TECHNISCHE UNIVERSITÄT WIEN

DISSERTATION

Advances in blind source separation for spatial data

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften unter der Leitung von

Klaus Nordhausen
Institute of Statistics and Mathematical Methods in Economics, TU Wien
Department of Mathematics and Statistics, University of Jyväskylä

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Christoph Mühlmann
Matrikelnummer: 01230151

Diese Dissertation haben begutachtet:

1. **Dipl.-Stat. University Lecturer Klaus Nordhausen PhD**
   Institute of Statistics and Mathematical Methods in Economics, TU Wien
   Department of Mathematics and Statistics, University of Jyväskylä
2. **Prof. Anne Ruiz-Gazen PhD**
   TSE Research Faculty, University of Toulouse Capitole
3. **Prof. Dr. Siegfried Hörmann**
   Institut für Statistik, TU Graz

Wien, am 13. Oktober 2021

Abstract

Multivariate data where each measurement is taken at a different location in space are encountered in many practical applications. Typical features of such data are that measurements in close proximity tend to be more similar than the ones further separated. Proper statistical tools need to account for this fact. Recently, the popular blind source separation methodology was formulated for such spatial data denoted as spatial blind source separation (SBSS). Specifically, it is assumed that the observed data is formed by linear combinations of unobserved components. Recovering these latent components is very beneficial as they are uncorrelated, most often show underlying physical processes that generated the data and for domain experts it is highly likely that only a few components are of interest. All advantages were proven on a geochemical dataset. However, the original SBSS only considers the case when all components are of equal interest and are invariant under translation in space (second-order stationarity). This thesis extends these methods for various features commonly present in spatial data. Namely, first and second order non-stationarity or noisy spatial data. Furthermore, the usefulness of blind source separation in spatial regression as well as spatial prediction is investigated.

Kurzfassung

Viele Datensätze bestehen aus multivariaten Messungen, die an verschiedenen geographischen Orten durchgeführt wurden. Typischerweise besitzen solche Datensätze die Eigenschaft, dass Messungen in unmittelbarer Nähe ähnlicher sind als Messungen, die eine große Entfernung aufweisen. In der statistischen Analyse solcher räumlichen Daten sollte diese spezielle Eigenschaft berücksichtigt werden. In letzter Zeit wurde in der statistischen Literatur die sogenannte Blind Source Separation Methode auf räumliche Daten erweitert. In diesem Model wird angenommen, dass die Daten aus Linearkombinationen von unbeobachteten Variablen bestehen, und das Ziel ist diese latenten Variablen zu bestimmen. Die weitere Analyse der Daten kann nun mit Hilfe dieser unbeobachteten Variablen durchgeführt werden. Dies bietet einige Vorteile: die latenten Variablen sind unkorreliert und zeigen oft physikalische Prozesse, welche die Daten generieren und in den meisten Fällen sind nicht alle latenten Variablen von Interesse. All diese Vorteile wurden anhand eines geochemischen Datensatzes eindrucksvoll nachgewiesen. Die originale räumliche Blind Source Separation Methode ist nur für latente Komponenten formuliert, die invariant unter räumlicher Translation (schwach stationär) und von gleichem Interesse sind. In dieser Dissertation wird die räumliche Blind Source Separation Methode für weitere Eigenschaften von räumlichen Daten adaptiert. Der Fokus liegt auf in erster oder zweiter Ordnung nicht stationären Daten und Daten die von weißem Rauschen beeinflusst sind. Weiters zeigt diese Arbeit den Nutzen der räumlichen Blind Source Separation Methode für räumliche Prognosen und Regression.

Acknowledgement

I want to sincerely thank my supervisor Klaus Nordhausen for giving me the chance to do a doctoral thesis, his support, guidance and teaching me the value of independent working. I wish to thank all my coauthors and project members: Markus Bögl, Kamila Fačevicová, Peter Filzmoser, Theresia Gschwandtner, Silvia Miksch, Hannu Oja, Nikolaus Piccolotto, Joni Virta, Mengxi Yi, and special thanks to François Bachoc who hosted my Toulouse visit. During my dissertation I had the honor to co-supervise two students, thank you Krisztián Harmat and Luzia Jorda for the nice work together. Thank you Anne Ruiz-Gazen and Siegfried Hörman for the willingness to act as referees for my thesis. As a part of the CSTAT group I want to thank all members of the group as well. Furthermore, I acknowledge the Austrian Science Fund which funded my position as a part of the P31881-N32 project.

I want to deeply thank my mother, stepfather, sister, grandparents and all other members of my family. They always supported my work and had time when I needed to talk and support. May my father rest in peace.

Surely, I want to thank my band and the TMK Judendorf which always warmly welcomed me when I had the chance to visit them. And lastly I want to thank all my friends and colleagues who supported me, had an open ear and went some parts of the way with me. Especially: Carina, Christopher, Domi, Herr von Tonsern, Kamilka, Liesl, Martin, Mathias, Philip, Tanja.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 13. Oktober 2021

Christoph Mühlmann

Original publications

This thesis consists of an introductory part and the following publications.

I C. Muehlmann, F. Bachoc, K. Nordhausen, and M. Yi. Test of the latent dimension of a spatial blind source separation model. Submitted. Preprint at arXiv:2011.01711, 2020.

II C. Muehlmann, P. Filzmoser, and K. Nordhausen. Local difference matrices for spatial blind source separation. To appear in Proceedings of the 3rd Conference of the Arabian Journal of Geosciences, 2020.

III C. Muehlmann, P. Filzmoser, and K. Nordhausen. Spatial blind source separation in the presence of a drift. Submitted. Preprint at arXiv:2108.13813, 2021.

IV C. Muehlmann, F. Bachoc, and K. Nordhausen. Blind source separation for non-stationary random fields. Submitted. Preprint at arXiv:2107.01916, 2021.

V C. Muehlmann, K. Nordhausen, and M. Yi. On Cokriging, neural networks, and spatial blind source separation for multivariate spatial prediction. *IEEE Geoscience and Remote Sensing Letters*, pages 1–5, 2020. doi: 10.1109/LGRS.2020.3011549.

VI C. Muehlmann, H. Oja, and K. Nordhausen. Sliced inverse regression for spatial data. In E. Bura and B. Li, editors, *Festschrift in Honor of R. Dennis Cook: Fifty Years of Contribution to Statistical Science*, pages 87–107. Springer, Cham, 2021. doi: 10.1007/978-3-030-69009-0_5.

Contents

Part I.

Summary

1. Introduction

Datasets consisting of measurements that are taken at different locations in space are a popular form of data. Mining engineers take drill wholes across a mine to determine sweet spots for mining, climate scientist measure concentrations of different pollutants at different roads in a city or geochemists analyze soil samples for detecting different geological formations. All of these examples share the property that most insights of the data can be gained when considering dependence in and in between measurements as a function of spatial separation. This leads to the fact that the usefulness of classical statistical tools designed for cross-sectional data is rather limited. Often scientist are interested in finding linear combinations of the spatial dataset at hand which are easy to interpret and might show physical processes that drive the data generation. Moreover, the best scenario is achieved when only a small number of such linear combinations needs to be considered which can be viewed as a form of dimension reduction.

Interest in finding the underlying structure of the data is generally given in statistical literature. E.g.: the well-known principal component analysis finds transformation of the data that maximize variance which might lead to more meaningful components. Another strategy is given by the methodology of blind source separation (BSS). The original motivation of BSS is to solve the so-called cocktail party problem. Several people are inside a room talking to each other and the corresponding sound signal is measured with some microphones that are placed all over the room. The goal is to transform the recorded signal in such a way that each voice can be recovered individually. Furthermore, to keep this example, not all of the found conversations might be of interest. This practical example translates to the motivations of BSS. Namely, it aims to find physical signals that led to the observed signal by assuming that the way of mixing these physical signals is linear. BSS has been used with great success for many forms of data, e.g.: cross-sectional data leading to independent component analysis, various time series data, tensorial data, etc. Or more specifically: separating the heart beat signal from a fetus and its mother based on the mothers electrocardiogram (ECG) signal or it is successfully used to extract neural signals of the brain from electroencephalography (EEG) data to name some examples.

Clearly, the interest of geostatisticians to explore a spatial dataset by finding a small number of meaningful linear combinations that are easy to interpret coincides with the motivations of the BSS methodology. This is also pointed out in more detail in the recent literature (Nordhausen et al., 2015; Bachoc et al., 2020) which extends successfully BSS for stationary multivariate spatial data. The findings of the former two publications act as a starting point for the present thesis where the goal is to refine the recently introduced spatial BSS methodology to account for various features commonly present in spatial data. Namely, extensions for first or second order spatial dependence that vary in space (non-stationary spatial data), determining the actual signals of the found components, dimension reduction in spatial regression and the usefulness of this methodology in spatial prediction. In total the present thesis consists of six scientific publications as well as software for the various introduced methods.

This summary part consists of four sections that follow this introductory part. Section 2

introduces the commonly considered statistical framework of spatial data where the emphasis is laid on characterizing second order spatial dependence and (linear) spatial prediction. Similarly, Section 3 reviews the BSS methodology for independently and identically distributed data as well as time series data by stating various models and estimators and additionally considering the case when noise is present in the data. The summary part is concluded by Section 4 which summarizes the main scientific contributions of the six scientific publications of this thesis and Section 5 that outlines directions for future research. Software for almost all methods of Section 4 is provided for the statistical computing software R (Team, 2021) by the package SpatialBSS (Muehlmann et al., 2021d).

2. Geostatistics

Geostatistics is concerned with the analysis of data where each measurement is taken at a different geographical location. For example geochemists take soil samples and measure concentrations of typical elements (Reimann et al., 2010, 2014) or climate scientist measure certain climate indicators such as soil moisture via remote sensing (Bauer-Marschallinger et al., 2013). All the beforehand datasets share the property that measurements are separated in space and therefore it is highly likely that nearby measurements are more related than the ones further apart. Moreover, often several quantities are measured at each sample location which leads to spatial dependencies also in-between variables. The goal of geostatistics is to provide a statistical framework that specifically accounts for this features of such multivariate spatial data. Great introductions are given in textbooks such as Cressie (1993); Wackernagel (2003); Chilés and Delfiner (2012).

The following sections aim to state the main principles of geostatistics where Section 2.1 introduces the notion of a random field and introduces stationarity concepts. The main tool of characterizing (second order) spatial dependence is the spatial covariance where properties and approaches of modeling this quantity are given in Section 2.2. Lastly, Section 2.3 is devoted to the task of predicting a value of the data at hand on an unobserved location where the focus lies on the well-known Kriging methodology. These three sections are mainly based on the discussions provided by Wackernagel (2003); Chilés and Delfiner (2012); Genton and Kleiber (2015).

2.1. Random fields and stationarity concepts

One central concept in the statistical analysis of spatial data is the assumption that a stochastic process generated the data at hand and the theoretical analysis is carried out by studying this stochastic process. Such a data generating process (DGP) is denoted as random field and defined as follows.

Definition 1 (Random field)**.** *A p-variate random field* $(\mathbf{x}(\mathbf{s}))_{\mathbf{s}\in\mathcal{S}}$ *(or equivalently* $\mathbf{x}(\mathbf{s})$*) is a family of p-variate random vectors indexed by the spatial domain* $\mathcal{S} \subseteq \mathbb{R}^d$ *which are defined on a probability space* $(\Omega, \mathcal{A}, P)$*.*

The spatial domain $\mathcal{S}$ contains all the possible locations where the random field is defined. Such locations might be pairs of longitude and latitude values or simply x and y values in some length unit. Often the dimension of the domain equals $d = 1, 2, 3$. In practical applications a finite set of sample locations $\mathcal{C} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathcal{S}$ is given and the geostatistician is concerned with a wide variety of tasks (e.g.: inference about parameters for spatial models, prediction of measurements at unobserved locations, etc.) which are based only on one realization of the random field on $\mathcal{C}$. Figure 2.1 depicts two different sets of sample locations for the case of $d = 2$.

In principle all the dependencies and the probabilistic behavior of the random field are given by its joint cumulative distribution function (cdf). However, exact modeling of the cdf is

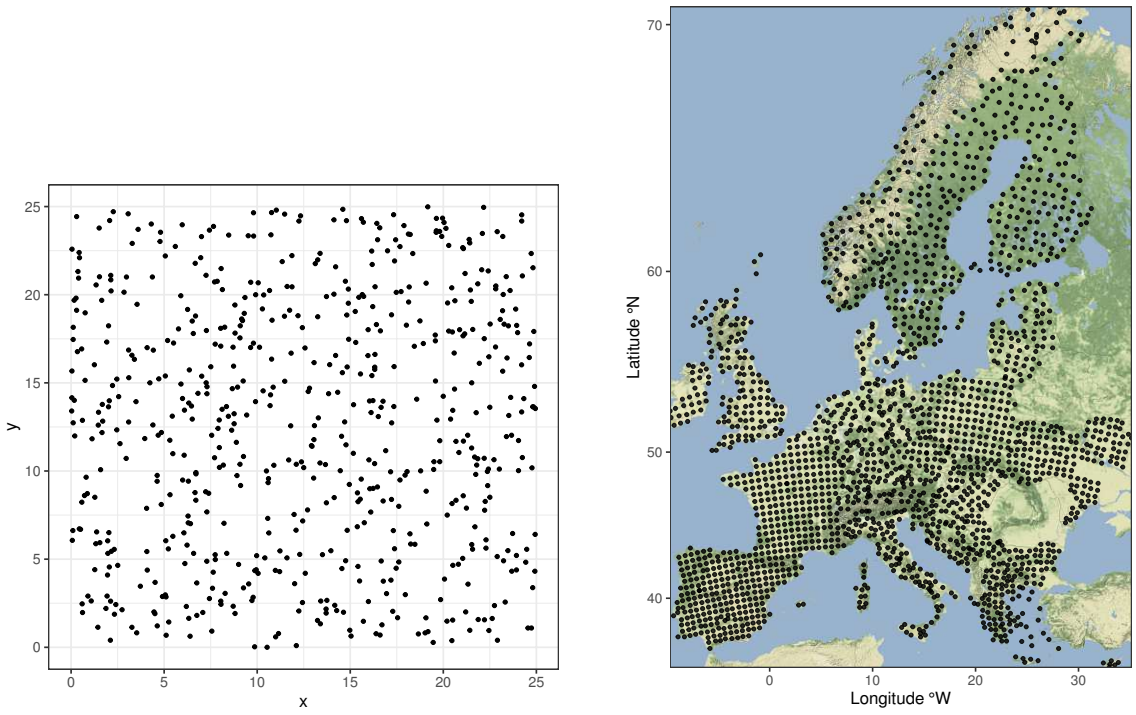

Figure 2.1.: Left panel: 625 uniformly sampled locations on a domain of $\mathcal{S} = [0, 25] \times [0, 25]$. Right panel: Sample locations for the GEMAS dataset (Reimann et al., 2014). Both figures are examples for two-dimensional domains ($d = 2$). Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

indefeasible as moments of any order need to be modeled. To overcome this issue it is common to only consider the first and second order spatial dependence structure (this procedure is often denoted as structural analysis) of the random field at hand. The first order dependence is given by the mean function (often denoted as drift function)

$$\begin{aligned} \mathbf{m}: \ & \mathcal{S} \to \mathbb{R}^p \\ & \mathbf{s} \mapsto \mathbb{E}\left(\mathbf{x}(\mathbf{s})\right), \end{aligned}$$

and the second order dependence is given by the (spatial) covariance function

$$\begin{aligned} \mathbf{C}: \ & \mathcal{S} \times \mathcal{S} \to \mathbb{R}^{p \times p} \\ & (\mathbf{s}, \mathbf{s}') \mapsto \mathrm{Cov}(\mathbf{x}(\mathbf{s}), \mathbf{x}(\mathbf{s}')). \end{aligned}$$

For the special case of a Gaussian random field the former two quantities specify the model completely as the Gaussian distribution is completely characterized by its first two moments. Formally, a Gaussian random field is defined as follows.

Definition 2 (Gaussian random field)**.** *A p-variate random field $(\mathbf{x}(\mathbf{s}))_{\mathbf{s}\in\mathcal{S}}$ is a Gaussian random field if for any finite set of sample locations $\mathcal{C} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathcal{S}$ the random vector $\left(\mathbf{x}(\mathbf{s}_1)^\top, \ldots, \mathbf{x}(\mathbf{s}_n)^\top\right)^\top$ is multivariate Gaussian distributed.*

Generally, a proper spatial covariance matrix functional $\mathbf{C}$ needs to possess the property that for any given set of n sample locations $\mathcal{C} \subset \mathcal{S}$ the $pn \times pn$ covariance matrix of the joint vector $(\mathbf{x}(\mathbf{s}_1)^\top, \ldots, \mathbf{x}(\mathbf{s}_n)^\top)^\top$ written as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{C}(\mathbf{s}_1, \mathbf{s}_1) & \mathbf{C}(\mathbf{s}_1, \mathbf{s}_2) & \ldots & \mathbf{C}(\mathbf{s}_1, \mathbf{s}_n) \\ \mathbf{C}(\mathbf{s}_2, \mathbf{s}_1) & \mathbf{C}(\mathbf{s}_2, \mathbf{s}_2) & \ldots & \mathbf{C}(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}(\mathbf{s}_n, \mathbf{s}_1) & \mathbf{C}(\mathbf{s}_n, \mathbf{s}_2) & \ldots & \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n) \end{pmatrix} \tag{2.1}$$

needs to be positive semi-definite (if it exists). Hence, for any vector $\mathbf{a} \in \mathbb{R}^{pn}$ it must hold that $\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} \geq 0$. Other properties and a thorough discussion about the covariance matrix functional is given in Chapter 2.2. Modeling of proper covariance matrices is one of the central tasks for spatial data analysis, it is unfortunately also a very challenging task as pointed out by Genton and Kleiber (2015). The covariance function of a random field depends specifically on the two location arguments, which allows a wide variety of second order dependencies across the whole spatial domain. Thus, a severe simplification which might be justified in practical applications and which acts as a starting point for more complex modeling, is the concept of second order stationarity (also denoted as weak stationarity or simply stationarity). Second order stationary states that the second order dependence is invariant under shifts and is formally defined as follows.

Definition 3 (Second-order stationary random field)**.** *A p-variate random field $\mathbf{x}(\mathbf{s})$ is said to be second order stationary if it fulfills the following conditions.*

- $\mathbb{E}\left(\|\mathbf{x}(\mathbf{s})\|^2\right) < \infty$ *for all* $\mathbf{s} \in \mathcal{S}$,
- $\mathbb{E}\left(\mathbf{x}(\mathbf{s})\right) = \mathbf{m}$ *for all* $\mathbf{s} \in \mathcal{S}$ *and*
- $\mathrm{Cov}(\mathbf{x}(\mathbf{s}), \mathbf{x}(\mathbf{s}')) = \mathrm{Cov}(\mathbf{x}(\mathbf{s}+\mathbf{s}''), \mathbf{x}(\mathbf{s}'+\mathbf{s}''))$ *for all* $\mathbf{s}, \mathbf{s}+\mathbf{s}'', \mathbf{s}', \mathbf{s}'+\mathbf{s}'' \in \mathcal{S}$.

The three conditions above state that all random vectors $\mathbf{x}(\mathbf{s})$ need to be square integrable, the drift is constant throughout the domain and the covariance function is only dependent on the difference between locations. The latter condition can be re-formulated by setting $\mathbf{s}'' = -\mathbf{s}$ which results in $\mathrm{Cov}(\mathbf{x}(\mathbf{0}), \mathbf{x}(\mathbf{s}-\mathbf{s}'))$. In this view the covariance depends only on the distance between sample locations $\mathbf{h} = \mathbf{s} - \mathbf{s}'$ leading to $\mathbf{C}(\mathbf{s}, \mathbf{s}') = \mathbf{C}(\mathbf{h})$. If additionally the covariance is independent of the direction of the spatial lag vector $\mathbf{h}$, then the random field is said to be isotropic and the covariance function reduces to a function with a scalar argument $\mathbf{C}(h)$ where $h = \|\mathbf{h}\|$. The simplest example for a stationary process is a white noise process that is defined as follows.

Definition 4 (White noise)**.** *A p-variate random field defined on a spatial domain $\mathcal{S}$ is denoted white noise if it satisfies*

- $\mathbb{E}\left(\|\mathbf{x}(\mathbf{s})\|^2\right) < \infty$ *for all* $\mathbf{s} \in \mathcal{S}$,
- $\mathbb{E}\left(\mathbf{x}(\mathbf{s})\right) = \mathbf{0}$ *for all* $\mathbf{s} \in \mathcal{S}$,
- $\mathbb{E}\left(\mathbf{x}(\mathbf{s})\mathbf{x}(\mathbf{s})^\top\right) = \boldsymbol{\Sigma}$ *for all* $\mathbf{s} \in \mathcal{S}$ *where* $\boldsymbol{\Sigma}$ *is a positive semi-definite matrix and*
- $\mathbb{E}\left(\mathbf{x}(\mathbf{s})\mathbf{x}(\mathbf{s}')^\top\right) = \mathbf{0}_p$ *for all* $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$ *and* $\mathbf{s} \neq \mathbf{s}'$ *where* $\mathbf{0}_p$ *is the* $p \times p$ *matrix containing only zeros.*

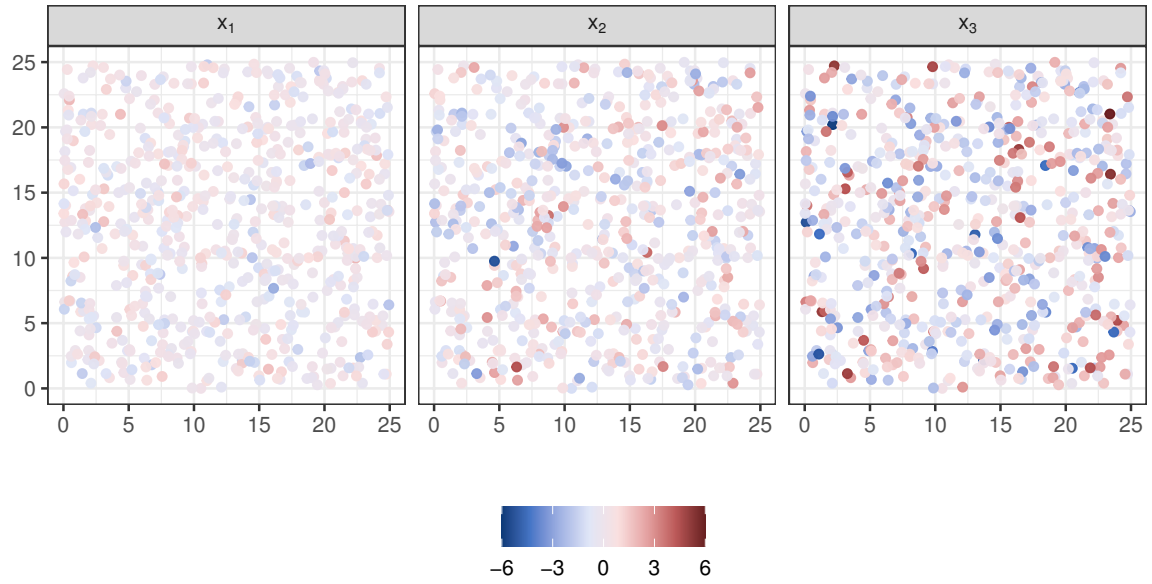

Figure 2.2.: Example of a three-variate Gaussian white noise process with a diagonal covariance matrix. The diagonal elements equal $0.5, 1.5$ and 3. The considered domain and the set of sample locations equal the ones from the left panel of Figure 2.1.

