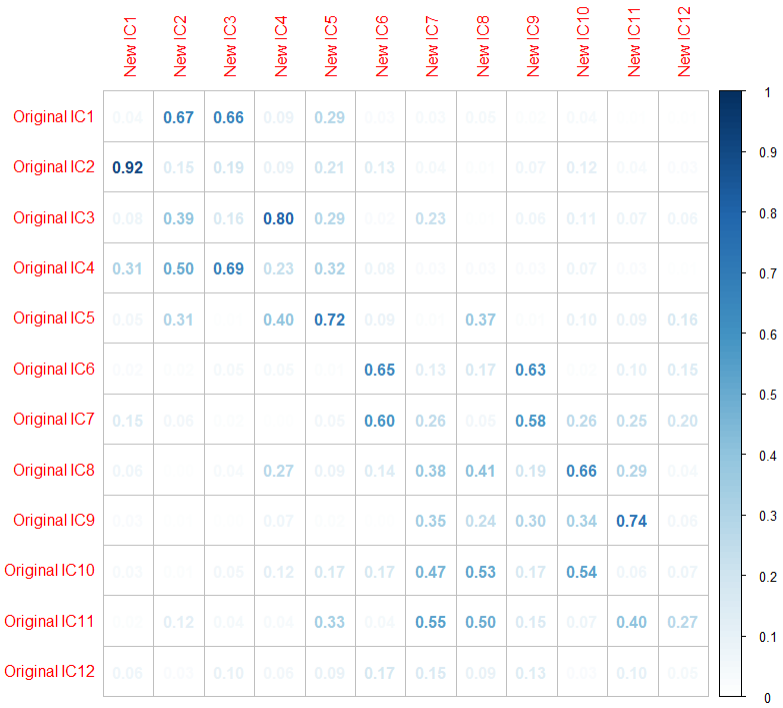

Figure 8: The absolute correlation coefficients between the first 12 independent components derived from the proposed method (New) and those obtained in Nordhausen et al. (2015) (Original).

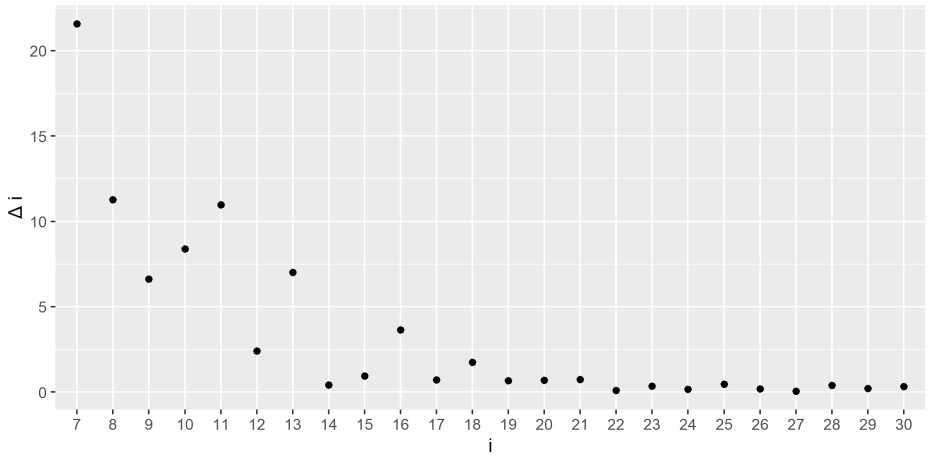

Figure 9: The estimated eigengaps $\Delta_i = \hat{\lambda}_{i-1} - \hat{\lambda}_i$ for $i = 7, \cdots, 30$ on real data example from proposed method with multiple kernel.