From this perspective white noise is a random field that does not exhibit second order spatial dependence as the spatial covariance matrix is the zero matrix for $\mathbf{h} \neq \mathbf{0}$ and any covariance matrix for $\mathbf{h} = \mathbf{0}$. The first and second property state that white noise is square integrable and zero mean, which reflects its weak stationarity property. Figure 2.2 depicts an example of a white noise process. Often a white noise random field is an additive term statistically independent from other summands describing a measurement error of some quantity of interest. The stationary and isotropic covariance matrix of white noise is given by

$$\mathbf{C}^{WN}(\mathbf{h}) = \mathbf{\Sigma}^{WN} I(\mathbf{h} = \mathbf{0}),$$

which is in geostatistical terms often denoted as nugget effect, here $\mathbf{\Sigma}^{WN}$ is a positive definite $p \times p$ matrix.

A more general class of random fields is given by the intrinsic stationarity concept which is based on working with difference processes rather than the original one.

Definition 5 (Intrinsic stationary random field)**.** *A p-variate random field $\mathbf{x}(\mathbf{s})$ is said to be intrinsic stationary if for every lag vector $\mathbf{h}$ the corresponding difference process $\mathbf{x}(\mathbf{s}+\mathbf{h}) - \mathbf{x}(\mathbf{s})$ is second order stationary.*

An instrinsic stationary random field is characterized by the drift function of the differences $\mathbf{m}'(\mathbf{h}) = \mathbb{E}\left(\mathbf{x}(\mathbf{s}+\mathbf{h}) - \mathbf{x}(\mathbf{s})\right)$ which is linear in its argument, i.e.: $\mathbf{m}'(\mathbf{h}) = \mathbf{a}^\top \mathbf{h}$ where $\mathbf{a} \in \mathbb{R}^p$ is an arbitrary vector. This is seen by

$$\mathbf{x}(\mathbf{s}+\mathbf{h}+\mathbf{h}') - \mathbf{x}(\mathbf{s}) = \left(\mathbf{x}(\mathbf{s}+\mathbf{h}) - \mathbf{x}(\mathbf{s})\right) + \left(\mathbf{x}(\mathbf{s}+\mathbf{h}+\mathbf{h}') - \mathbf{x}(\mathbf{s}+\mathbf{h})\right)$$

which leads to

$$\mathbf{m}'(\mathbf{h}+\mathbf{h}') = \mathbf{m}'(\mathbf{h}) + \mathbf{m}'(\mathbf{h}').$$

Often, $\mathbf{m}'(\mathbf{h})$ is simply set to the zero-vector. Furthermore, the natural measure of second order dependence for an instrinsic random field is given by the so called semivariogram

$$\begin{aligned}\boldsymbol{\gamma}: \ \mathcal{S} \times \mathcal{S} &\to \mathbb{R}^{p\times p}\\ (\mathbf{s}, \mathbf{s}') &\mapsto \frac{1}{2}\operatorname{Cov}(\mathbf{x}(\mathbf{s}) - \mathbf{x}(\mathbf{s}')),\end{aligned}$$

which only depends on the lag vector $\mathbf{h}$. Namely, $\boldsymbol{\gamma}(\mathbf{h}) = 1/2\operatorname{Cov}(\mathbf{x}(\mathbf{s}+\mathbf{h}) - \mathbf{x}(\mathbf{s}))$. Note that the variogram equals twice times the semivariogram. Sometimes it is assumed that the cross increments $x_i(\mathbf{s}) - x_j(\mathbf{s}')$ are stationary which leads to the so-called pseudo variogram

$$\begin{aligned}\boldsymbol{\pi}: \ \mathcal{S} \times \mathcal{S} &\to \mathbb{R}^{p\times p}\\ (\mathbf{s}, \mathbf{s}') &\mapsto \frac{1}{2}\left(\operatorname{Var}(x_i(\mathbf{s}) - x_j(\mathbf{s}'))\right)_{i,j=1,\dots,p}.\end{aligned}$$

Genton and Kleiber (2015) argue that the pseudo variogram is favorable as it has advantage when used for multivariate spatial prediction. In contrast, Wackernagel (2003, Chapter 21) states that the stationarity of the cross-increments is an unrealistic assumption and that often the entries of a random field measure different physical quantities which leads to the fact that differences make no sense.

As outlined in Chilés and Delfiner (2012, Chapter 2) for the univariate case the variogram has two benefits over the covariance function. Firstly, as it relies only on the difference process; the mean of the random field does not need to be estimated which is a practical argument. Secondly, as every weakly stationary field is also intrinsic stationary but the opposite does not hold true it follows that the class of intrinsic stationary random fields is richer as the weakly stationary one. Therefore, more random fields are covered by the intrinsic stationary class which is a theoretical argument. In contrast, for the multivariate case Genton and Kleiber (2015) argue that interpretation of the variogram is more difficult contrarily to the covariance. Furthermore, for an intrinsic stationary random field that is also weakly stationary it holds that

$$\boldsymbol{\gamma}(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \frac{1}{2}\left(\mathbf{C}(-\mathbf{h}) + \mathbf{C}(\mathbf{h})\right).$$

Thus, the variogram is an even function in its argument and does not cover the odd term of the covariance (Wackernagel, 2003, Chapter 21). Therefore, in the multivariate case the covariance might be favored. Generally, it seems that most authors favor the variogram for the univariate case but the covariance for the multivariate case.

One example of an intrinsic stationary field that is not weakly stationary is given by the fractional Brownian field (Dobrić and Ojeda, 2006). Its covariance function is defined by

$$C(\mathbf{s}, \mathbf{s}'; H) = \frac{1}{2}\left(\|\mathbf{s}\|^{2H} + \|\mathbf{s}'\|^{2H} - \|\mathbf{s} - \mathbf{s}'\|^{2H}\right).$$

Here, $H \in (0, 1]$ is denoted as Hurst parameter. Realizations of such a random field for different Hurst parameters can be seen in Figure 2.3.

Finally, a random field where the drift and/or the covariance function specifically depends on the location is denoted as non-stationary and acts as the opposite of weak stationarity.

Definition 6 (Non-stationary random field)**.** *A p-variate random field* $\mathbf{x}(\mathbf{s})$ *is said to be non-stationary if the mean and/or the covariance function specifically depends on the spatial location.*

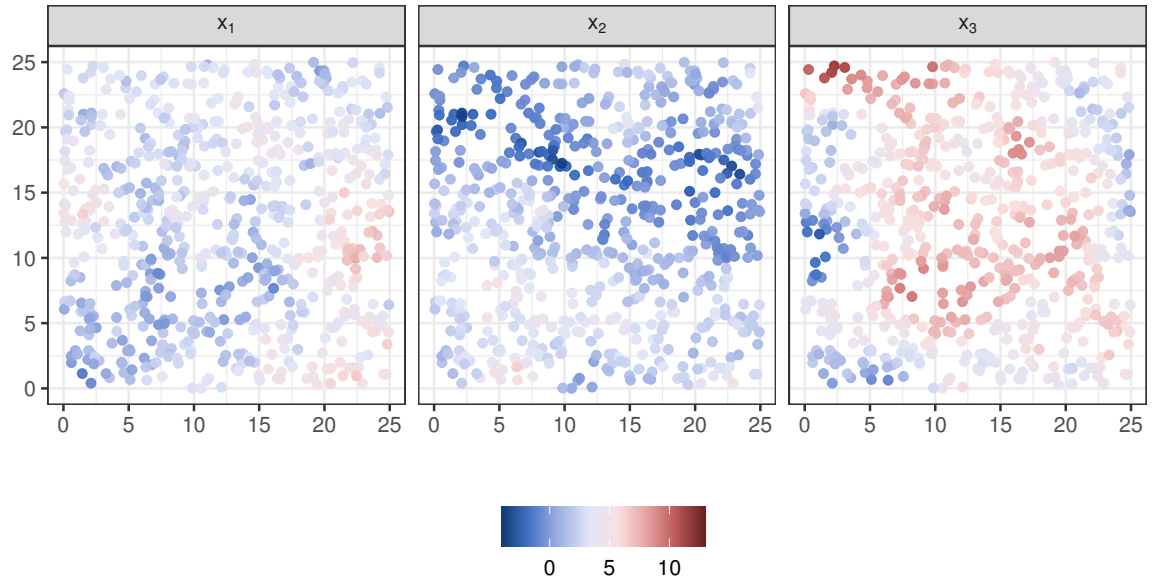

Figure 2.3.: Three examples of a Gaussian fractional Brownian field where the Hurst parameters equal $H = 0.3, 0.5$ and 0.8. The considered domain and the set of sample locations equal the ones from the left panel of Figure 2.1.

2.2. Covariance models

By the positive definite condition of proper covariance functionals it follows that $\mathbf{\Sigma}$ of Equation (2.1) needs to be symmetric. Therefore it holds that

$$\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{C}(\mathbf{s}_2, \mathbf{s}_1)^\top.$$

But if only the coordinates are switched or the variables are interchanged the covariance generally also changes

$$\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2) \neq \mathbf{C}(\mathbf{s}_2, \mathbf{s}_1) \quad \text{and} \quad \mathbf{C}(\mathbf{s}_1, \mathbf{s}_2) \neq \mathbf{C}(\mathbf{s}_1, \mathbf{s}_2)^\top.$$

Therefore, the covariance between different locations is not symmetric in general but the on-site covariance is. Furthermore, for a vector $\mathbf{a} \in \mathbb{R}^{2p}$ (considering only two coordinates $\mathbf{s}_1, \mathbf{s}_2$) where the i-th element is set to one and the $j+p$ element is set to $(\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2))_{ij}/(\mathbf{C}(\mathbf{s}_2, \mathbf{s}_2))_{jj}$ for $i, j \in \{1, \ldots, p\}$ the quadratic form $\mathbf{a}^\top \mathbf{\Sigma} \mathbf{a} \geq 0$ leads to

$$(\mathbf{C}(\mathbf{s}_1, \mathbf{s}_1))_{ii} (\mathbf{C}(\mathbf{s}_2, \mathbf{s}_2))_{jj} \geq |(\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2))_{ij}|^2.$$

Under stationary this translates to

$$(\mathbf{C}(\mathbf{0}))_{ii} (\mathbf{C}(\mathbf{0}))_{jj} \geq |(\mathbf{C}(\mathbf{h}))_{ij}|^2.$$

Therefore, the covariance matrices for each entry of the random field are bounded by the on-site variance, i.e.: $(\mathbf{C}(\mathbf{0}))_{ii} \geq |(\mathbf{C}(\mathbf{h}))_{ii}|$. In contrast, no such bounds exist for the cross covariances, i.e.: $|(\mathbf{C}(\mathbf{h}))_{ij}|$ is not bounded by $|(\mathbf{C}(\mathbf{0}))_{ij}|$ which is often referred to delay effect.

This is very common in time series where the highest influence of one variable to another might arises delayed in time.

Estimation of the covariance matrix for a given dataset with n irregularly spaced sample locations collected in the set $\mathcal{C}$ is usually done by the following estimator (often denoted as empirical covariogram or empirical covariance)

$$\hat{\mathbf{C}}(\mathbf{h}) = |N(\mathbf{h})|^{-1} \sum_{(i,j)\in\mathcal{N}(\mathbf{h})} (\mathbf{x}(\mathbf{s}_i) - \bar{\mathbf{x}})(\mathbf{x}(\mathbf{s}_j) - \bar{\mathbf{x}})^\top,$$

where $\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^{n} \mathbf{x}(\mathbf{s}_i)$ is the sample mean and the set $\mathcal{N}(\mathbf{h}) = \{(i,j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}$ covers all sample locations that are separated by the lag vector $\mathbf{h}$. Usually, for irregularly spaced sample locations the separation vectors between each coordinates are unique. To overcome this issue the set $\mathcal{N}(\mathbf{h})$ is adapted in the sense that it considers sample locations which are separated between $\mathbf{h}-\mathbf{a}$ and $\mathbf{h}+\mathbf{a}$, often also angles are considered to model anisotropy. After the estimation usually a parametric covariance function is fitted by different least squares approaches. In contrast, fitting of the covariance function directly to the data by maximum likelihood or within the Bayesian framework are also widely used in the literature.

There exist many approaches on covariance modeling in the literature. Genton and Kleiber (2015) give an thorough investigation of multivariate stationary and non-stationary covariances, Sampson (2010) review different strategies for univariate non-stationary covariance models. In the following two prominent approaches to model the covariance matrix functional are given. Section 2.2.1 covers the rich family of Matérn covariance functions which are adapted for the uni- and multivariate as well as the stationary and non-stationary case and Section 2.2.2 details how more complex covariance functions can be build upon simple univariate ones in a linear fashion.

2.2.1. The Matérn family

One covariance function that gets great attention in the literature is the so-called Matérn covariance function (Guttorp and Gneiting, 2006). It is stationary and isotropic and formally defined by

$$C^M(h;\sigma^2,\nu,\phi) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{h}{\phi}\right)^\nu K_\nu \left(\frac{h}{\phi}\right). \tag{2.2}$$

The Matérn covariance function is parametrized by $\sigma^2 > 0$, $\nu > 0$ and $\phi > 0$ which are the variance, shape and range parameters respectively. K_ν denotes the Bessel function of second kind with shape parameter ν and Γ denotes the gamma function. The main advantage of this covariance function is that the shape parameter controls the behavior at small lags. Or equivalently ν controls the smoothness of realizations of random fields. Lower values lead to sharper realizations. Another advantage lies in the fact that other popular covariance functions arise if the smoothness parameter takes values of $k + 1/2$ where k is a natural number. To name an example: for $\nu = 1/2$ it results in the popular exponential covariance function

$$C^{exp}(h) = \sigma^2 \exp\left(-h/\phi\right),$$

more examples are provided by Guttorp and Gneiting (2006). Figure 2.4 illustrates the Matérn covariance function for different parameter sets and Figure 2.5 depicts corresponding realizations.

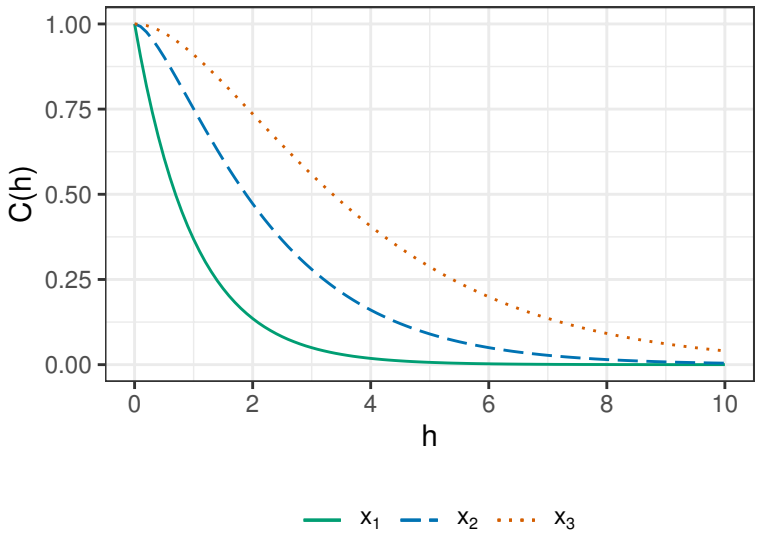

Figure 2.4.: Three different univariate stationary Matérn covariance functions. The parameters equal $(\sigma^2, \nu, \phi) = (1.0, 0.5, 1.0), (1.0, 1.0, 1.5)$ and $(1.0, 1.5, 2.0)$ for x_1, x_2 and x_3 respectively. Example realizations for these covariance models are depicted in Figure 2.5.

Extensions to the multivariate case are also provided in the literature. Gneiting et al. (2010) introduce models where the marginal as well as the cross covariance functions are all Matérn functions, i.e.:

$$C_{ii}^M(h) = \sigma_{ii}^2 C^M(h; 1, \nu_{ii}, \phi_{ii}) \quad \text{and} \quad C_{ij}^M(h) = \rho_{ij}\sigma_{ii}\sigma_{jj} C^M(h; 1, \nu_{ij}, \phi_{ij}),$$

for $i, j = 1, \ldots, p$ with $i \neq j$. Sufficient conditions on the parameters to provide a valid covariance functional are also given in the publication. For the case of $p > 2$ a so-called parsimonious Matérn is described by setting the range parameters equal across all functions ($\phi_{ij} = \phi$ for all $i, j = 1, \ldots, p$) and the cross shape parameters are set to the average of the marginal ones ($\nu_{ij} = (\nu_{ii} + \nu_{jj})/2$ for all $i, j = 1, \ldots, p$ with $i \neq j$). Conditions on the correlation parameters ρ_{ij} are rather involved and the interested reader is referred to the original publication. The publication also introduces the full bivariate Matérn model which relaxes the former conditions for different scale and shape parameters for the cross covariances in the case of $p = 2$. Apanasovich et al. (2012) extend the parsimonious Matérn model and allow under mild conditions different scale and shape parameters as well. Genton and Kleiber (2015) review the stationary Matérn family in more detail.

Anderes and Stein (2011) provide a univariate non-stationary Matérn model. This is achieved by allowing spatial varying shape $\nu : \mathbb{R}^d \to \mathbb{R}^+$ and variance $\sigma^2 : \mathbb{R}^d \to \mathbb{R}^+$ as well as spatial varying scale by $\boldsymbol{\phi} : \mathbb{R}^d \to \mathbb{R}_{pd}^{d \times d}$ where $\mathbb{R}^+$ is the set of positive real numbers and $\mathbb{R}_{pd}^{d \times d}$ is the set of real positive definite $d \times d$ matrices. The non-stationary Matérn model is defined by

$$C^M(\mathbf{s}, \mathbf{s}') = \sigma(\mathbf{s})\sigma(\mathbf{s}') \det\left(\boldsymbol{\phi}_{\mathbf{s},\mathbf{s}'}^{-1/2}\right) \left\|\boldsymbol{\phi}_{\mathbf{s},\mathbf{s}'}^{-1/2}(\mathbf{s} - \mathbf{s}')\right\|^{\nu_{\mathbf{s},\mathbf{s}'}} K_{\nu_{\mathbf{s},\mathbf{s}'}}\left(\left\|\boldsymbol{\phi}_{\mathbf{s},\mathbf{s}'}^{-1/2}(\mathbf{s} - \mathbf{s}')\right\|\right),$$

with $\boldsymbol{\phi}_{\mathbf{s},\mathbf{s}'} = (\boldsymbol{\phi}(\mathbf{s}) - \boldsymbol{\phi}(\mathbf{s}'))/2$ and $\nu_{\mathbf{s},\mathbf{s}'} = (\nu(\mathbf{s}) - \nu(\mathbf{s}'))/2$. The matrix valued scale function $\boldsymbol{\phi}$ can be restricted to map only onto the set of positive-valued diagonal matrices which leads to

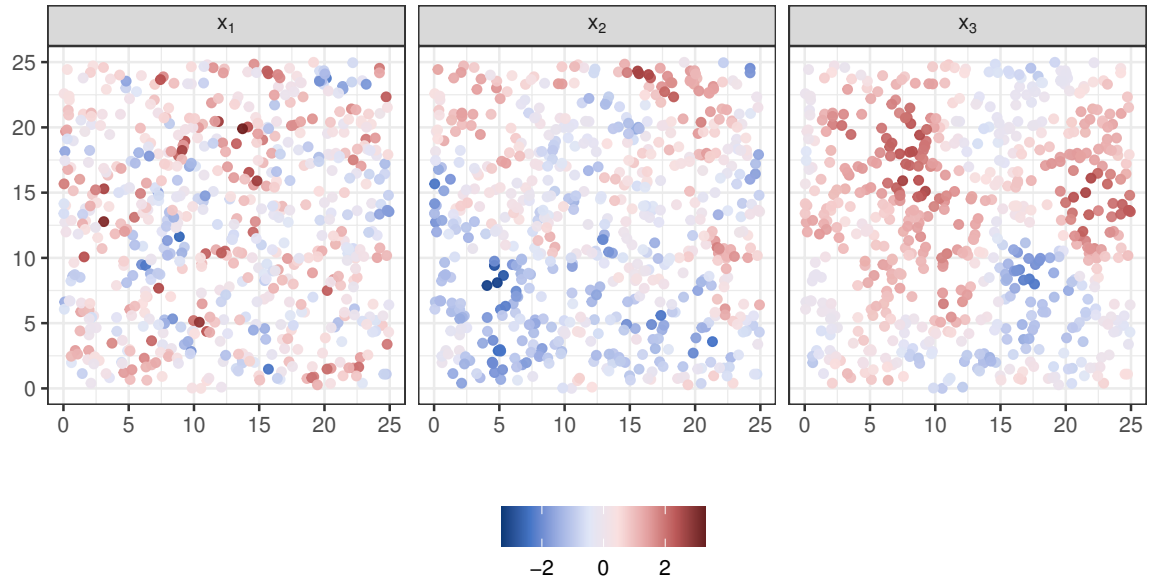

Figure 2.5.: Example realization of Gaussian random fields with Matérn covariance functions seen in Figure 2.4. The considered domain and the set of sample locations equal the ones from the left panel of Figure 2.1.

an isotropic model. Anderes and Stein (2011) also provide a slightly different version of the above model by only allowing isotropy and they show that for very smooth scale, shape and variance functions the above form (almost) simplifies to the stationary Matérn model.

Kleiber and Nychka (2012) combine the notions of Gneiting et al. (2010) and Anderes and Stein (2011) to provide a multivariate non-stationary covariance model where the marginal and cross covariances are of the non-stationary Matérn form.

As pointed out by Gneiting et al. (2010) the least squares framework should be avoided when fitting Matérn models as equal weight is put on each spatial lag which results in suboptimal estimation of the shape parameter. Gneiting et al. (2010) and Apanasovich et al. (2012) provide different parameter estimation algorithms based on maximum likelihood for the multivariate stationary case. For the non-stationary case parameter estimation is naturally an even more demanding task as parameter functions need to be estimated. Anderes and Stein (2011) use a local maximum likelihood variant for the non-stationary univarate case and Kleiber and Nychka (2012) introduce a two-step procedure that only relies on assuming that the first and second moments exist.

2.2.2. Linear model of coregionalization

Building more complex covariance models from simple univariate ones is a popular approach. For that purpose one can decompose the random field at hand into several independent ones by

$$x(\mathbf{s}) = \sum_{i=1}^{N} x_i(\mathbf{s}) + m,$$

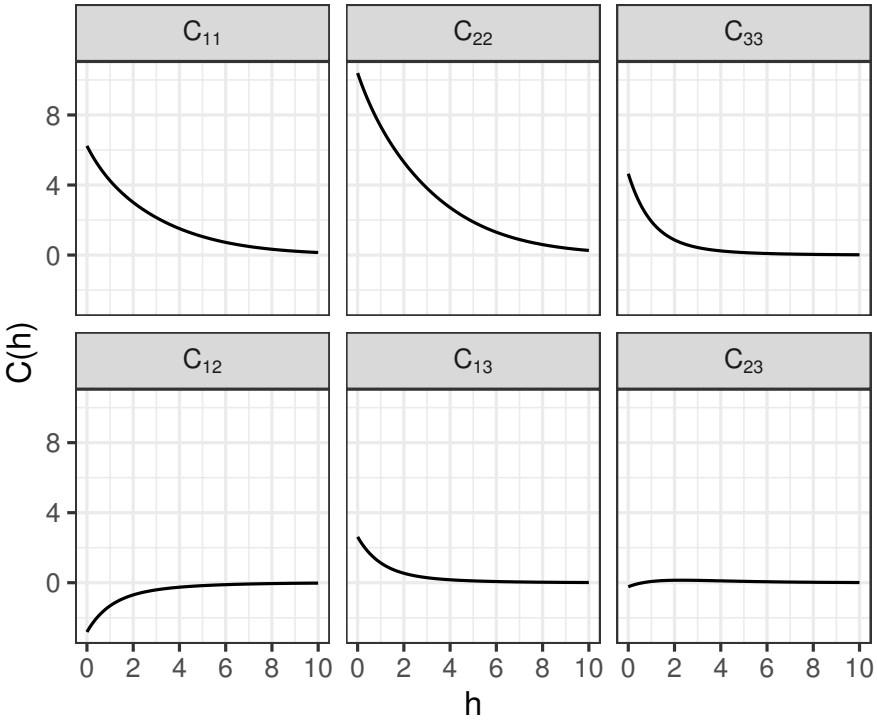

Figure 2.6.: Covariance structure for a three-variate LMC with $N = 2$. The univariate correlation functions are Matérn correlation functions with parameters $(\sigma^2, \nu, \phi) = (1.0, 0.5, 1.0)$ and $(1.0, 1.5, 2.0)$ (which equal the parameters of x_1 and x_3 seen in Figure 2.4). The entries for the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are drawn iid from $N(0, 1)$. An example realization of this model can be seen in Figure 2.7.

where m is a constant and each $x_i(\mathbf{s})$ is stationary, centered and for $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$, $i, j = 1, \ldots, N$ and $i \neq j$ it holds that $\text{Cov}(x_i(\mathbf{s}), x_j(\mathbf{s}')) = 0$. Leading to $\mathbb{E}(x(\mathbf{s})) = m$ and the linear model of regionalization (LMR) defined by

$$C^{LMR}(\mathbf{h}) = \sum_{i=1}^{N} C_i(\mathbf{h}) = \sum_{i=1}^{N} \sigma_i^2 \rho_i(\mathbf{h}).$$

Here, C_i, ρ_i and σ_i^2 are the covariance, correlation functions and the variance parameter for the i-th random field. In practical applications the covariance functions are chosen in such a way that they act at different spatial scales. For example, a suiting model might consists of $N = 3$ summands where the first one is a nugget effect and the second and third ones are exponential correlation functions with different range parameters to account for short and long distance effects reflecting the regionalization aspect of this model.

This idea is extended to the multivariate case by considering a separable covariance function defined by

$$\mathbf{C}(\mathbf{h}) = \mathbf{M}\rho(\mathbf{h}),$$

where $\mathbf{M}$ is a $p \times p$ covariance matrix (it is positive semi-definite) often denoted as coregionalization matrix and $\rho(\mathbf{h})$ is a univariate spatial correlation function. For this model each

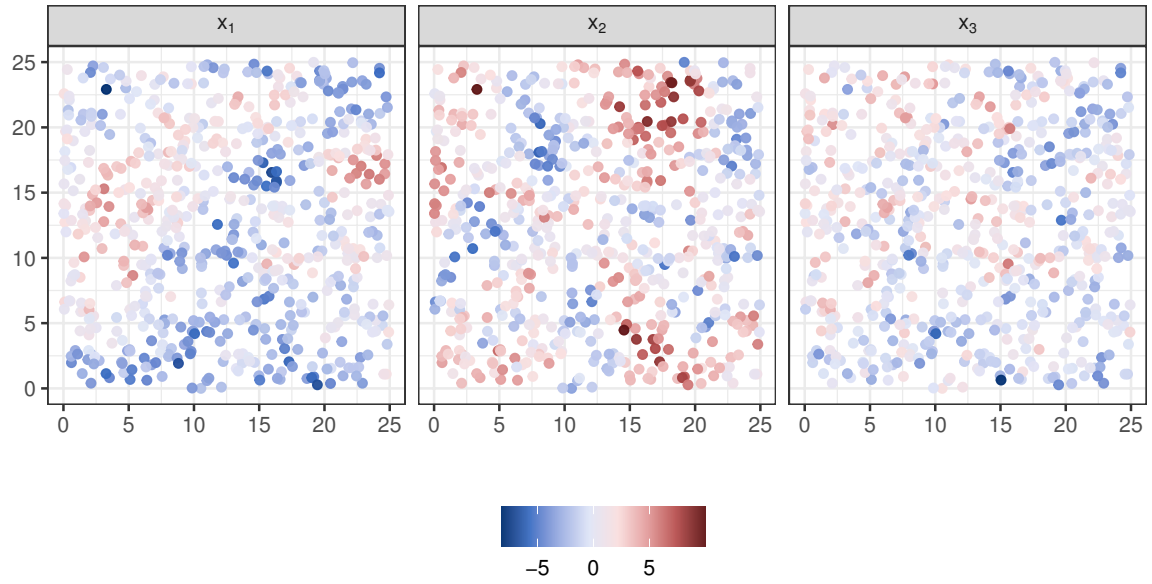

Figure 2.7.: Example realization of a Gaussian random field with the LMC structure seen in Figure 2.6. The considered domain and the set of sample locations equal the ones from the left panel of Figure 2.1.

marginal and cross covariance is formed by the corresponding entry of the matrix $\mathbf{M}$ multiplied by a univariate correlation function which is equal for each marginal and cross covariance. This model arises when the multivariate random field is build by

$$\mathbf{x}(\mathbf{s}) = \mathbf{A}\mathbf{z}(\mathbf{s}) + \mathbf{m}.$$

In the above form $\mathbf{A}$ is any $p \times p$ matrix, $\mathbf{m}$ is any p-dimensional vector and $\mathbf{z}(\mathbf{s})$ is a p-variate stationary and centered random field with independent entries (often denoted as factors) where $\rho(\mathbf{h})$ is the spatial correlation function for each factor. This leads to $\mathbb{E}(\mathbf{x}(\mathbf{s})) = \mathbf{0}$ and $\mathbf{M} = \mathbf{A}\mathbf{A}^\top$ which is positive semi-definite and its rank equals the one of $\mathbf{A}$. Note that this model is closely related to the spatial blind source separation model discussed in Chapter 4. In less recent literature such as Helterbrand and Cressie (1994) or Wackernagel (2003, Chapter 23) a separable covariance function is referred to as intrinsic correlation model.

Taking one step further the principle of regionalization is used for the cross covariances as well leading to the prominent linear model of coregionalization (LMC), see for example Goulard and Voltz (1992); Vargas-Guzmán et al. (2002); Emery (2010) or Wackernagel (2003, Chapter 26). It is defined by

$$\mathbf{C}^{LMC}(\mathbf{h}) = \sum_{i=1}^{N} \mathbf{M}_i \rho_i(\mathbf{h}). \tag{2.3}$$

Here, $\rho_i(\mathbf{h})$ are univariate correlation functions and $\mathbf{M}_i$ are the positive semi-definite coregionalization matrices. The LMC is symmetric in its distance argument as it is build from univariate stationary correlation functions. Note that the separable model is a special case of the LMC for $N = 1$. Similar as for the LMR the LMC arises when the random field is formed

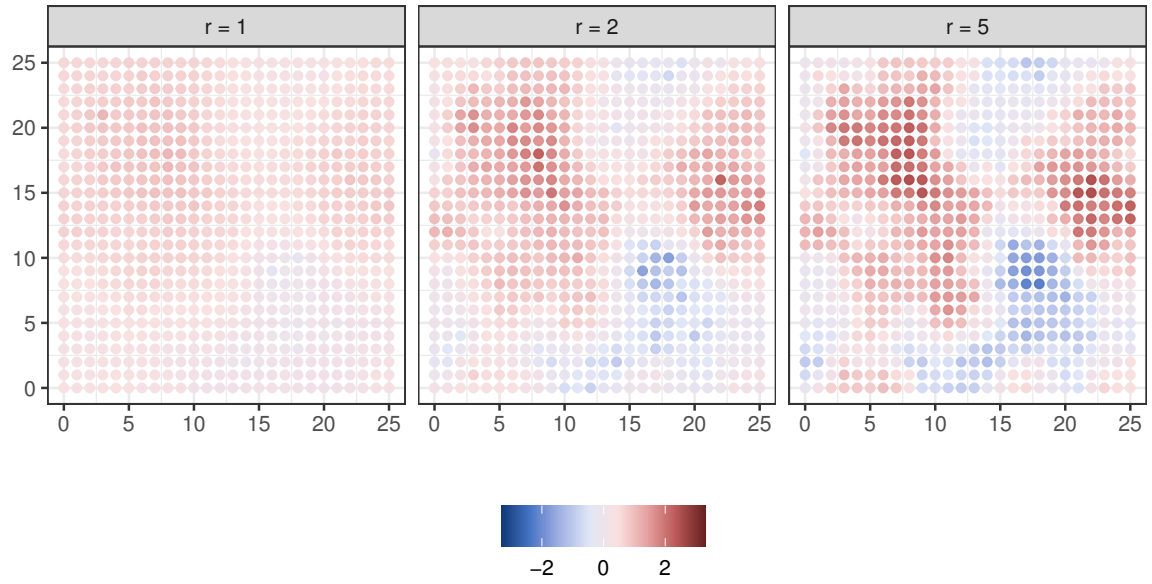

Figure 2.8.: IDW interpolation of x_3 of Figure 2.5 on the integer grid $(\mathcal{S} \cap \mathbb{Z}^2)$ for different parameters $r = 1, 2$ and 5.

by a sum of independent ones written as

$$\mathbf{x}(\mathbf{s}) = \sum_{i=1}^{N} \mathbf{A}_i \mathbf{z}_i(\mathbf{s}) + \mathbf{m}.$$

In the above formula $\mathbf{A}_i$ are $p \times p$ matrices, $\mathbf{m}$ is any p-dimensional vector and $\mathbf{x}_i(\mathbf{s})$ are p-variate stationary and independent random fields which consist of independent entries (factors) with equal spatial correlation. Yielding $\mathrm{Cov}((\mathbf{z}_i(\mathbf{s}+\mathbf{h}))_k(\mathbf{z}_i(\mathbf{s}))_k) = \rho_i(\mathbf{h})$ and $\mathrm{Cov}((\mathbf{z}_i(\mathbf{s}+\mathbf{h}))_k(\mathbf{z}_j(\mathbf{s}))_l) = 0$ for all $i, j = 1, \ldots, N$ and $k, l = 1, \ldots, p$ with $i \neq j$ and $k \neq l$ and $\mathbf{s}, \mathbf{s}+\mathbf{h} \in \mathcal{S}$. Therefore, $\mathbb{E}(\mathbf{x}(\mathbf{s})) = \mathbf{m}$ and the coregionalization matrices are formed by $\mathbf{M}_i = \mathbf{A}_i \mathbf{A}_i^\top$ which are of the same rank as $\mathbf{A}_i$. Figure 2.6 depicts the covariance matrix for an example LMC and Figure 2.7 shows one corresponding realization. Fitting of a LMC is usually done in a two-step procedure. Firstly, the number N and the parameters of the univariate correlation functions are chosen based on visual inspection of the empirical semivariogram or covariogram. Secondly, the coregionalization matrices are fitted by keeping the parameters of the univariate correlation functions untouched. Such a procedure is outlined for example by Goulard and Voltz (1992).

Gelfand et al. (2004) extended the LMC to the non-stationary case by letting the coefficient matrices depend on the spatial location, i.e.: $\mathbf{A}_i(\mathbf{s})$, but keeping the univariate spatial correlation functions still stationary.

2.3. Spatial prediction

One central task in the analysis of geostatistical datasets is predicting a measurement at an unobserved location, namely spatial interpolation or prediction. Specifically, the univariate

random field $x(\mathbf{s})$ is observed on n sample locations $\mathcal{C} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. The task is now to predict the value of $x(\mathbf{s})$ at some unobserved location $\mathbf{s}_0$.

Many methods form the prediction in a linear fashion by

$$\hat{x}(\mathbf{s}_0) = \sum_{i=1}^{n} a_i x(\mathbf{s}_i). \tag{2.4}$$

For example, the very simple inverse distance weighting (IDW) method determines the weights by the inverse of the distance, i.e.: $a_i = \|\mathbf{s}_0 - \mathbf{s}_i\|^{-r}$, with the restriction that the weights sum up to one. IDW has the advantage that no algorithms to fit the weights have to be utilized and the weights decrease quickly as a function of distance which leads to so-called local interpolation. However, the weights will always be positive, there is no error to characterize the prediction quality and the choice of r seems to be (mostly) arbitrary. Figure 2.8 shows the IDW interpolation for different r values of x_3 from Figure 2.5. Another class of methods that predict in such a linear way is provided by the Kriging methodology discussed below.

Countless variants that predict in a different fashion are also considered in the literature. For example, Wang et al. (2019) use neural networks with different forms of feature selection for univariate prediction. Generally, Li and Heap (2014); Jiang (2019) give an overview of the countless possibilities.

Ordinary Kriging Kriging is the standard tool for spatial prediction and extensively reviewed in the literature. There exist many variants, e.g.: for stationary, intrinsic stationary random fields, for multivariate random fields, with and without an external drift, etc. but all share the same principle. Namely, a linear predictor as seen in Equation (2.4) where the weights are determined by requiring that the bias is zero and minimizing the prediction mean squared error (MSE). Table 2.3 gives an overview and corresponding literature for prominent Kriging methods. As the principle of Kriging is equal for each method ordinary Kriging (OK) acts as an illustrative example and is detailed in the following.

For OK it is assumed that a univariate random field $x(\mathbf{s})$ is observed on n sample locations $\mathcal{C} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ with a constant but unknown mean m. The OK predictor can either be expressed in terms of the variogram (Wackernagel, 2003, Chapter 12) or the covariance function (Cressie, 1993, Chapter 3.2) where the former one relies on the assumptions that the random field is intrinsic stationary and the latter one on second-order stationarity assumptions. As stated above the prediction of the random field at the unobserved location $\mathbf{s}_0$ is formed by a linear combination of the observed values as seen in Equation (2.4). The quantity which is minimized to determine the weights a_i is the MSE which can be deduced into a variance and a bias term by

$$\mathbb{E}\left((\hat{x}(\mathbf{s}_0) - x(\mathbf{s}_0))^2\right) = \mathrm{Var}\left(\hat{x}(\mathbf{s}_0) - x(\mathbf{s}_0)\right) + \mathbb{E}\left(\hat{x}(\mathbf{s}_0) - x(\mathbf{s}_0)\right)^2,$$

where the last term is the squared bias. The desired property of zero bias

$$\mathbb{E}\left(\hat{x}(\mathbf{s}_0) - x(\mathbf{s}_0)\right) = m \sum_{i=1}^{n} a_i - m = 0$$

leads to the following conditions on the weights

$$\sum_{i=1}^{n} a_i = 1.$$

Table 2.1.: Overview of popular Kriging methods

Method	Description/Assumptions	Selected literature
Simple Kriging	Random field is stationary, the mean is constant and known	Wackernagel (2003, Chapter 3)
Ordinary Kriging	Random field is stationary or intrinsic stationary, the mean is constant but unknown	Cressie (1993, Chapter 3.2), Wackernagel (2003, Chapter 12)
Universal Kriging	The random field is formed by a deterministic drift function and an intrinsic random field with mean zero, the drift function is a linear combination of basis functions, often low order monomials in the coordinates	Cressie (1993, Chapter 3.4), Wackernagel (2003, Chapter 31)
Factorial Kriging	Prediction of the factors that lead to the LMC, see also Section 2.2.2	Goovaerts (1992), Ma et al. (2014)
Cokriging	Multivariate extension, all variables are measured at all sample locations is denoted as isotopic Cokriging, variables are measured at different locations is denoted as heterotopic Cokriging, the latter is often found in practice when one variable is expensive and others are cheap to measure	Stein and Corsten (1991), Wackernagel (1994), Goovaerts (1998)

Under this condition minimizing the MSE is equal to minimizing the prediction error variance which writes as

$$\mathrm{Var}\left(\hat{x}(\mathbf{s}_0) - x(\mathbf{s}_0)\right) = \sum_{i,j=1}^{n} a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) - 2\sum_{i=1}^{n} a_i C(\mathbf{s}_i - \mathbf{s}_0) + C(\mathbf{0}).$$

This can be equivalently rewritten in vector notation by letting $\mathbf{a} = (a_1, \ldots, a_n)^\top$, $\boldsymbol{\delta} = (C(\mathbf{s}_1 - \mathbf{s}_0), \ldots, C(\mathbf{s}_n - \mathbf{s}_0))^\top$ and $\boldsymbol{\Sigma}$ as seen in Equation (2.1) leading to

$$\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} - 2\mathbf{a}^\top \boldsymbol{\delta} + C(\mathbf{0}).$$

Now the MSE can be minimized with respect to the constraint on the weights by using the method of Lagrange multipliers which defines the Lagrange function to be

$$\phi(\mathbf{a}, a_0) = \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} - 2\mathbf{a}^\top \boldsymbol{\delta} + C(\mathbf{0}) + 2a_0(\mathbf{a}^\top \mathbf{1} - 1),$$

where $\mathbf{1}$ is the n-vector of ones and a_0 is the Lagrange multiplier. Setting the partial derivatives of the Lagrange function with respect to the weights to zero leads to the optimization equations

$$\boldsymbol{\Sigma}\mathbf{a} + a_0 \mathbf{1} = \boldsymbol{\delta} \quad \text{and} \quad \mathbf{a}^\top \mathbf{1} = 1,$$

where the latter equation is again the zero-bias condition. The above $n+1$ linear equations can be comfortably written in a single matrix multiplication as

$$\underbrace{\begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix}}_{=\tilde{\boldsymbol{\Sigma}}} \begin{pmatrix} \mathbf{a} \\ a_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\delta} \\ 1 \end{pmatrix}.$$

The weights can be simply computed by inverting $\tilde{\boldsymbol{\Sigma}}$. $\tilde{\boldsymbol{\Sigma}}$ is invertible if $\boldsymbol{\Sigma}$ is positive definite which is ensured when a proper covariance matrix functional is chosen and each sample location is only present once in the set of sample locations. The uncertainty of the prediction can be characterized by the prediction error variance which can be expressed using the optimization equations as

$$\sigma^2_{err} = C(\mathbf{0}) - a_0 - \mathbf{a}^\top \boldsymbol{\delta}.$$

Ordinary Kriging ensures exact prediction. Considering $\mathbf{s}_0 = \mathbf{s}_i$ then the i-th column of $\boldsymbol{\Sigma}$ is equal to $\boldsymbol{\delta}$. $\boldsymbol{\Sigma}$ is still full-rank ensuring that the only solution is the zero vector with the only non-zero element $a_i = 1$ leading to $\hat{x}(\mathbf{s}_i) = x(\mathbf{s}_i)$.

So far it is assumed that the spatial covariance is known which is in practical applications rarely encountered. Usually, a parametric model is fitted to the data as described in Section 2.2 leading to the fact that the Kriging prediction is additionally affected by the possible variation of the model parameters. Several cases can be characterized as follows. The semivariogram usually approaches a constant value when its argument tends to infinity as the spatial dependence usually vanishes at very large distances, this is denoted as sill. For a stationary random field the sill corresponds to the covariance at lag zero. A higher sill does not affect the Kriging weights but increases the prediction error variance. Usually, the sill is the sum of the nugget effect and the on-site variances of other considered covariance models where a higher relative nugget effect leads to more similar Kriging weights. This is intuitively explained by the fact that each Kriging weight equals $1/n$ when only a nugget effect is considered leading to weights approaching $1/n$ as the relative share of the nugget effect in the sill increases. The highest effect on the Kriging weights and on the error variance is given by possible anisotropies which are not included in the model, the overall range of spatial dependence present in the model and the behavior of the semivariogram (or covariance) near the origin. Due to the high influence of the latter the Matérn family is a very popular covariance choice as the shape parameter ν can control the behavior at the origin with high flexibility.

Generally, a thorough discussion about the influence of estimated parameters on the Kriging methodology is far beyond the scope of the present thesis. As pointed out by Stein (1999, Chapter 6) characterization of such effects is very involved due to the fact that linear predictions formed by the data at hand are no longer linear when the weights depend on parameters that are estimated from the same data.

3. Blind source separation

Blind source separation (BSS) is a multivariate statistical tool which is well-established for many forms of data. The foundation of the BSS methodology forms the so-called location-scatter model

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \mathbf{b}, \tag{3.1}$$

where $\mathbf{x}$ and $\mathbf{z}$ are p-variate random vectors denoted as observable and source or latent vector, both forming the random parts of the model. $\mathbf{A}$ is the invertible $p \times p$ mixing matrix and $\mathbf{b}$ is a location vector where both are deterministic. The aim of all BSS methods is to find a location vector functional $\mathbf{T} = \mathbf{T}(\mathbf{x})$ as well as a so-called unmixing matrix functional $\mathbf{W} = \mathbf{W}(\mathbf{x})$ which together recover the source vector by $\mathbf{W}(\mathbf{x} - \mathbf{T})$ given only the observable $\mathbf{x}$. One example for $\mathbf{W}$ would be simply the inverse of the mixing matrix $\mathbf{A}^{-1}$, indeed under sufficient assumptions on the source it is possible to recover $\mathbf{A}^{-1}$ up to certain ambiguities. Chapter 3.1 discusses different BSS models for different kinds of data with proper assumptions on the source and Chapter 3.2 outlines properties of unmixing matrix functionals. In Chapter 3.3 a two step procedure to find the unmixing matrix applicable for almost all BSS models is detailed and unmixing matrix functionals for the models discussed in Section 3.1 are given in Section 3.4. Lastly, Section 3.5 discusses the implications of noise and how the signal subspace can be found. Profound discussions on BSS are given in textbooks such as Cichocki and Amari (2002); Comon and Jutten (2010) and in review papers such as Nordhausen and Oja (2018); Pan et al. (2021).

Still, a natural question arises: What are the advantages of recovering the source over analyzing the original data? The advantages of BSS are threefold. (i) The entries of the source are at least uncorrelated or even independent. This leads to the great advantage that the data can be pre-processed by some BSS method and then each entry of the source can be studied individually. For example, in prediction tasks this procedure discards the need of building one multivariate model in favor of p univariate ones. (ii) The model states a linear connection between the source and observable. Hence, interpretations of the results can follow the same loadings-scores interpretation scheme as in principle component analysis (PCA), where the loadings correspond to the unmixing matrix and the scores to the source. But in contrast to PCA the BSS methods find features of the data beyond the correlation structure, the unmixing matrix is not restricted to be orthogonal and the results of the methods do not depend on prior linear transformations. The latter attribute is denoted as affine equivariance (discussed in Chapter 3.2). (iii) The dimension of the observable and the source is for both cases p. However, in most applications only a few components of the source are of interest. This acts as a form of dimension reduction which additionally also simplifies the univariate modeling as only less than p models need to be formulated.

3.1. Blind source separation models

Equation (3.1) forms the basic equation for the (instantaneous) mixture model, the aim of this chapter is to discuss different assumptions on the unobserved source vector which leads to different models for different forms of data.

3.1.1. Independent component model

For independently and identically distributed (iid) data BSS models are based on statistical independence assumptions on the source which results in the well-known independent component analysis (ICA). As outlined by Virta (2018) to motivate the independent component (IC) model consider the random vector $\mathbf{z}$ to be p-variate standard normal distributed, i.e.: $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$. $\mathbf{z}$ has two characteristic properties:

1. Spherical symmetry: $\mathbf{z} \sim \mathbf{V}\mathbf{z}$, where $\mathbf{V}$ is a $p \times p$ orthogonal matrix.
2. Independence property: all components of $\mathbf{z}$ are statistically independent.

Generalizations of the standard normal model that show the spherical symmetry property in conjunction with the location-scatter model lead to the class of elliptical models and the ones that posses the independence property lead to the class of IC models. The standard normal distribution is the only distribution that possesses these two former properties. If it is combined with the location-scatter model it forms the multivariate Gaussian model which is the only model in the intersection of the IC and elliptical models. Figure 3.1 gives examples for these three discussed models. A thorough discussion of the location-scatter model and its implications is given for example by Oja et al. (2006); Nordhausen and Oja (2018). Formally, the considered IC model is defined as follows.

Definition 7 (Independent component (IC) model)**.** *Consider a p-variate random vector $\mathbf{x}$. It follows an independent component model if it can be written as*

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \mathbf{b},$$

where $\mathbf{x}$ is the p-variate observable random vector, $\mathbf{z}$ is the p-variate latent random vector, $\mathbf{A}$ is the deterministic $p \times p$ mixing matrix of full-rank and $\mathbf{b}$ is a p-variate deterministic location vector. Furthermore, $\mathbf{z}$ fulfills the following conditions.

(IC 1) $\mathbb{E}(\mathbf{z}) = \mathbf{0}$ *and* $\mathrm{Cov}(\mathbf{z}) = \mathbb{E}\left(\mathbf{z}\mathbf{z}^\top\right) = \mathbf{I}_p$,

(IC 2) *the entries of the latent vector are mutually independent and*

(IC 3) *at most one component of the latent vector is Gaussian distributed.*

In the above model minimal distributional assumptions on the source vector $\mathbf{z}$ are stated. The first assumption fixes the location and scale leading to $\mathbb{E}(\mathbf{z}) = \mathbf{b}$ and $\mathrm{Cov}(\mathbf{x}) = \mathbf{A}\mathbf{A}^\top$ and the second one states the key independence property of the latent components. (IC 3) is in its essence an identifiability condition for the model parameters (namely the mixing matrix) and will be discussed in a subsequent chapter. Stronger assumptions than (IC 3) need to be stated for different estimators of the unmixing matrix as seen in Chapter 3.4.

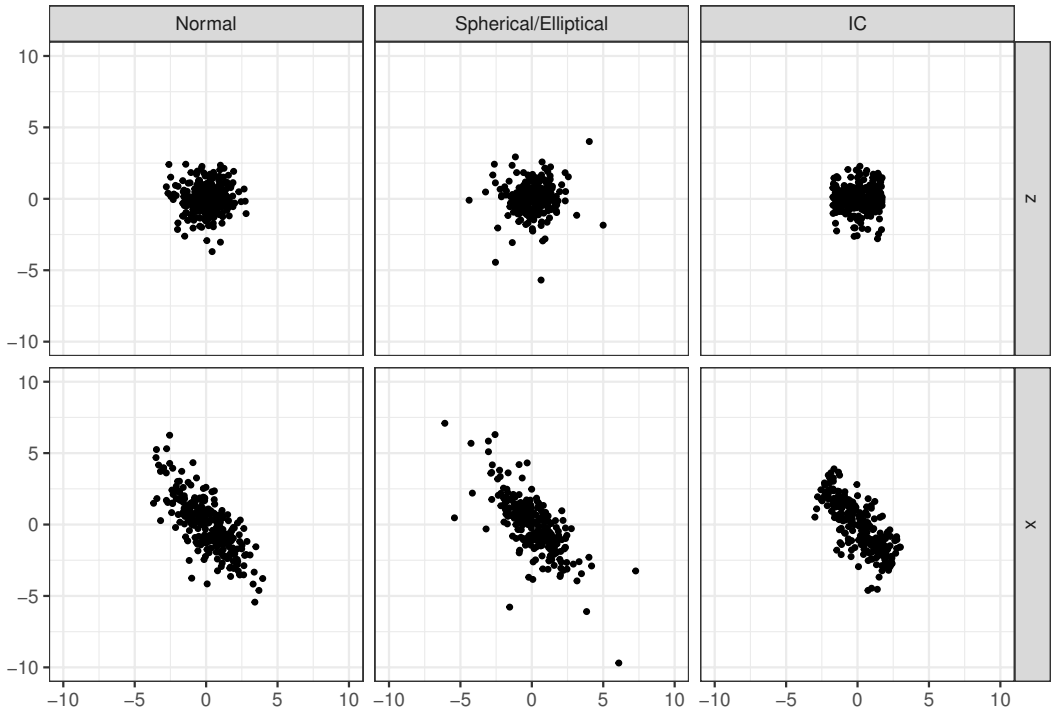

Figure 3.1.: Examples for the location-scatter model (Equation (3.1)) with different bivariate latent components $\mathbf{z}$. The first row corresponds to the latent vector where the first one follows a bivariate standard normal distribution, the second one a bivariate spherical t_6 distribution and the third one consists of a uniform and a univariate t_5 distribution. All latent vectors are centered and have unit covariance matrix. The second row shows the corresponding vectors $\mathbf{x}$ when the latent components are inserted into the location-scatter model with $\mathbf{b} = \mathbf{0}$ and the mixing matrix entries are drawn iid from $N(0, 1)$. This yields a multivariate normal, an elliptical and an IC model respectively. The sample size equals 250 for all six panels.

3.1.2. Second order source separation model

In contrast to iid data, time series data shows dependence between the random vectors at different time stamps, hence, at least the independence assumptions of iid data is not fulfilled. This property is exploited for time series BSS models in favor of the non-Gaussianity and the independence property of the IC model. The second order source separation (SOS) model is build around second-order stationary time series which are defined in similar fashion as in the spatial case. Specifically, a (univariate) time series $(x_t)_{t\in\mathbb{Z}}$ is said to be second-order stationary if (1) $\mathbb{E}\left(x_t^2\right) < \infty$, (2) $\mathbb{E}\left(x_t\right) = \mathbb{E}\left(x_{t'}\right)$ and (3) $\mathrm{Cov}(x_{t+\tau}, x_t) = \mathrm{Cov}(x_{t'+\tau}, x_{t'})$ for all $t, t', \tau \in \mathbb{Z}$. Figure 3.2 illustrates three different stationary time series. More details on time series can be found in textbooks such as Shumway and Stoffer (2011).

Definition 8 (Second order source separation (SOS) model)**.** *Consider a p-variate time series $\mathbf{x}_t$ for all $t \in \mathbb{Z}$. $\mathbf{x}_t$ follows a second order source separation model if it can be written as*

$$\mathbf{x}_t = \mathbf{A}\mathbf{z}_t + \mathbf{b},$$

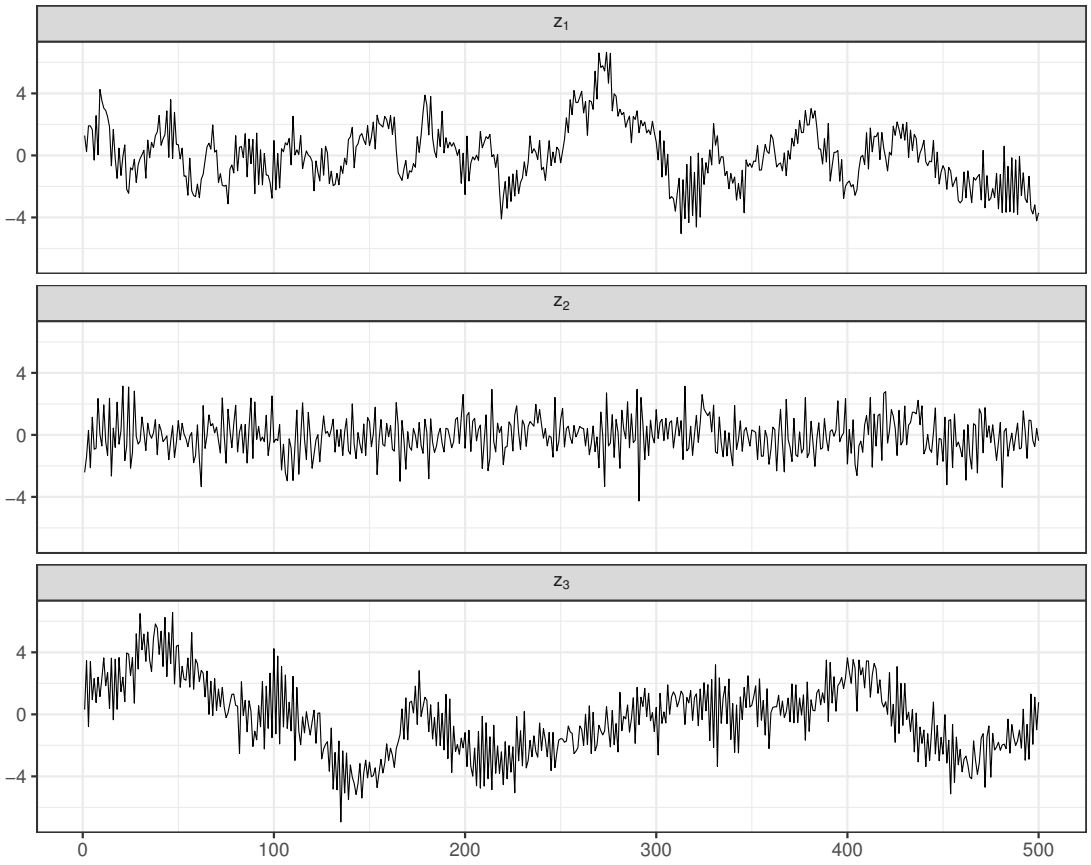

Figure 3.2.: Sample paths for three different stationary time series models. The first and third models are autoregressive AR(2) and AR(4) models where the parameters equal $(0.2, 0.7)$ and $(-0.1, 0.6, 0.2, 0.2)$ respectively. The second model is a MA(4) moving average process with parameters $(-0.4, 0.3, 0.7, -0.1)$. The lengths of the time series equal 500.

where $\mathbf{x}_t$ *is the* p*-variate observable time series,* $\mathbf{z}_t$ *is the* p*-variate latent time series,* $\mathbf{A}$ *is the deterministic* $p \times p$ *mixing matrix of full-rank and* $\mathbf{b}$ *is a* p*-variate deterministic location vector. Furthermore,* $\mathbf{z}_t$ *fulfills the following conditions.*

(SOS 1) $\mathbb{E}(\mathbf{z}_t) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{z}_t) = \mathbb{E}\left(\mathbf{z}_t \mathbf{z}_t^\top\right) = \mathbf{I}_p$ *for all* $t \in \mathbb{Z}$ *and*

(SOS 2) $\mathrm{Cov}(\mathbf{z}_{t+\tau}, \mathbf{z}_t) = \mathbb{E}\left(\mathbf{z}_{t+\tau} \mathbf{z}_t^\top\right) = \mathbf{D}_\tau$ *for all* $t \in \mathbb{Z}$ *and* $\tau = \pm 1, \pm 2, \ldots$ *where* $\mathbf{D}_\tau$ *is a diagonal matrix containing the (univariate) autocovariance functions for each entry of* $\mathbf{z}_t$ *which only depend on the lag* τ *on its diagonal elements.*

This model implies that the latent time series is formed by weakly stationary uncorrelated time series and that the way of mixing is constant for all time stamps (instantaneous mixing). Similar as in the iid model the location and scale is fixed by (SOS 1) but in contrast the separation of the signals is now based on their second order serial dependence (SOS 2). Hence, it is allowed that the source time series are Gaussian distributed. Further conditions for the source separation to be possible are different for the different estimators and discussed in Section 3.4. Figure 3.3 depicts a three-variate example of a SOS model. The SOS model

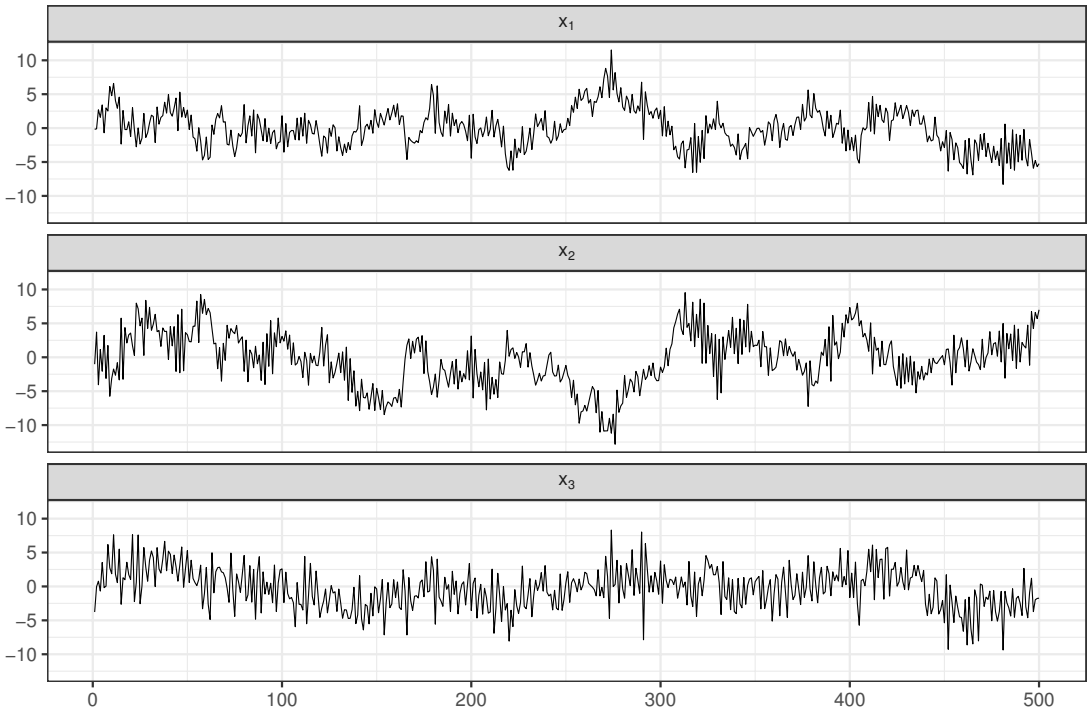

Figure 3.3.: Example of an SOS model which is formed by mixing the time series seen in Figure 3.2. The entries of the mixing matrix are drawn iid from $N(0,1)$ and $\mathbf{b} = \mathbf{0}$.

is widely considered in the literature, see for example Tong et al. (1990); Belouchrani et al. (1997); Miettinen et al. (2016). Pan et al. (2021) extensively review BSS for time series data.

3.1.3. Non-stationary source separation model

As BSS originates from the signal processing community it was soon noticed that for example in speech signals the stationarity assumption is not justified. Matsuoka et al. (1995) argued that rather a time varying variance of the source is a more realistic scenario. In that fashion a non-stationary time series can be defined as a time series that either shows a non-constant mean and/or a time varying second-order dependence structure. The latter case is considered in the non-stationary source separation (NSS) model as follows.

Definition 9 (Non-stationary source separation (NSS) model)**.** *Consider a p-variate time series $\mathbf{x}_t$ for all $t \in \mathbb{Z}$. $\mathbf{x}_t$ follows a non-stationary source separation model if it can be written as*

$$\mathbf{x}_t = \mathbf{A}\mathbf{z}_t + \mathbf{b},$$

where $\mathbf{x}_t$ is the p-variate observable time series, $\mathbf{z}_t$ is the p-variate latent time series, $\mathbf{A}$ is the deterministic $p \times p$ mixing matrix of full-rank and $\mathbf{b}$ is a p-variate deterministic location vector. Furthermore, $\mathbf{z}_t$ fulfills the following conditions.

(NSS 1) $\mathbb{E}(\mathbf{z}_t) = \mathbf{0}$ *for all* $t \in \mathbb{Z}$,

(NSS 2) $\mathrm{Cov}(\mathbf{z}_t) = \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t^\top\right) = \mathbf{D}_t$ *for all* $t \in \mathbb{Z}$ *where* $\mathbf{D}_t$ *are a positive definite diagonal matrices and*

(NSS 3) $\mathrm{Cov}(\mathbf{z}_t, \mathbf{z}_{t'}) = \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_{t'}^\top\right) = \mathbf{D}_{t,t'}$ *for all* $t, t' \in \mathbb{Z}$ *with* $t \neq t'$ *where* $\mathbf{D}_{t,t'}$ *are a diagonal matrices depending on* t *and* t'.

In the above model definition (NSS 1) fixes the location of the source but the scale cannot be fixed as it is allowed to vary in time by (NSS 2). A strategy to impose a scale restriction is achieved by assuming that the "global" covariance $\sum_{t=1}^{T} \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t^\top\right)$ equals $\mathbf{I}_p$ which comes without loss of generality as a global scale is not identifiable (see Section 3.2). Assumption (NSS 3) states that the serial second order dependence does depend on the specific times and not just on the lag between. Both assumptions, (NSS 2) and (NSS 3), reflect the non-stationary second-order dependence of the source vector. In some situations (NSS 2) and (NSS 3) are replaced by the stricter so-called block-stationary assumption. It states that the time series can be divided into non-overlapping sub-time series where each is second-order stationary with different (constant) covariance. For more details on the NSS model see Choi and Cichocki (2000a,b); Choi et al. (2001); Pham (2002); Pham and Cardoso (2001); Tanaka et al. (2006); Nordhausen (2014); Pan et al. (2021) and references therein.

The block-stationary model can also be viewed in a different way. Namely, each block time series might be seen as a realization of a BSS model where the sources are different but the mixing matrix is equal across blocks. This is often encountered in EEG datasets where the placement of sensors is equal for each patient. BSS for such form of data is often denoted as group ICA, details are for example given in Cong et al. (2013); Pfister et al. (2019).

3.1.4. Models in the context of regression analysis

In regression analysis the main aim is to find the relationship between a response random variable y (often univariate) and an explaining random vector $\mathbf{x}$ in the presence of some error term ϵ which in its most general form can be written as

$$y = f(\mathbf{x}, \epsilon).$$

In such a model the function f is assumed to be unknown and the error term ϵ is assumed to be independent of $\mathbf{x}$. With increasing dimension of the explaining vector modeling of the function f is increasingly difficult. Therefore, it would come with great advantage if r (linear) transformations of the explaining vector can be found which cover as much of the relationship between $\mathbf{x}$ and y as possible. In the best case $r \ll p$. The problem of finding such transformations is denoted as sufficient dimension reduction (SDR) and extensively considered in the literature, a review is provided by Ma and Zhu (2013) and a textbook by Li (2017). Certain SDR methods are also set into the context of the invariant coordinate system methodology (see Section 3.2) by Liski et al. (2014) as well as BSS by Matilainen et al. (2017a, 2019); Nordhausen et al. (2021b). The latter three publications consider the following model.

Definition 10 (SDR model in a BSS context)**.** *Consider a* $(p+1)$*-variate random vector* $(y, \mathbf{x}^\top)^\top$ *where* y *is the (univariate) response and* $\mathbf{x}$ *is its explaining vector which has the representation*

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \mathbf{b} = \mathbf{A}\begin{pmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{pmatrix} + \mathbf{b},$$

where $\mathbf{A}$ *is the invertible* $p \times p$ *mixing matrix,* $\mathbf{b}$ *is the deterministic* p*-dimensional location vector and* $\mathbf{z} = (\mathbf{z}_1^\top, \mathbf{z}_2^\top)^\top$ *is the* p*-variate latent random vector.* $\mathbf{z}$ *is partitioned into a* r*-variate subvector* $\mathbf{z}_1$ *and a* $(p-r)$*-variate subvector* $\mathbf{z}_2$ *which satisfy:*

(SDR 1) $\mathbb{E}(\mathbf{z}) = \mathbf{0}$ *and* $\mathrm{Cov}(\mathbf{z}) = \mathbf{I}_p$ *and*

(SDR 2) $\left(y, \mathbf{z}_1^\top\right)^\top \perp\!\!\!\perp \mathbf{z}_2$.

The dimension r *and the partitioning are minimal in the sense that, for projection matrices satisfying* $(y, \mathbf{Pz}) \perp\!\!\!\perp (\mathbf{I} - \mathbf{P})\mathbf{z}$*, the rank of* $\mathbf{P}$ *is larger than or equal to* r *and* $\mathbf{P}\left(\mathbf{z}_1^\top, \mathbf{0}^\top\right)^\top = \left(\mathbf{z}_1^\top, \mathbf{0}^\top\right)^\top$.

This models is formed around the location-scatter model and the typical BSS assumptions on the first two moments of the source (SDR 1). (SDR 2) is the key assumption which states that the relationship of the observable and the response variable is only given by the r-variate source subvector $\mathbf{z}_1$. The dimension r is the smallest for which all projections of the source still satisfy the key independence assumption (SDR 2). (SDR 2) is usually weakened in the SDR literature, e.g.: one of the first works on SDR (Li, 1991; Cook and Weisberg, 1991) state the following assumption.

(SDR 2') $\mathbf{z}_2 \perp\!\!\!\perp y | \mathbf{z}_1$ and $\mathbb{E}(\mathbf{z}_2 | \mathbf{z}_1) = \mathbf{0}$ (a.s.).

(SDR 2) implies (SDR 2') and is favored in the BSS literature as tests on the dimension r can be easily formulated (Nordhausen et al., 2021b).

After the source vector is found one can consider the regression model

$$y = g(\mathbf{z}_1, \epsilon),$$

where g is unknown and (possibly) different from f. Modeling of g is considerably simplified as the dimension of $\mathbf{z}_1$ is hopefully much less as the one of the original data $\mathbf{x}$ ($r \ll p$).

Matilainen et al. (2017a, 2019) consider a similar outline for (stationary) time series. In the above spirit one is interested in modeling a univariate response time series $(y_t)_{t\in\mathbb{Z}}$ based on a p-variate explaining time series $(\mathbf{x}_t)_{t\in\mathbb{Z}}$. But in contrast to the iid setting it is natural to additionally consider a link between the regressors present value and the predictors past values

$$y_{t+1} = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \epsilon_t, \epsilon_{t-1}, \ldots).$$

Here, the function f is again unknown and $(\epsilon_t)_{t\in\mathbb{Z}}$ is an unobserved time series that is independent from $(\mathbf{x}_t)_{t\in\mathbb{Z}}$. If the dimension of the time series is high it would be again of great advantage to find a small number of linear combinations of the predictor time series that capture as much of the dependency as possible between $\mathbf{x}_t$ and y_t. Formally, the time series BSS model for SDR writes as follows.

Definition 11 (Time series SDR model in a BSS context (TSDR))**.** *Consider a* $(p+1)$*-variate time series* $(y_t, \mathbf{x}_t^\top)^\top$ *with* $t \in \mathbb{Z}$ *where* y_t *is the (univariate) response and* $\mathbf{x}_t$ *is its explaining time series which has the representation*

$$\mathbf{x}_t = \mathbf{A}\mathbf{z}_t + \mathbf{b} = \mathbf{A}\begin{pmatrix}\mathbf{z}_{1,t}\\ \mathbf{z}_{2,t}\end{pmatrix} + \mathbf{b},$$

where $\mathbf{A}$ *is the invertible* $p \times p$ *mixing matrix,* $\mathbf{b}$ *is the deterministic* p*-dimensional location vector and* $\mathbf{z}_t = (\mathbf{z}_{1,t}^\top, \mathbf{z}_{2,t}^\top)^\top$ *is the* p*-variate stationary latent time series.* $\mathbf{z}_t$ *is partitioned into a* r*-variate time series* $\mathbf{z}_{1,t}$ *and a* $(p-r)$*-variate time series* $\mathbf{z}_{2,t}$ *which satisfy:*

(TSDR 1) $\mathbb{E}(\mathbf{z}_t) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{z}_t) = \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t^\top\right) = \mathbf{I}_p$ *for all* $t \in \mathbb{Z}$ *and*

(TSDR 2) $\left(y_{t''}, \mathbf{z}_{1,t'}^\top\right)^\top \perp\!\!\!\perp \mathbf{z}_{2,t}$ *for all* $t, t', t'' \in \mathbb{Z}$.

The dimension r *and the partitioning are minimal as in Definition 10.*

As in the SOS model the assumed mixture is instantaneous (the mixing matrix is independent of the time) and the latent time series is centered and has identity covariance matrix (TSDR 1). (TSDR 2) is the key assumptions which states that all dependency of the observable time series goes through the r-variate part of the latent time series $\mathbf{z}_{1,t}$, hence, the $(p-r)$-variate part $\mathbf{z}_{2,t}$ can be considered as noise and discarded for further regression analysis. If the dimension of the subvector is much smaller than the original predictor time series the goal of dimension reduction is achieved and the (hopefully) much simpler model

$$y_{t+1} = g(\mathbf{z}_{1,t}, \mathbf{z}_{1,t-1}, \ldots, \epsilon_t, \epsilon_{t-1}, \ldots).$$

can be considered. Here, g is possibly different from the former link function f.

In SDR only some of the latent components are considered as signal, i.e., those that carry information about the response. In similar fashion this can be formulated generally for BSS without a response variable. Namely, only some components might be of interest and the remaining ones can be discarded as they are noise. The specific definition of noise depends on the considered BSS model as follows.

3.1.5. Internal and external noise model

All the former BSS models share the same property that the source dimension as well as the observable dimension is equal. However, often only a few components of the source are of interest and the remaining ones can be discarded as noise. This circumstance leads to the internal noise model (INM). In contrast, in the external noise model (ENM) it is assumed that an additional p-variate noise term is added to the location scatter model.

Internal noise model The INM is formulated and discussed in the context of the SOS model (Definition 8) as considered in Matilainen et al. (2018); Nordhausen and Virta (2018); Virta and Nordhausen (2021) and defined as follows.

Definition 12 (Internal noise model in the SOS context (INM))**.** *Let* $\mathbf{x}_t$ *for all* $t \in \mathbb{Z}$ *be a* p*-variate time series.* $\mathbf{x}_t$ *follows an internal noise model if it can be written as*

$$\mathbf{x}_t = \mathbf{A}\begin{pmatrix}\mathbf{z}_{sg,t}\\ \mathbf{z}_{ns,t}\end{pmatrix} + \mathbf{b},$$

where $\mathbf{A}$ *is the deterministic* $p \times p$ *mixing matrix of full-rank and* $\mathbf{b}$ *is a* p*-variate deterministic location vector. The latent time series* $\mathbf{z}_t = \left(\mathbf{z}_{sg,t}^\top, \mathbf{z}_{ns,t}^\top\right)^\top$ *can be divided into a* q*-variate signal time series* $\mathbf{z}_{sg,t}$ *and a* $(p-q)$*-variate noise time series* $\mathbf{z}_{ns,t}$ *satisfying the following assumptions.*

(INM 1) $\mathbb{E}(\mathbf{z}_{sg,t}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{z}_{sg,t}) = \mathbb{E}\left(\mathbf{z}_{sg,t}\mathbf{z}_{sg,t}^\top\right) = \mathbf{I}_p$ *for all* $t \in \mathbb{Z}$ *and*

(INM 2) $\mathrm{Cov}(\mathbf{z}_{sg,t+\tau}, \mathbf{z}_{sg,t}) = \mathbb{E}\left(\mathbf{z}_{sg,t+\tau}\mathbf{z}_{sg,t}^\top\right) = \mathbf{D}_\tau$ *for all* $t \in \mathbb{Z}$, $\tau = \pm 1, \pm 2, \ldots$ *where* $\mathbf{D}_\tau$ *is a diagonal matrix containing the (univariate) autocovariance functions for each entry of* $\mathbf{z}_{sg,t}$ *which only depend on the lag* τ *on its diagonal.*

(INM 3) *The noise time series* $\mathbf{z}_{ns,t}$ *is white noise with identity covariance matrix and it is independent of* $\mathbf{z}_{sg,t}$.

The above model states that the latent time series is divided into a signal time series and noise time series where the former exhibits second order stationary dependence in time (the signal time series obeys an SOS model) and the latter does not. Additionally, the signal and the noise time series are statistically independent. The main goal for such a model is to estimate the signal subspace (namely $\mathbf{z}_{sg,t}$ and the corresponding dimension q). In further analysis only the signal part of the latent time series can be considered which leads to dimension reduction and further simplification of univariate modeling as only q time series need to be considered.

Nordhausen et al. (2017); Nordhausen and Oja (2018) consider a similar outline for the IC model (Definition 7) where the components of interest (signal) are assumed to be non-Gaussian and the noise part is assumed to be Gaussian. Nordhausen et al. (2017) provide bootstrap and asymptotic strategies for the signal subspace estimation, Luo and Li (2016, 2021) determine the signal subspace by using an information criterion based on eigenvalues and the variation of eigenvectors.

External noise model In the following the ENM is formulated in the context of the SOS model (Definition 8) as for example considered in Belouchrani and Cichocki (2000); Georgiev and Cichocki (2001) or in Cichocki and Amari (2002, Chapter 4).

Definition 13 (External noise model in the SOS context (ENM))**.** *Let* $\mathbf{x}_t$ *for all* $t \in \mathbb{Z}$ *be a* p*-variate time series.* $\mathbf{x}_t$ *follows an external noise model if it can be written as*

$$\mathbf{x}_t = \mathbf{A}\mathbf{z}_t + \mathbf{b} + \boldsymbol{\epsilon}_t,$$

where $\mathbf{A}$ *is the deterministic* $p \times q$ *mixing matrix of full-rank and* $\mathbf{b}$ *is a* p*-variate deterministic location vector.* $\mathbf{x}_t$ *is the* p*-variate observable time series,* $\mathbf{z}_t$ *is the* q*-variate latent source time series and* $\boldsymbol{\epsilon}_t$ *is the* p*-variate noise fulfilling the following assumptions.*

(ENM 1) $\mathbb{E}(\mathbf{z}_t) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{z}_t) = \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t^\top\right) = \mathbf{I}_q$ *for all* $t \in \mathbb{Z}$,

(ENM 2) $\mathrm{Cov}(\mathbf{z}_{t+\tau}, \mathbf{z}_t) = \mathbb{E}\left(\mathbf{z}_{t+\tau}\mathbf{z}_t^\top\right) = \mathbf{D}_\tau$ *for all* $t \in \mathbb{Z}$ *and* $\tau = \pm 1, \pm 2, \ldots$ *where* $\mathbf{D}_\tau$ *is a* $q \times q$ *diagonal matrix containing the (univariate) autocovariance functions for each entry of* $\mathbf{z}_t$ *which only depend on the lag* τ *on its diagonal,*

(ENM 3) *The noise* $\boldsymbol{\epsilon}_t$ *is independent of the source* $\mathbf{z}_t$*, centered and exhibits zero serial dependence and*

(ENM 4) $\mathrm{Cov}(\boldsymbol{\epsilon}_t) = \mathbb{E}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right) = \mathbf{R}$ *for all* $t \in \mathbb{Z}$ *where* $\mathbf{R}$ *is a non-negative definite* $p \times p$ *matrix.*

In contrast to the INM in the ENM a white noise time series is an additional additive term to the location-scatter model and the source obeys a SOS model but the dimension q is less or equal to p. The ENM model assumptions differ in the literature based on the context of the analysis. For the case of $q < p$ the goal is to infer the number q and provide an orthogonal transformation to q components which still cover the signal. For this case Virta and Nordhausen (2019) outlined a similar model for iid data. For the case of $q = p$ the goal is to properly estimate the unmixing matrix without influence of the noise which is outlined in Belouchrani and Cichocki (2000); Georgiev and Cichocki (2001). Choi et al. (2002) combine the ENM with the NSS model (Definition 9).

3.2. Unmixing matrix functional properties

Two properties are usually formulated in the literature that BSS unmixing matrix functionals are supposed to fulfill. Identifiability is concerned with assumptions on the latent vector in order to make the model parameters identifiable (up to certain ambiguities). Affine equivariance ensures that the same latent components are found independently of the way of mixing, it is a property desired for many statistical tools.

Identifiability A natural discussion regards the question if (or under which conditions) the model parameters, namely the mixing matrix $\mathbf{A}$ (or its counterpart the unmixing matrix $\mathbf{W}$) and the location vector $\mathbf{b}$ can be identified. Without any further assumptions on the source the location-scatter model (Equation (3.1)) is totally ill-defined in the sense that the pairs $(\mathbf{A}, \mathbf{z})$ and $(\mathbf{AM}, \mathbf{M}^{-1}\mathbf{z})$ lead to the same observable $\mathbf{x}$ for any invertible $p \times p$ matrix $\mathbf{M}$. To make the parameters more identifiable almost all BSS methods state assumptions on the first two moments (e.g.: (IC 1) from Definition 7 or (SOS 1) from Definition 8). These assumptions ensure an identifiable location vector and that the pairs $(\mathbf{A}, \mathbf{z})$ and $(\mathbf{AV}, \mathbf{V}^\top\mathbf{z})$ lead to the same observable $\mathbf{x}$ for any orthogonal $p \times p$ matrix $\mathbf{V}$, where $\mathbf{z}^* = \mathbf{V}^\top\mathbf{z}$ fulfills for example (IC 1).

Still, the set of orthogonal matrices is too rich, therefore, the ambiguities are reduced by further assumptions on the latent vector. These further assumptions are specific to the form of data and lead to a rich family of BSS models (see Section 3.1 for some examples). As an example the IC model (Definition 7) is examined in more detail as follows. (IC 2) states that the source is formed by independent components. On the one hand it is clear that if more than one component of the latent vector is Gaussian distributed still an orthogonal matrix forms the ambiguity of the model by the spherical symmetry of the standard normal distribution but on the other hand it is not clear if (non-trivial) linear combinations of independent non-Gaussian random variables can be constructed that are statistically independent. This would again result in an orthogonal matrix ambiguity of the model. However, the Skitovich-Darmon theorem (Ghurye and Olkin, 1962) states that if (non-trivial) linear combinations of independent random variables that are themselves independent can be constructed then the original random variables are Gaussian distributed. Both arguments above form the foundation of the assumption (IC 3) which results in (almost) identifiable parameters as only pairs $(\mathbf{A}, \mathbf{z})$ and $(\mathbf{APJ}, \mathbf{JP}^\top\mathbf{z})$ lead to the same observable $\mathbf{x}$, where $\mathbf{z}^* = \mathbf{JP}^\top\mathbf{z}$ fulfills (IC 1) - (IC 3). Here, $\mathbf{P}$ is any permutation matrix which contains exactly one 1 in each column and row and zero otherwise and $\mathbf{J}$ is any sign-change matrix which is diagonal and contains either -1 or 1 as diagonal elements. Therefore, (IC 3) results in a totally identifiable IC model meaning that the unmixing matrix is now identifiable up to sign and order of its rows. Virta et al. (2016); Nordhausen and Oja (2018) discuss implications for the IC model with more than one Gaussian component which is then naturally an INM for the iid case.

Note that often it is said that the latent vector is identifiable up to sign and order, which in its essence does not change the statements above. In some BSS models (e.g.: Definition 9) also the scale is not identifiable, resulting in replacing $\mathbf{PJ}$ by $\mathbf{PJD}$ where $\mathbf{D}$ is any positive-definite diagonal matrix. General discussions of the ambiguities in BSS models are given for example in Tong et al. (1991); Eriksson and Koivunen (2004).

Affine equivariance Besides the discussion about indentifiability, BSS methods are also desired to possess the affine equivariance property. Namely, if the observable is affine transformed

by $\mathbf{x}^{af} = \mathbf{M}\mathbf{x} + \mathbf{a}$ where $\mathbf{M}$ is any invertible $p \times p$ matrix and $\mathbf{a} \in \mathbb{R}^p$ is any vector then the unmixing matrix functional $\mathbf{W}(\mathbf{x}^{af})$ should equal $\mathbf{W}(\mathbf{x})\mathbf{M}^{-1}$ up to the ambiguities discussed before. Hence, the same latent vector $\mathbf{z}$ is obtained up to sign, order (and possible scale) independently of the way of mixing, or alternatively independently of the coordinate system. BSS methods are one of many statistical methods that possess the affine equivariance property, the class of such methods is denoted as invariant coordinate system (ICS) and discussed for example in Tyler et al. (2009); Ilmonen et al. (2010a, 2012). This property can be motivated also by the fact that BSS methods were designed to recover physical processes that are measured by sensors. The placement of the sensors defines the way of mixing (mixing matrix). Classical physical processes take place equally independently of the actual way of measuring, therefore recovering of these processes should be independent of the way of measuring as well.

One example of the great advantage of the affine equivariance property is given by data where relative information is of importance. Such data is denoted as compositional data, for details see textbooks such as Aitchison (2003). For compositional data it is common to linearly transform the raw data into different coordinate systems which have advantages in interpretation. However, the affine equivariance property of BSS methods leads to the advantage that always the same latent components are found but interpretations can be carried out with respect to the favored coordinate system. Details on that approach are given in Nordhausen et al. (2015, 2021a); Muehlmann et al. (2021b).

Both properties above are formalized by the following definition (see also Miettinen et al. (2015, Definition 1) for the iid case and Matilainen et al. (2015, Definition 1) for the time series case).

Definition 14 (Unmixing matrix functional)**.** *Let* $\mathbf{x}$ *be a* p*-variate random vector following an IC model from Definition 7. A* $p \times p$ *matrix-valued functional* $\mathbf{W}(\mathbf{x})$ *is an unmixing matrix functional if it possesses the following identifiability and affine equivariance properties.*

(Identifiability) $\mathbf{W}(\mathbf{z}) = \mathbf{P}\mathbf{J}$ *for any* $p \times p$ *permutation matrix* $\mathbf{P}$ *and* $p \times p$ *sign change matrix* $\mathbf{J}$*.*

(Affine equivariance) $\mathbf{W}(\mathbf{M}\mathbf{x} + \mathbf{a}) = \mathbf{P}\mathbf{J}\mathbf{W}(\mathbf{x})\mathbf{M}^{-1}$ *where* $\mathbf{M}$ *is any invertible* $p \times p$ *matrix,* $\mathbf{a}$ *is a* p*-dimensional vector,* $\mathbf{P}$ *is any* $p \times p$ *permutation matrix and* $\mathbf{J}$ *is any* $p \times p$ *sign change matrix.*

Note that Definition 14 is formulated for the special case of an IC model, but in its essence the outline is equal for other BSS models. As discussed above for some BSS functionals it is not possible to recover the scale. Therefore, the definition above would be adapted in the sense that additionally to the indeterminacies of sign and order also the scale is added by a positive-definite diagonal matrix.

Minimum distance index In simulation situations it is desirable to quantify the performance of a BSS method. This can be done by measuring the distance between $\mathbf{WA}$ and the identity matrix up to sign, order and scale. Note that due to the affine equivariance property $\mathbf{WA}$ yields always $\mathbf{PJ}$ (or $\mathbf{PJD}$), therefore, in simulations it comes without loss of generality to only consider the trivial mixing $\mathbf{A} = \mathbf{I}_p$.

One popular performance measure is the so-called minimum distance index (MDI) studied by Ilmonen et al. (2010b); Lietzen et al. (2020) and defined as follows (again formulated for the IC model but easily adapted to any BSS model).

Definition 15 (Minimum distance index)**.** *For a p-variate vector* $\mathbf{x}$ *generated by a IC model (Definition 7) with corresponding mixing matrix* $\mathbf{A}$ *the minimum distance index (MDI) for an unmixing matrix functional* $\mathbf{W} = \mathbf{W}(\mathbf{x})$ *is defined as*

$$\mathrm{MDI}(\mathbf{W}) = \frac{1}{\sqrt{p-1}} \inf_{\mathbf{H} \in \mathcal{H}} \|\mathbf{HWA} - \mathbf{I}_p\|_F,$$

where $\mathcal{H}$ *is the set of all* $p \times p$ *matrices that contain exactly one non-zero element on each row an column.*

In the former definition $\|\cdot\|_F$ denotes the Frobenius matrix norm and $\mathcal{H}$ contains all possible matrix products of sign change, permutation and positive-definite diagonal matrices (the indeterminacies outlined above). The MDI is a function $\mathrm{MDI} : \mathbb{R}^{p \times p} \to [0, 1]$ where zero indicates a perfect recovery of the unmixing matrix. Publications III, IV, and VI use this quantity to validate the performance of the introduced BSS functionals in extensive simulation studies.

3.3. How to find an unmixing matrix functional?

The former sections review popular BSS models and state desirable properties of an unmixing matrix functional. In this section approaches of finding such an unmixing matrix functional are discussed. In general, this is a demanding task as an unmixing matrix is only restricted to be invertible. Therefore, many BSS methods utilize the fact that the problem of finding any invertible matrix can be reduced to only finding an orthogonal matrix after a proper transformation. This is based on a theorem which is for example stated and proved in Miettinen et al. (2015) for the IC model as follows.

Lemma 1. *Let* $\mathbf{x}$ *be a p-variate random vector obeying an IC model (Definition 7). The mean vector equals* $\mathbb{E}(\mathbf{x}) = \mathbf{b}$ *and the covariance matrix equals* $\mathrm{Cov}(\mathbf{x}) = \mathbf{A}\mathbf{A}^\top$*. The whitened version of* $\mathbf{x}$ *writes as* $\mathbf{x}^{wh} = \mathrm{Cov}^{-1/2}(\mathbf{x})(\mathbf{x} - \mathbf{b})$*. Then, it holds that*

$$\mathbf{x}^{wh} = \mathbf{U}^\top \mathbf{z},$$

where $\mathbf{U}$ *is some* $p \times p$ *orthogonal matrix.*

Based on the former lemma the outline to find an unmixing matrix for almost all BSS models can be divided into two steps. Whitening step: Whiten the observable with respect to the covariance matrix (or suitable other matrices as discussed below). Rotation step: Find an orthogonal transformation that maximizes some information criterion of interest. To obtain the resulting unmixing matrix functional the matrices from the first and second step are simply combined via matrix multiplication. Scatter matrices as defined below play an important role in the discussion of BSS methods.

Definition 16 (Scatter matrix functional)**.** *Let* $\mathbf{x}$ *be a p-variate random vector. A* $p \times p$ *matrix functional* $\mathbf{S} = \mathbf{S}(\mathbf{x})$ *is a scatter matrix if it is symmetric, positive definite and possess the affine equivariance property*

$$\mathbf{S}(\mathbf{Mx} + \mathbf{a}) = \mathbf{M}\mathbf{S}(\mathbf{x})\mathbf{M}^\top,$$

for any choice of $\mathbf{x}$ *where* $\mathbf{M}$ *is any* $p \times p$ *invertible matrix and* $\mathbf{a}$ *is any p-dimensional vector.*

Scatter matrices are discussed in the context of BSS for example by Oja et al. (2006) and Nordhausen and Tyler (2015) where the latter publications relaxes the positive definite property to positive semi-definiteness. Scatter matrices play also an important role in ICS (Ilmonen et al., 2012).

3.3.1. Whitening step

The whitening step is most often carried out with respect to the covariance matrix $\mathrm{Cov}(\mathbf{x})$ which is the most prominent example of a scatter matrix. As it is positive (semi)-definite its eigen-decomposition given by $\mathbf{V}\mathbf{D}\mathbf{V}^\top$ always exists, here $\mathbf{V}$ is an orthogonal matrix containing the eigenvectors of $\mathrm{Cov}(\mathbf{x})$ as columns and $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathrm{Cov}(\mathbf{x})$ as diagonal entries. If the diagonal elements of $\mathbf{D}$ are pairwise distinct then the eigen-decomposition is unique. Additionally, if the diagonal elements of $\mathbf{D}$ are strictly positive ($\mathrm{Cov}(\mathbf{x})$ is positive definite) the inverse square-root of $\mathrm{Cov}(\mathbf{x})$ can be defined as $\mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^\top$ which is symmetric and unique as well. One advantage of the symmetric choice is its behavior when the observable is transformed under affine transformations. As outlined by Ilmonen et al. (2012) it holds that $\mathrm{Cov}^{-1/2}(\mathbf{M}\mathbf{x}+\mathbf{a}) = \mathbf{O}\,\mathrm{Cov}^{-1/2}(\mathbf{x})\mathbf{M}^{-1}$ for a unique orthogonal matrix $\mathbf{O}$, any invertible matrix $\mathbf{M}$ and any vector $\mathbf{a}$. This property together with proper choices of information criteria in the following discussed rotation step ensures the affine equivariance property of BSS functionals generally. A thorough discussion of whitening is provided by Ilmonen et al. (2012); Virta (2018).

The whitening step with respect to the covariance matrix also shows the relation of BSS methods to PCA. The principal components are obtained by the orthogonal transformation defined by the matrix of eigenvectors $\mathbf{V}^\top$. The additional transformation given by $\mathbf{D}^{-1/2}$ transforms all the principal components to unit variance. In total the whitening step removes the correlation structure of the observable where the last orthogonal transformation given by $\mathbf{V}$ has no influence on the unit covariance. The followup rotation step in BSS methods yields another rotation $\mathbf{U}$ which finds structure of the data past its correlation structure. In that sense BSS methods might be seen as an extension of PCA. The relation between ICA and PCA is also discussed by Nordhausen and Oja (2018); Virta (2018) and summarized by

$$\mathbf{V}^\top(\mathbf{x}-\mathbb{E}(\mathbf{x})), \quad \mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^\top(\mathbf{x}-\mathbb{E}(\mathbf{x})) \quad \text{and} \quad \mathbf{U}^\top\mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^\top(\mathbf{x}-\mathbb{E}(\mathbf{x})).$$

Here the first form corresponds to the the principal components, the second form yields the whitened version of $\mathbf{x}$ and the last form corresponds to the independent components. Figure 3.4 depicts the resulting principal components as well as the whitened version of the three models seen in the second row of Figure 3.1.

The whitening step can in principle be carried out with respect to any scatter matrix functional $\mathbf{S} = \mathbf{S}(\mathbf{x})$. However, it must fulfill that the so-called global mixing matrix $\mathbf{S}^{-1/2}(\mathbf{x})\mathbf{A}$ is orthogonal which is equivalent to the condition that $\mathbf{S}(\mathbf{z}) = \mathbf{D}$ where $\mathbf{D}$ is a positive definite diagonal matrix. Note that usually $\mathbf{S}$ needs to additionally fulfill the independence property which is in more detail discussed in Section 3.4.1. This ensures that the rotation step can be carried out in a meaningful way. If so, it holds that $\mathbf{S}^{-1/2}(\mathbf{x})(\mathbf{x}-\mathbf{b}) = \mathbf{S}^{-1/2}(\mathbf{x})\mathbf{A}\mathbf{z} = \mathbf{U}\mathbf{z}$ which reflects Lemma 1, where $\mathbf{U}$ is an orthogonal matrix and $\mathbf{S}^{-1/2}(\mathbf{x})(\mathbf{x}-\mathbf{b})$ is the whitened version of $\mathbf{x}$ with respect to the scatter matrix $\mathbf{S}$. This procedure is for example used in BSS methods with external noise (more details in Section 3.5.2) where it is denoted as robust orthogonalization, see Belouchrani and Cichocki (2000); Georgiev and Cichocki (2001); Choi et al. (2002) or the textbook Cichocki and Amari (2002, Chapter 4).

3.3.2. Rotation step

The rotation step is concerned with finding an orthogonal matrix that maximizes some information criterion of interest which is suited to the BSS model at hand. Two popular strategies

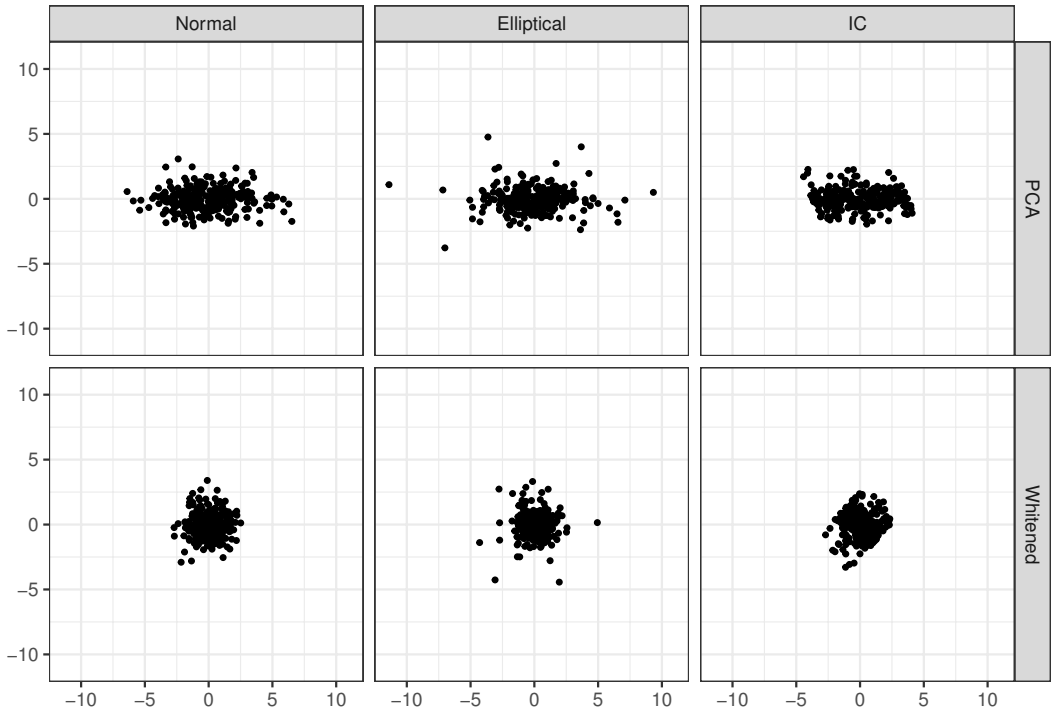

Figure 3.4.: Principal components and the whitened version of the data seen in the second row of Figure 3.1.

are utilized by almost all BSS methods. The first approach that is exclusively followed in the present thesis is based on moment assumptions and simultaneous/joint diagonalization of two/many suitable matrices. Such methods are sometimes denoted as algebraic BSS methods and detailed in Section 3.4.

Another popular strategy is based on projection pursuit (PP) (Huber, 1985) ideas and originally formulated for the IC model leading to the well-known fastICA methodology. The main idea of PP is to find linear combinations of the multivariate data at hand that maximize some information criterion. This information criterion is also denoted as projection index and typically has the following form for fastICA

$$\sum_{i=1}^{p} |G(\mathbf{u}_i^\top \mathbf{x}^{wh})|^a,$$

where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$ is the $p \times p$ orthogonal maximizer of the above projection index. The so called non-linearity $G : \mathbb{R} \to \mathbb{R}$ is supposed to be non-linear and non-quadratic and aims to measure the non-Gaussianity of its argument. Therefore, it possesses $\mathbb{E}\,(G(x)) = 0$ if $x \sim N(0, 1)$. Popular choices for the non-linearity are the pow3, tanh and gauss function which write as $(x^4 - 3)/4$, $\log(\cos(x)) - c_t$ and $-\exp(-x^2/2)c_g(x^4 - 3)/4$ respectively, where c_t and c_g are constants to ensure the minimum condition. Algorithms that find the orthogonal maximizer $\mathbf{U}$ are usually fixed-point algorithms and rely only on the derivatives of G which act also as the names for the non-linearities.

FastICA methods can be roughly classified by the way of finding the orthogonal maximizer

and the value of $a \in \{1, 2\}$. The first introduced fastICA method relies on the pow3 non-linearity, $a = 1$ and finds each column of $\mathbf{U}$ one-by-one (this procedure is denoted as deflation-based), see Hyvärinen and Oja (1997); Hyvärinen (1999a); Ollila (2010). The drawback of the deflation-based approach is that the estimation quality depends heavily on the order of the found directions as the error of estimation progresses through the directions. This issue is addressed by the reloaded fast ICA method (Nordhausen et al., 2011) which uses firstly FOBI or k-JADE (see the subsequent chapter) and then finds the optimal component order based on asymptotic arguments. One step further is taken by Miettinen et al. (2014) who suggest a fastICA variant where not only the optimal order but also the optimal non-linearity is chosen from a broad set of functions based on asymptotic arguments, denoted as adaptive deflation-based fastICA. The first row of Figure 3.5 illustrates the independent components found by applying the adaptive deflation-based fastICA method on the data seen in the second row of Figure 3.1. Another workaround is provided by the symmetric fastICA approach (Hyvärinen, 1999b; Wei, 2015), it uses $a = 1$ and finds all directions simultaneously (denoted as symmetric). In similar fashion the squared symmetric fastICA variant (Miettinen et al., 2017) finds the directions in a symmetric way and replaces the absolute value with the square ($a = 2$) for the projection index leading to optimization equations that put more emphasis on non-Gaussian components. For the above fastICA variants the reloaded, adaptive deflation-based, symmetric and squared symmetric variants are affine equivariant whereas the original deflation-based fastICA method lacks this property.

Also FastICA ideas are adapted for time series data for example by Hyvärinen (2001); Shi et al. (2009).

3.4. Unnmixing matrix functionals using simultaneous/joint diagonalization

Algebraic BSS methods are usually based on the diagonalization of two (simultaneous diagonalization) or many (joint diagonalization) matrices. The details on either approach as well as specific methods for the models of Section 3.1 are discussed in the following.

Simultaneous diagonalization Simultaneous diagonalization is the problem of finding a matrix that diagonalizes two symmetric matrices. Details on this problem can be found for example in Harville (1997, Chapter 21).

Definition 17 (Simultaneous diagonalization problem)**.** *Let* $\mathbf{M}_1, \mathbf{M}_2$ *be two symmetric* $p \times p$ *dimensional matrices. A* $p \times p$ *simultaneous diagonalization matrix* $\mathbf{W}$ *satisfies*

$$\mathbf{W}\mathbf{M}_1\mathbf{W}^\top = \mathbf{I}_p \quad \textit{and} \quad \mathbf{W}\mathbf{M}_2\mathbf{W}^\top = \mathbf{D},$$

where $\mathbf{D}$ *is a diagonal matrix.*

The two conditions above are equivalently rewritten as

$$\mathbf{M}_2\mathbf{W}^\top = \mathbf{M}_1\mathbf{W}^\top\mathbf{D},$$

which can be identified as a generalized eigenproblem (GEVD) where $\mathbf{W}^\top$ contains the generalized eigenvectors as columns and $\mathbf{D}$ carries the generalized eigenvalues as its diagonal elements. The GEVD can always be solved, if $\mathbf{M}_1$ is positive-definite then the generalized eigenvalues are

all positive and the generalized eigenvectors are orthogonal with respect to the inner product defined by $\mathbf{M}_1$.

By further inversion of the involved matrices ($\mathbf{M}_1$ is from now on assumed to be positive definite) the GEVD can be formulated as an eigenproblem by

$$\mathbf{M}_1^{-1}\mathbf{M}_2 = \mathbf{W}^\top \mathbf{D}\mathbf{W}^{-\top},$$

where $\mathbf{W}^\top$ and $\mathbf{D}$ contain the eigenvectors (as columnvectors) and the eigenvalues (as diagonal elements) of $\mathbf{M}_1^{-1}\mathbf{M}_2$. Note that the latter matrix is not symmetric which leads to non-orthogonal eigenvectors. In this view it is also clear that the solution is unique if the eigenvalues of $\mathbf{M}_1^{-1}\mathbf{M}_2$ are pairwise distinct.

It is also possible to find the simultaneous diagonalizer by utilizing a two step procedure. In the first step the eigendecomposition of $\mathbf{M}_1 = \mathbf{V}\mathbf{D}_1\mathbf{V}^\top$ is performed and its square-root is computed by $\mathbf{M}_1^{-1/2} = \mathbf{V}\mathbf{D}_1^{-1/2}\mathbf{V}^\top$. Secondly, the eigendecomposition of $\mathbf{M}_1^{-1/2}\mathbf{M}_2\mathbf{M}_1^{-1/2} = \mathbf{U}\mathbf{D}_{1,2}\mathbf{U}^\top$ is performed which leads to the joint diagonalizer $\mathbf{W} = \mathbf{U}^\top\mathbf{M}_1^{-1/2}$. This two-step procedure reflects also the condition on $\mathbf{M}_1$ to be positive definite, otherwise the inversion of $\mathbf{M}_1$ in the first step would be impossible. Note that in a BSS context $\mathbf{M}_1$ and $\mathbf{M}_2$ are scatter matrices which leads to the fact that this two-step procedure is equal to the two-step outline in Section 3.3. The first step corresponds to whitening of the observable and the second step corresponds to the eigendecomposition of $\mathbf{M}_2$ evaluated on the whitened data.

Joint diagonalization Joint diagonalization is the extension of the simultaneous diagonalization principle for more than two symmetric matrices.

Definition 18 (Joint diagonalization problem)**.** *Let $\mathbf{M}_1, \ldots, \mathbf{M}_K$ be K symmetric $p \times p$ dimensional matrices with $K > 2$. A $p \times p$ joint diagonalization matrix $\mathbf{W}$ satisfies*

$$\mathbf{W}\mathbf{M}_k\mathbf{W}^\top = \mathbf{D}_k \quad \textit{for} \quad k = 1, \ldots, K,$$

where $\mathbf{D}_k$ are diagonal matrices.

As shown in Harville (1997, Chapter 21) exact joint diagonalization is only possible if the matrices commute, i.e.: $\mathbf{M}_k\mathbf{M}_l = \mathbf{M}_l\mathbf{M}_k$ for all $k, l = 1, \ldots, K$ and $k \neq l$. This is equivalent to the statement that the eigenvectors are equal for all matrices $\mathbf{M}_k$. The exact diagonalizer is unique if for each pair $i, j = 1, \ldots, p$ and $i \neq j$ there exists a $k \in \{1, \ldots, K\}$ for which the diagonal elements of $\mathbf{D}_k$ are non-equal $(\mathbf{D}_k)_{ii} \neq (\mathbf{D}_k)_{jj}$, see for example Belouchrani et al. (1997).

In statistical applications the set of matrices to be jointly diagonalized do usually commute on the population level but most often not on the sample level due to the estimation error. Therefore, algorithms that approximately jointly diagonalize the matrices at hand need to be utilized. Such methods find the joint diagonalizer by maximizing some diagonalization criterion where most often the squared Frobenius matrix norm of the diagonal elements

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

is maximized. To avoid finding the trivial solution ($\mathbf{W} = \mathbf{0}$) many algorithms use one of the K matrices to transform the remaining $K - 1$ by $\mathbf{M}_k^{tr} = \mathbf{M}_1^{-1/2}\mathbf{M}_k\mathbf{M}_1^{-1/2}$ for $k = 2, \ldots, K$ and

find an approximate joint diagonalizer $\mathbf{U}$ of the transformed matrices that is restricted to be orthogonal. The joint diagonalizer of the original matrices is then given by $\mathbf{W} = \mathbf{U}^\top \mathbf{M}_1^{-1/2}$ Note that for this procedure $\mathbf{M}_1$ additionally needs to be positive definite. Yeredor (2002) discusses procedures of finding the approximate joint diagonalizer without the orthogonality constraint.

The orthogonal joint diagonalizer can be either found deflation based (each column is found one by another) see Nordhausen et al. (2012) or each direction simultaneously. The latter approach is exclusively followed by Publications I - VI. Specifically, iteratively Givens rotations are performed on each 2×2 submatrix set until the angles of the rotations are very close to zero. See Clarkson (1988); Cardoso (1989) for details on this approach and Theis and Inouye (2006); Illner et al. (2015) for others.

For the BSS methods the set of matrices usually consists of scatter matrices. Therefore, the transformation step above can be seen as a whitening step with respect to $\mathbf{M}_1$ and the orthogonal joint diagonalizer is then found by diagonalizing the matrices $\mathbf{M}_2, \ldots, \mathbf{M}_K$ which are evaluated on the whitened data. This is equivalent to the outline in Section 3.3.

3.4.1. FOBI and JADE

Algebraic ICA methods use cumulants as the central information criterion for characterizing the distribution of the observable as outlined by Virta (2018). Similar as moments the cumulants are defined by the coefficients of the cumulant generating function

$$\log(M_{\mathbf{x}}(\mathbf{t})) \quad \text{with} \quad M_{\mathbf{x}}(\mathbf{t}) = \mathbb{E}\left(\exp(\mathbf{t}^\top \mathbf{x})\right).$$

In the above formula $M_{\mathbf{x}}(\mathbf{t})$ is the moment-generating function for a random vector $\mathbf{x}$. As the following methods rely on cumulants up to order four but usually one is used to work with moments the relationship between the cumulants and the moments are given in the following. For the marginal fourth order cumulants this yields

$$\kappa_2(x) = \mathbb{E}\left(x^2\right), \quad \kappa_3(x) = \mathbb{E}\left(x^3\right) \quad \text{and} \quad \kappa_4(x) = \mathbb{E}\left(x^4\right) - 3\mathbb{E}\left(x^2\right)^2.$$

In similar fashion the excess kurtosis of a random variable can be expressed in terms of its cumulants by

$$\beta_2 = \frac{\kappa_4}{\kappa_2^2},$$

which yields zero if the random variable is standard normal distributed, i.e.: $x \sim N(0,1)$. For the joint cumulants of random variables x_1, x_2, x_3 and x_4 it holds that

$$\begin{aligned}
\kappa(x_1, x_2) &= \mathbb{E}\left(x_1 x_2\right) \\
\kappa(x_1, x_2, x_3) &= \mathbb{E}\left(x_1 x_2 x_3\right) \\
\kappa(x_1, x_2, x_3, x_4) &= \mathbb{E}\left(x_1 x_2 x_3 x_4\right) - \mathbb{E}\left(x_1 x_2\right)\mathbb{E}\left(x_3 x_4\right) \\
&\quad - \mathbb{E}\left(x_1 x_3\right)\mathbb{E}\left(x_2 x_4\right) - \mathbb{E}\left(x_1 x_4\right)\mathbb{E}\left(x_2 x_3\right).
\end{aligned}$$

Based on the last line of the above equation all possible fourth order joint cumulants can be captured by the following p^2 matrices

$$\mathbf{K}_{ij}(\mathbf{x}) = \mathbb{E}\left(x_i^{wh} x_j^{wh} \mathbf{x}^{wh} \mathbf{x}^{wh\top}\right) - \mathbf{E}_{ij} - \mathbf{E}_{ij}^\top - \operatorname{tr}(\mathbf{E}_{ij})\mathbf{I}_p,$$

for $i, j = 1, \ldots, p$, where $\mathbf{E}_{ij}$ is the zero matrix where only the ij-th element caries an one and $\mathbf{x}^{wh}$ is the whitened version of $\mathbf{x}$ given by $\mathbf{x}^{wh} = \mathrm{Cov}^{-1/2}(\mathbf{x})(\mathbf{x} - \mathbb{E}(\mathbf{x}))$. As the number of different joint cumulant matrices rises quickly with the dimension of the data the information of the cumulants is often somewhat compressed by considering the matrix of fourth moments

$$\mathrm{Cov}_4(\mathbf{x}) = \sum_{i=1}^{p} \mathbf{K}_{ii}(\mathbf{x}) = \mathbb{E}\left(\mathbf{x}^{wh\top}\mathbf{x}^{wh}\mathbf{x}^{wh}\mathbf{x}^{wh\top}\right) - 3\mathbf{I}_p.$$

The key property of joint cumulants for the use in ICA states that if the argument of the joint cumulant contains (at least) two independent random variables then the cumulant equals zero. This property together with the IC model (Definition 7) leads to the fact that the joint cumulant matrices as well as the matrix of fourth moments is diagonal when evaluated on the source which motivates the diagonalization of this quantities as follows. More informations on the use of cumulants in the context of ICA is given by Miettinen et al. (2015); Virta (2018).

FOBI One of the first ICA methods is the fourth order blind identification (FOBI) method which is originally introduced by Cardoso (1989). It simultaneously diagonalizes the covariance matrix and the matrix of fourth moments as follows.

Definition 19 (Fourth order blind identification (FOBI))**.** *Let* $\mathbf{x}$ *be a* p*-variate random vector following the IC model (Definition 7). The FOBI unmixing matrix functional* $\mathbf{W} = \mathbf{W}(\mathbf{x})$ *satisfies*

$$\mathbf{W}\,\mathrm{Cov}(\mathbf{x})\mathbf{W}^\top = \mathbf{I}_p \quad \textit{and} \quad \mathbf{W}\,\mathrm{Cov}_4(\mathbf{x})\mathbf{W}^\top = \mathbf{D},$$

where $\mathbf{D}$ *is a diagonal matrix with decreasingly ordered diagonal elements.*

Note that in the FOBI method the order of the source components is determined by the decreasingly ordered diagonal elements of the matrix $\mathbf{D}$. As the matrix of fourth moments is orthogonal invariant it immediately follows that FOBI is affine equivariant (Miettinen et al., 2015). Additionally, the matrix of fourth moments evaluated on the source vector $\mathbf{z} = (z_1, \ldots, z_p)$ equals

$$\mathrm{Cov}_4(\mathbf{z}) = \sum_{i=1}^{p} (\kappa_4(z_i) + p + 2)\,\mathbf{E}_{ii},$$

which results in identifiable latent components if their fourth order cumulants are pairwise distinct. As the fourth order cumulants are closely related with the kurtosis (see the above outline) it can be equivalently stated that the kurtosis values of the source must be pairwise distinct. This identifiability assumption is stricter than (IC 3) from the IC model. For more details on FOBI see Miettinen et al. (2015); Nordhausen and Virta (2019).

JADE Joint approximate diagonalization of eigenmatrices (JADE) originally introduced by Cardoso and Souloumiac (1993) is an extension of the FOBI methods which jointly diagonalizes all p^2 joint cumulants matrices as follows.

Definition 20 (Joint approximate diagonalization of eigenmatrices (JADE))**.** *Let* $\mathbf{x}$ *be a* p*-variate random vector following the IC model (Definition 7). Define the whitened version of*

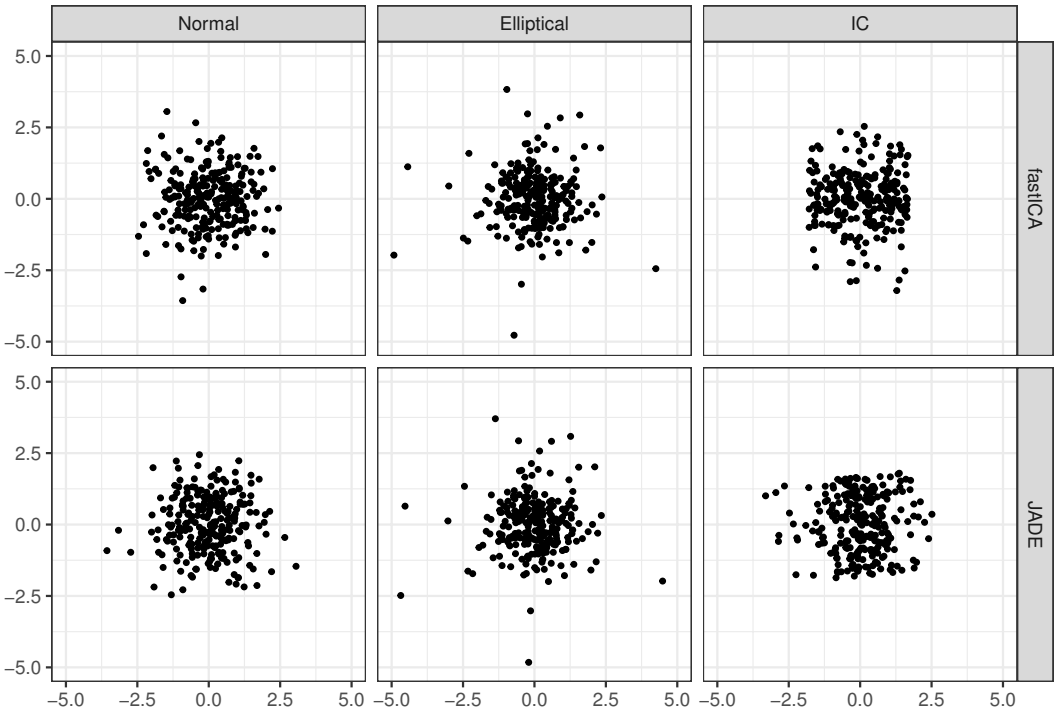

Figure 3.5.: Independent components of the data seen in the second row of Figure 3.1 which are found by the adaptive deflation-based fastICA or JADE method. Note that for the multivariate normal model and for the elliptical model the found independent components are only a random rotation of the whitened versions seen in the second row of Figure 3.4. For the IC model both methods find the true independent components up to sign and permutation.

$\mathbf{x}$ *by* $\mathbf{x}^{wh} = \mathrm{Cov}^{-1/2}(\mathbf{x})(\mathbf{x} - \mathbf{b})$. *Let* $\mathbf{U} = \mathbf{U}(\mathbf{x})$ *be the orthogonal* $p \times p$ *joint diagonalization matrix of* $\mathbf{K}_{ij}(\mathbf{x})$ *for* $i, j = 1, \ldots, p$, *maximizing*

$$\sum_{i,j=1}^{p} \|\mathrm{diag}(\mathbf{U}^\top \mathbf{K}_{ij}(\mathbf{x})\mathbf{U})\|_F^2.$$

The JADE unmixing matrix functional $\mathbf{W} = \mathbf{W}(\mathbf{x})$ *equals* $\mathbf{W} = \mathbf{U}^\top \mathrm{Cov}^{-1/2}(\mathbf{x})$.

As shown in Miettinen et al. (2015) JADE is affine equivariant and the source is identifiable if at most one latent component has zero fourth order cumulant (or equivalently kurtosis). The last row of Figure 3.5 illustrates the found independent components when FOBI is applied on the data seen in the second row of Figure 3.1.

As the number of joint cumulant matrices rises quickly with the dimension of the source Miettinen et al. (2013) introduce the k-JADE method. It uses the FOBI result for whitening and then diagonalizes all joint cumulant matrices $\mathbf{K}_{ij}$ for $i, j = 1, \ldots, p$ with the condition that $|i - j| < k$.

Related methods A related approach to find an unmixing matrix functional for the IC model (Definition 7) is followed by Oja et al. (2006); Taskinen et al. (2007); Nordhausen et al. (2016).

All publications consider simultaneous diagonalization of scatter matrices that posses the independence property defined as follows.

Definition 21 (Independence property)**.** *Consider any p-variate random vector* $\mathbf{x}$ *and a scatter matrix functional* $\mathbf{S} = \mathbf{S}(\mathbf{x})$ *(Definition 16). If* $\mathbf{S}(\mathbf{x})$ *is diagonal for any choice of* $\mathbf{x}$ *with independent components then it is said to posses the independence property.*

As proven in Oja et al. (2006); Nordhausen and Tyler (2015) scatter matrices symmetrized in terms of differences posses the independence property. A symmetrized version of a scatter matrix is given by

$$\mathbf{S}_s(\mathbf{x}) = \mathbf{S}(\mathbf{v} - \mathbf{w}),$$

where $\mathbf{v}$ and $\mathbf{w}$ are independent copies of the random vector $\mathbf{x}$. Naturally, the covariance matrix is already symmetrized as $\mathrm{Cov}(\mathbf{x}) = 2^{-1}\mathbb{E}\left((\mathbf{v} - \mathbf{w})(\mathbf{v} - \mathbf{w})^\top\right)$. For example a popular robust counterpart of the covariance matrix is the minimum determinant covariance (MCD) (Hubert et al., 2018) which does not posses the independence property unless it is symmetrized. Therefore, a procedure to construct robust BSS functionals can be based on robust scatter matrices which are evaluated on differences. Nordhausen and Tyler (2015) argued theoretically and in simulations that this procedure is valid. Note that the number of differences rises with the square of the sample size resulting in very expensive computations of the symmetrized versions.

3.4.2. AMUSE and SOBI

In contrast to the ICA methods BSS for stationary time series relies on the dependence between the random vectors at different time stamps. Under the weak stationarity assumptions the autocovariance can be written as

$$\mathbf{S}_\tau(\mathbf{x}_t) = \mathbb{E}\left((\mathbf{x}_{t+\tau} - \mathbb{E}(\mathbf{x}_t))(\mathbf{x}_t - \mathbb{E}(\mathbf{x}_t))^\top\right).$$

As the former quantity is not symmetric, (i.e.: $\mathbb{E}(x_{i,t+\tau}x_{j,t}) \neq \mathbb{E}(x_{j,t+\tau}x_{i,t})$) it is often symmetrized by

$$\mathbf{S}'_\tau(\mathbf{x}_t) = \frac{1}{2}\left(\mathbf{S}_\tau(\mathbf{x}_t) + \mathbf{S}_\tau^\top(\mathbf{x}_t)\right),$$

which forms the basis for the algorithm for multiple unknown signals extraction (AMUSE) and the second-order blind identification (SOBI) methods where both solve the SOS model seen in Definition 8.

AMUSE One of the first unmixing matrix functionals for the SOS model is the AMUSE method (Tong et al., 1990) that simultaneously diagonalizes the covariance matrix as well as one autocovariance matrix as follows.

Definition 22 (Algorithm for multiple unknown signals extraction (AMUSE))**.** *Let* $\mathbf{x}_t$ *be a p-variate time series following the SOS model (Definition 8). Let* $\tau > 0$ *be some lag. The AMUSE unmixing matrix functional* $\mathbf{W} = \mathbf{W}(\mathbf{x}_t)$ *satisfies*

$$\mathbf{W}\,\mathrm{Cov}(\mathbf{x}_t)\mathbf{W}^\top = \mathbf{I}_p \quad \textit{and} \quad \mathbf{W}\mathbf{S}'_\tau(\mathbf{x}_t)\mathbf{W}^\top = \mathbf{D},$$

where $\mathbf{D}$ *is a diagonal matrix with decreasingly ordered diagonal elements.*

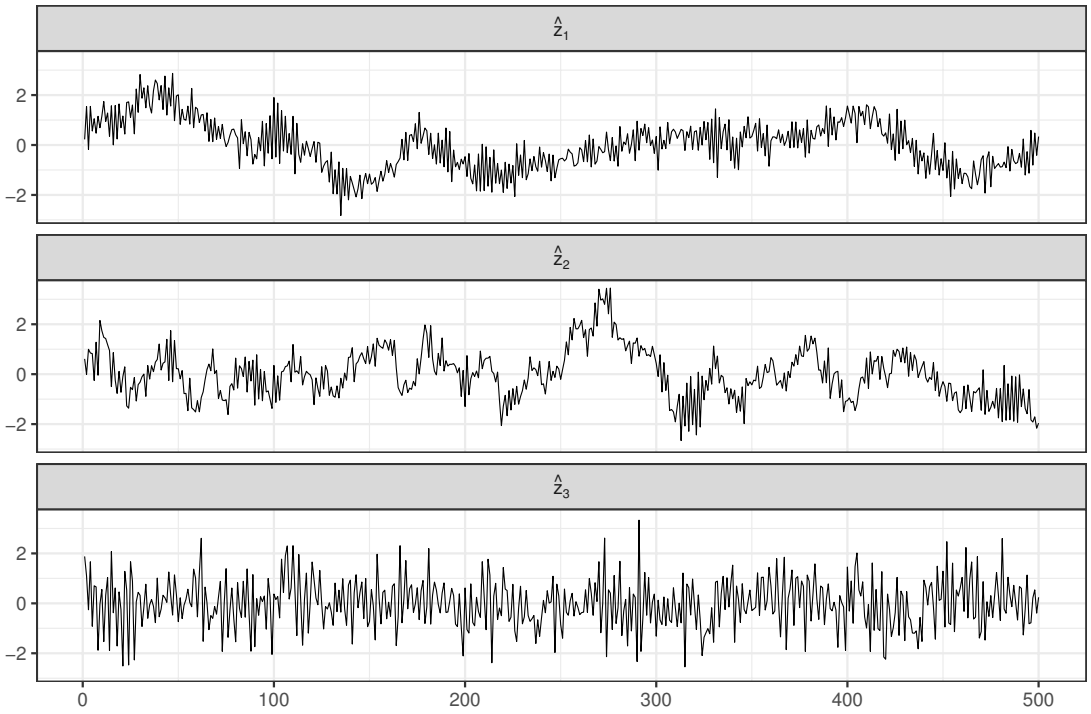

Figure 3.6.: Latent time series which are found by applying SOBI with the set of lags $\mathcal{L} = \{1, \ldots, 12\}$ on the data seen in Figure 3.3.

The AMUSE unmixing matrix functional is affine equivariant (Miettinen et al., 2012) and it is identifiable if the diagonal elements of $S'_\tau(\mathbf{z}_t)$ are pairwise distinct. In practical applications the considered lag is often chosen to be $\tau = 1$ as the autocovariances of the source components might be most different for this choice. Generally, the performance of AMUSE crucially depends on the choice of the lag, hence, the identifiability condition might be too restricting. Therefore, a method that jointly diagonalizes several autocovariances is introduced.

SOBI The SOBI methods (Belouchrani et al., 1997) solves the SOS model by diagonalizing several autocovariance matrices as follows.

Definition 23 (Second-order blind identification (SOBI))**.** *Let $\mathbf{x}_t$ be a p-variate time series following the SOS model (Definition 8). Define the whitened version of $\mathbf{x}_t$ by $\mathbf{x}_t^{wh} = \mathrm{Cov}^{-1/2}(\mathbf{x}_t)(\mathbf{x}_t - \mathbf{b})$. Let $\mathcal{L}$ be a finite set of strictly positive integer numbers which act as lags. Let $\mathbf{U} = \mathbf{U}(\mathbf{x}_t)$ be the orthogonal $p \times p$ joint diagonalization matrix of $\mathbf{S}'_\tau(\mathbf{x}_t^{wh})$ for $\tau \in \mathcal{L}$, maximizing*

$$\sum_{\tau \in \mathcal{L}} \|\mathrm{diag}(\mathbf{U}^\top \mathbf{S}'_\tau(\mathbf{x}_t^{wh})\mathbf{U})\|_F^2.$$

The SOBI unmixing matrix functional $\mathbf{W} = \mathbf{W}(\mathbf{x}_t)$ equals $\mathbf{W} = \mathbf{U}^\top \mathrm{Cov}^{-1/2}(\mathbf{x}_t)$.

The SOBI method is affine equivariant and the parameters are identifiable if for each pair $i, j = 1, \ldots, p$ with $i \neq j$ there exists a lag $\tau \in \mathcal{L}$ for which $(\mathbf{S}'_\tau(\mathbf{z}_t))_{ii} \neq (\mathbf{S}'_\tau(\mathbf{z}_t))_{jj}$ which is far less restricting than the AMUSE identifiability condition (Miettinen et al., 2016). In practical

applications often the lags from one to twelve are used, however, as pointed out by Tang et al. (2005) this choice is not always the optimal one. Figure 3.6 depicts the latent time series when SOBI is applied on the the data from Figure 3.3. More details on the SOBI (and AMUSE) method are given by Miettinen et al. (2012, 2016).

As the order of the latent components is not identifiable often the order of the sources is dictated by the decreasing order of the so-called pseudo-eigenvalues

$$\lambda_i^2 = \sum_{\tau \in \mathcal{L}} \left(\mathbf{u}_i^\top \mathbf{S}'_\tau(\mathbf{x}_t^{wh}) \mathbf{u}_i \right)^2,$$

where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$ meaning that $\mathbf{u}_i$ is the i-th column of the orthogonal joint diagonalizer. Based on this definition it might be viewed in the way that the order is determined by the found source components that show the highest serial second order dependence averaged over the chosen lags $\mathcal{L}$.

Related methods The SOS model and the above two unmixing matrix functionals are based on the use of second order time dependence structure and processes that are weakly stationary (e.g.: ARMA processes). Different forms of processes are more suited towards financial time series. Specifically, financial time series exhibit so-called stochastic volatility which leads to the fact that the information is captured in higher order time dependence. Certain time series models exhibiting stochastic volatility have a vanishing second-order serial dependence (e.g.: ARCH, GARCH or SV processes see for example Matteson and Ruppert (2011) and references below) which are clearly not captured by the SOS model leading to the fact that AMUSE or SOBI delivers meaningless results. To overcome this issue Matilainen et al. (2015) extended the IC model (Definition 7) by stating that the source time series are independent and introduce generalized FOBI and JADE (gFOBI and gJADE) which utilize joint cumulant matrices (and the matrix of fourth moments) for different lag combinations.

A different direction is taken by introducing non-linearities in the optimization problem of the original SOBI by

$$\sum_{\tau \in \mathcal{L}} \sum_{i=1}^{p} \left| \mathbb{E}\left(G(\mathbf{u}_i^\top \mathbf{x}_{t+\tau}^{wh}) G(\mathbf{u}_i^\top \mathbf{x}_t^{wh}) \right) - \mathbb{E}\left(G(\mathbf{u}_i^\top \mathbf{x}_{t+\tau}^{wh}) \right) \mathbb{E}\left(G(\mathbf{u}_i^\top \mathbf{x}_t^{wh}) \right) \right|^a,$$

where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$ is an orthogonal maximizer. For the choice of $a = 1$ the above function leads to the fixed-point algorithm for maximizing the non-linear autocorrelation (FixNA) and a variant of it (FixNA2) introduced by Hyvärinen (2001); Shi et al. (2009). For the choice $a = 2$ Matilainen et al. (2017b) introduce the variant of SOBI (vSOBI) method, this publication gives also statistical properties and comparisons of the three methods.

The most general method is introduced by Miettinen et al. (2020) which is designed for a source containing components that show stochastic volatility as well as second order serial dependence. This generalized SOBI (gSOBI) method combines the original SOBI with a vSOBI variant ($G(x) = x^2$) leading to

$$w \sum_{\tau_1 \in \mathcal{L}_1} \sum_{i=1}^{p} \left(\mathbf{u}_i^\top \mathbb{E}\left(\mathbf{x}_t \mathbf{x}_{t+\tau_1}^\top \right) \mathbf{u}_i \right)^2 + (1-w) \sum_{\tau_2 \in \mathcal{L}_2} \sum_{i=1}^{p} \left(\mathbb{E}\left((\mathbf{u}_i^\top \mathbf{x}_t)^2 (\mathbf{u}_i^\top \mathbf{x}_{t+\tau_2})^2 - 1 \right) \right)^2,$$

where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$ is an orthogonal maximizer. The left part of the above equation is the SOBI part which uses the lag set $\mathcal{L}_1$ and the right part is the vSOBI part using the lag set $\mathcal{L}_2$

where the strength of either part is controlled by the weight $w \in [0,1]$. The proposed default values are given by $\mathcal{L}_1 = \{1,\dots,12\}$, $\mathcal{L}_2 = \{1,2,3\}$ and $w = 0.9$ and Piccolotto et al. (2020) provide visual analytic tools for choosing appropriate parameters for the problem at hand.

Effort has been put in developing a robust counterpart to the classical SOBI estimator. To ensure robust BSS estimators the whitening step has to be adapted as well as the information criteria that are used in the rotation step. This needs robust location and scatter functionals that are diagonal for the source under the SOS model, as pointed out above and by Nordhausen and Tyler (2015) symmetrized scatter matrices can be used. Theis et al. (2010) used spatial sign autocovariance matrices defined by

$$\mathbf{S}_\tau^{sp}(\mathbf{x}_t) = \mathbb{E}\left(\frac{\mathbf{x}_{t+\tau}}{\|\mathbf{x}_{t+\tau}\|}\frac{\mathbf{x}_t}{\|\mathbf{x}_t\|}\right)$$

for the whitening as well as the rotation step. These matrices are robust as the absolute value of the random vector at time t is canceled by the division through the norm of it. This sign autocovariance SOBI (SAM-SOBI) method is not affine equivariant as the $\mathbf{S}_\tau^{sp}$ matrices are only orthogonal invariant which leads to problems in the whitening step. Ilmonen et al. (2015) introduce an affine equivariant extension denoted as eSAM-SOBI which uses the robust Hettmansperger-Randels estimates of location and scatter (Hettmansperger and Randles, 2002) that are affine equivariant.

3.4.3. NSS-SD, NSS-JD and NSS-TD-JD

The NSS model (Definition 9) allows for time varying variance as well as autocovariances that specifically depend on the time stamps. Therefore, the covariance is not equal along time which leads to the need of an adapted whitening procedure and additionally a new information criterion for the rotation step. For this purpose Choi and Cichocki (2000a,b) introduce the notion of (auto)covariances that are computed only for a subset of the whole time series. Specifically, for a finite set $\mathcal{T} = \{t_1, t_1+1, \dots, t_2\} \subset \mathbb{Z}$ with $t_1 < t_2$ the autocovariance for this interval can be defined by

$$\mathbf{S}_{\mathcal{T},\tau}(\mathbf{x}_t) = \frac{1}{|\mathcal{T}| - \tau}\sum_{t\in\mathcal{T}} \mathbb{E}\left((\mathbf{x}_{t+\tau} - \mathbb{E}(\mathbf{x}_t))(\mathbf{x}_t - \mathbb{E}(\mathbf{x}_t))^\top\right) I(t+\tau \in \mathcal{T}),$$

where again the mean is assumed to be constant (see the NSS model). As introduced above $\mathbf{S}'_{\mathcal{T},\tau}(\mathbf{x}_t)$ is the symmetrized version of the above quantity. For a lag of $\tau = 0$ this reduces to the average covariance in the finite sub time series $\mathcal{T}$

$$\mathbf{S}_{\mathcal{T},0}(\mathbf{x}_t) = \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} \mathbb{E}\left((\mathbf{x}_t - \mathbb{E}(\mathbf{x}_t))(\mathbf{x}_t - \mathbb{E}(\mathbf{x}_t))^\top\right),$$

which is also a reasonable choice for whitening if the set $\mathcal{T}$ is chosen to be the whole available time series. Based on the former quantities Choi and Cichocki (2000a,b) introduce the following three methods that solve the NSS problem.

NSS-SD Two average covariance matrices for two different sub time series are simultaneously diagonalized by the non-stationary source separation simultaneous diagonalization (NSS-SD) method.

Definition 24 (Non-stationary source separation - simultaneous diagonalization (NSS-SD))**.** *Let* $\mathbf{x}_t$ *be a p-variate time series following the NSS model (Definition 9). Consider two non-overlapping finite time intervals* $\mathcal{T}_1, \mathcal{T}_2 \subset \mathbb{Z}$*. The NSS-SD unmixing matrix functional* $\mathbf{W} = \mathbf{W}(\mathbf{x}_t)$ *satisfies*

$$\mathbf{W}\mathbf{S}_{\mathcal{T}_1,0}(\mathbf{x}_t)\mathbf{W}^\top = \mathbf{I}_p \quad \textit{and} \quad \mathbf{W}\mathbf{S}_{\mathcal{T}_2,0}(\mathbf{x}_t)\mathbf{W}^\top = \mathbf{D},$$

where $\mathbf{D}$ *is a diagonal matrix with decreasingly ordered diagonal elements.*

The NSS-SD functional is affine equivariant and the parameters are identifiable if the diagonal elements of $\mathbf{S}^{-1}_{\mathcal{T}_1,0}(\mathbf{z}_t)\mathbf{S}_{\mathcal{T}_2,0}(\mathbf{z}_t)$ are pairwise distinct (Nordhausen, 2014). Similar as in the stationary case the identifiability condition might be too restricting which leads to the following extension.

NSS-JD The non-stationary source separation joint diagonalization (NSS-JD) method jointly diagonalizes several average covariance matrices for different sub time series.

Definition 25 (Non-stationary source separation - joint diagonalization (NSS-JD))**.** *Let* $\mathbf{x}_t$ *be a p-variate time series following the NSS model (Definition 9). Consider* $K+1$ *finite time intervals* $\mathcal{T}, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_K \subset \mathbb{Z}$*, where* $\mathcal{T}_1, \ldots, \mathcal{T}_K$ *are non-overlapping and* $\mathcal{T}$ *corresponds to the whole considered time. Define the whitened version of* $\mathbf{x}_t$ *by* $\mathbf{x}_t^{wh} = \mathbf{S}^{-1/2}_{\mathcal{T},0}(\mathbf{x}_t)(\mathbf{x}_t - \mathbf{b})$*. Let* $\mathbf{U} = \mathbf{U}(\mathbf{x}_t)$ *be the orthogonal* $p \times p$ *joint diagonalization matrix of* $\mathbf{S}_{\mathcal{T}_k,0}$ *for* $k = 1, \ldots, K$*, maximizing*

$$\sum_{k=1}^{K} \|\mathrm{diag}(\mathbf{U}^\top \mathbf{S}_{\mathcal{T}_k,0}(\mathbf{x}_t^{wh})\mathbf{U})\|_F^2.$$

The NSS-JD unmixing matrix functional $\mathbf{W} = \mathbf{W}(\mathbf{x}_t)$ *equals* $\mathbf{W} = \mathbf{U}^\top \mathbf{S}^{-1/2}_{\mathcal{T},0}(\mathbf{x}_t)$*.*

Again the NSS-JD functional is affine equivariant but the identifiability condition is less restrictive as the parameters are identifiable if for each pair $i, j = 1, \ldots, p$ with $i \neq j$ the exists a $k \in \{1, \ldots, K\}$ for which $(\mathbf{S}^{-1}_{\mathcal{T},0}(\mathbf{z}_t)\mathbf{S}_{\mathcal{T}_k,0}(\mathbf{z}_t))_{ii} \neq (\mathbf{S}^{-1}_{\mathcal{T},0}(\mathbf{z}_t)\mathbf{S}_{\mathcal{T}_k,0}(\mathbf{z}_t))_{jj}$. The above two methods utilize only information from the covariance that is possibly different across time. Information useful for the signal separation might also be present in the autocovariance structure which is exploited in the following method.

NSS-TD-JD The non-stationary source separation time delayed joint diagonalization (NSS-TD-JD) method jointly diagonalizes many autocovariance matrices for different sub time series.

Definition 26 (Non-stationary source separation - time delayed - joint diagonalization (NSS-TD-JD))**.** *Let* $\mathbf{x}_t$ *be a p-variate time series following the NSS model (Definition 9). Consider* $K+1$ *finite time intervals* $\mathcal{T}, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_K \subset \mathbb{Z}$*, where* $\mathcal{T}_1, \ldots, \mathcal{T}_K$ *are non-overlapping and* $\mathcal{T}$ *corresponds to the whole considered time and a finite set of non-negative integers* $\mathcal{L}$ *acting as lags. Define the whitened version of* $\mathbf{x}_t$ *by* $\mathbf{x}_t^{wh} = \mathbf{S}^{-1/2}_{\mathcal{T},0}(\mathbf{x}_t)(\mathbf{x}_t - \mathbf{b})$*. Let* $\mathbf{U} = \mathbf{U}(\mathbf{x}_t)$ *be the orthogonal* $p \times p$ *joint diagonalization matrix of* $\mathbf{S}'_{\mathcal{T}_k,\tau}$ *for* $k = 1, \ldots, K$ *and all* $\tau \in \mathcal{L}$*, maximizing*

$$\sum_{k=1}^{K} \sum_{\tau \in \mathcal{L}} \|\mathrm{diag}(\mathbf{U}^\top \mathbf{S}'_{\mathcal{T}_k,\tau}(\mathbf{x}_t^{wh})\mathbf{U})\|_F^2.$$

The NSS-TD-JD unmixing matrix functional $\mathbf{W} = \mathbf{W}(\mathbf{x}_t)$ *is written as* $\mathbf{W} = \mathbf{U}^\top \mathbf{S}^{-1/2}_{\mathcal{T},0}(\mathbf{x}_t)$*.*

The NSS-TD-JD functional is affine equivariant and the parameters are identifiable if for each pair $i, j = 1, \ldots, p$ with $i \neq j$ the exists a $k \in \{1, \ldots, K\}$ and $\tau \in \mathcal{L}$ for which $(\mathbf{S}^{-1}_{\mathcal{T},0}(\mathbf{z}_t)\mathbf{S}_{\mathcal{T}_k,\tau}(\mathbf{z}_t))_{ii} \neq (\mathbf{S}^{-1}_{\mathcal{T},0}(\mathbf{z}_t)\mathbf{S}_{\mathcal{T}_k,\tau}(\mathbf{z}_t))_{jj}$. This is the most general condition for these three different estimators. Note that the NSS-SD and the NSS-JD are special cases of the NSS-TD-JD method and if in the set of lags $\mathcal{L}$ the element zero is contained then NSS-TD-JD is the most general method. However, the number of matrices that need to be jointly diagonalized equals $|\mathcal{L}|K$ which might lead to noise for the utilized approximate joint diagonalization algorithms. In similar fashion when only one time interval is considered ($K = 1$) then NSS-TD-JD reduces to AMUSE/SOBI (Definition 22/23) if the set of lags consists of one/many elements. Furthermore, to properly estimate the autocovariance matrices weak stationary needs to hold in each of the intervals $\mathcal{T}_k$. In practical applications the choice of this intervals is not clear a-priori, often the time series is divided into a small number of equal length intervals in order to keep the sample size (for each interval) relatively high for proper estimation of the involved quantities.

Related methods Nordhausen (2014) introduces robust counterparts for the above three methods by again using spatial sign matrices and the Hettmansperger-Randels estimates for whitening as done in the stationary counterpart method eSAM-SOBI described above.

3.4.4. SIR in a blind source separation context

One of the first SDR methods is sliced inverse regression (SIR) originally introduced by (Li, 1991). It is based on inverse regression meaning that the role of the predictor $\mathbf{x}$ and the regressor y are interchanged. The key quantity in SIR is the covariance of the inverse regression curve

$$\operatorname{Cov}(\mathbb{E}(\mathbf{x}|y)).$$

Under the model assumptions of Definition 10 (and also the less restrictive version stated in the original SIR publication) it holds that

$$\operatorname{Cov}(\mathbb{E}(\mathbf{z}|y)) = \begin{pmatrix} \operatorname{Cov}(\mathbb{E}(\mathbf{z}_1|y)) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

which motivates to diagonalize this quantity evaluated on the whitened vector $\mathbf{x}$ leading to the SIR method embedded in the BSS methodology as described in Matilainen et al. (2017a).

Definition 27 (SIR in a BSS framework)**.** *Let $(y, \mathbf{x}^\top)^\top$ be a $(p+1)$-variate random vector following the BSS SDR model (Definition 10). The SIR unmixing matrix functional $\mathbf{W} = \mathbf{W}((y, \mathbf{x}^\top)^\top)$ satisfies*

$$\mathbf{W}\operatorname{Cov}(\mathbf{x})\mathbf{W}^\top = \mathbf{I}_p \quad \textit{and} \quad \mathbf{W}\operatorname{Cov}(\mathbb{E}(\mathbf{x}|y))\mathbf{W}^\top = \mathbf{D},$$

where $\mathbf{D}$ is a diagonal matrix with decreasingly ordered diagonal elements.

Computation of the inverse regression curve is rather complicated. Therefore, the real line is usually split into K non-overlapping intervals ($\mathbb{R} = \mathbb{R}_1 \cup \cdots \cup \mathbb{R}_K$) and the regressor is replaced by its sliced counterpart $y^{sl}(y) = \sum_{k=1}^{K} y_k I(y \in \mathbb{R}_k)$ for some choices y_k. Often ten slices are used and y_k are chosen to be the sample mean of each slice. The found components $\mathbf{Wx}$ are whitened, the components of the corresponding inverse regression curve $\mathbb{E}(\mathbf{Wx}|y)$

are uncorrelated and their order is given by their variances contained as diagonal elements of the matrix $\mathbf{D}$. Moreover, the variances hint the importance of the dependence between the regressor and the components of $\mathbf{Wx}$ and the last $p - r$ values should be zero under the model assumptions. Tests for the actual dimension of r are discussed in Bura and Cook (2001); Nordhausen et al. (2021b). Note that the original SIR method is equal to the BSS SIR method above. The core difference is that the model in the original SIR publication is less restrictive than the one state in Definition 10.

For the time series SDR model seen in Definition 11 Matilainen et al. (2017a) extended the former framework to also consider lagged inverse regression curves as the serial dependence present in time series data is something to consider naturally. Specifically, the quantities of interest write as

$$\operatorname{Cov}(\mathbb{E}\left(\mathbf{x}_t|y_{t+\tau}\right)) \quad \text{for all} \quad \tau \geq 0.$$

As seen before the model assumptions lead to

$$\operatorname{Cov}(\mathbb{E}\left(\mathbf{z}_t|y_{t+\tau}\right)) = \begin{pmatrix} \operatorname{Cov}(\mathbb{E}\left(\mathbf{z}_{1,t}|y_{t+\tau}\right)) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{for all} \quad \tau \geq 0,$$

which motivates to jointly diagonalize a number of such matrices as introduced by Matilainen et al. (2017a).

Definition 28 (Time series SIR in a BSS framework (TSIR))**.** *Let $(y_t, \mathbf{x}_t^\top)^\top$ be a $(p+1)$-variate time series following the time series BSS SDR model (Definition 11). Define the whitened version of $\mathbf{x}_t$ by $\mathbf{x}_t^{wh} = \operatorname{Cov}^{-1/2}(\mathbf{x}_t)(\mathbf{x}_t - \mathbf{b})$. Let $\mathcal{L}$ be a finite set of non-negative integer numbers. Let $\mathbf{U} = \mathbf{U}((y_t, \mathbf{x}_t^\top)^\top)$ be the orthogonal $p \times p$ joint diagonalization matrix of $\operatorname{Cov}\left(\mathbb{E}\left(\mathbf{x}_t^{wh}|y_{t+\tau}\right)\right)$ for $\tau \in \mathcal{T}$, maximizing*

$$\sum_{\tau \in \mathcal{L}} \left\| \operatorname{diag}\left(\mathbf{U}^\top \operatorname{Cov}\left(\mathbb{E}\left(\mathbf{x}_t^{wh}|y_{t+\tau}\right)\right)\mathbf{U}\right)\right\|_F^2.$$

The TSIR unmixing matrix functional $\mathbf{W} = \mathbf{W}((y_t, \mathbf{x}_t^\top)^\top)$ is written as $\mathbf{W} = \mathbf{U}^\top \operatorname{Cov}^{-1/2}(\mathbf{x}_t)$.

In the same fashion as in the iid case the response time series y_t is sliced to estimate the different lagged inverse regression curves. The key advantage of jointly diagonalizing the covariance matrices of the lagged inverse regression curves is given by the fact that the squared pseudo-eigenvalues

$$\lambda_{i,\tau} = \left(\mathbf{u}_i^\top \operatorname{Cov}\left(\mathbb{E}\left(\mathbf{x}_t^{wh}|y_{t+\tau}\right)\right)\mathbf{u}_i\right)^2$$

hold the desired information about the dependence structure between the response and the latent time series. The components of $\mathbf{Wx}_t$ are order according to $\lambda_{i,\cdot} = \sum_{\tau \in \mathcal{L}} \lambda_{i,\tau}$ where high values indicate a strong dependence between the component and the response time series. High values of $\lambda_{i,\tau}$ indicate a strong dependence between the i-th found component and the lagged response time series $y_{t+\tau}$. Matilainen et al. (2017a) introduce several strategies of finding the most important components and lag combinations for further regression analysis based on the pseudo-eigenvalues.

Related methods One disadvantage is that the SIR methodology fails completely when the relationship between the response and source is symmetric (i.e.: $y = f(\mathbf{z}_1) + \epsilon$ where f is symmetric in its argument). To overcome this issue Cook (2000) replace the inverse regression

curve $\mathbb{E}(\mathbf{x}|y)$ by the conditional variance $\mathrm{Cov}(\mathbf{x}|y)$ where the resulting method is denoted as sliced average variance estimation (SAVE). In the same fashion as in the above outline for SIR Matilainen et al. (2019) reformulate SAVE in a BSS framework. Additionally, the publication introduces its time series variant TSAVE as well as a novel estimator that combines TSIR and TSAVE denoted as time series SIR and SAVE hybrid (TSSH) which uses a combination of the variation of the inverse regression curve and the conditional variance.

3.5. Determining the signal subspace in noisy models

For the INM (Definition 12) as well as the ENM (Definition 13) the first goal is to find the signal dimension q which can be done for example as follows.

3.5.1. Asymptotic and bootstrap tests and ladle estimator for the internal noise model

Considering the INM (Definition 12) the main interest is to find the number of signal components of the latent time series. This can be achieved by formulating a statistical test with the following null hypothesis for $q \in \{0, \cdots, p-1\}$

$$H_0 : \text{There are exactly } p-q \text{ white noise time series in } \mathbf{z}_t.$$

A procedure for testing this null hypothesis as considered in Virta and Nordhausen (2021) can be motivated as follows. Under sufficient conditions (stated in the former publication), for a given finite set of strictly positive integers (set of lags) $\mathcal{L}$ a given observable time series of length T the AMUSE or SOBI method (Definition 22 or 23) estimates an unmixing matrix $\hat{\mathbf{W}}$ which makes the matrices

$$\hat{\mathbf{D}}_\tau = \hat{\mathbf{W}}\hat{\mathbf{S}}'_\tau(\mathbf{x}_t)\hat{\mathbf{W}}^\top \text{ for all } \tau \in \mathcal{L},$$

as diagonal as possible. Under the null the last $p-q$ latent components do not exhibit second order serial dependence. This motivates to split the above matrices into blocks leading to the form

$$\hat{\mathbf{D}}_\tau = \begin{pmatrix} \hat{\mathbf{D}}_\tau^{qq} & \hat{\mathbf{D}}_\tau^{q0} \\ \hat{\mathbf{D}}_\tau^{0q} & \hat{\mathbf{D}}_\tau^{00} \end{pmatrix} \text{ for all } \tau \in \mathcal{L},$$

where $\hat{\mathbf{D}}_\tau^{qq}$ is a $q \times q$ and $\hat{\mathbf{D}}_\tau^{00}$ is a $(p-q)\times(p-q)$ matrix corresponding to the $(p-q)\times(p-q)$ matrix $\hat{\mathbf{S}}'_\tau(\hat{\mathbf{z}}_{ns,t})$. The latter quantity should be (approximately) the zero matrix as it corresponds to the (estimated) autocovariance at lag τ of the estimated noise part $\hat{\mathbf{z}}_{ns,t}$ which is white noise under the null (it does not exhibit serial dependence). Therefore, the following test statistics can be motivated

$$t_q = \frac{1}{|\mathcal{L}|(p-q)^2} \sum_{\tau \in \mathcal{L}} \|\hat{\mathbf{D}}_\tau^{00}\|_F^2. \tag{3.2}$$

The distribution of the null hypothesis can be either approximated by bootstrap strategies or by large sample theory results (both strategies are discussed below) and the test can be carried out for a specific number of signal components q.

Estimation of the actual signal dimension can be obtained by sequentially carrying out the tests for different hypothetical noise dimensions where Matilainen et al. (2018); Virta and Nordhausen (2021) consider a forward, backward and divide-and-conquer strategy. A forward

strategy starts with $q = 0$ and decreases this number by one until the first test accepts the null hypothesis, a backward strategy starts with $q = p - 1$ and the signal dimension is estimated by $q' - 1$ where q' is the hypothetical noise dimension of the first rejecting test. Another approach is the divide-and-conquer strategy which halves the interval of possible hypothetical noise dimensions until the border between rejection and acceptance of the tests is found.

Asymptotic results Virta and Nordhausen (2021) provide and proof the following result based on the former outline.

Proposition 1. *Under sufficient conditions it holds that*

$$T|\mathcal{L}|(p-q)^2 t_q \xrightarrow{\mathcal{L}} \chi^2_{|\mathcal{L}|(p-q)(p-q+1)/2} \text{ as } n \to \infty,$$

where χ^2_ν denotes the chi-square distribution with ν degrees of freedom.

The sufficient conditions state that each signal time series has at least for one lag $\tau \in \mathcal{L}$ non-vanishing autocovariance and that the latent time series is a linear process having a MA(∞) representation. The exact conditions are found in the original publication. The closed form asymptotic distribution of the test statistic can be easily used to carry out an (approximate) test for the signal dimension q.

Virta and Nordhausen (2021) provide also a result that the signal dimension can be consistently estimated. However, the practical use of this result is rather limited. Therefore, estimation of the signal dimension is based on strategies that apply sequential hypothesis test as described above.

Bootstrap strategies The bootstrap is a non-parametric statistical tool which is developed to estimate the distribution or population quantities (e.g. the mean squared error (MSE)) of some considered estimator. Details on the bootsrap can be found in monographs such as Lahiri (2003). In the following the idea of the bootstrap is explained for the iid case.

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with common cdf F and let $X_1, \ldots, X_n$ be the n random variables which generate the sample of size n. Furthermore, θ is a parameter of the random process of interest (hence $\theta = \theta(F)$) and $\hat{\theta}_n = g(X_1, \ldots, X_n; F)$ is a sample estimate of θ. The goal is to estimate population quantities or the distribution of $\hat{\theta}_n$ without stating any assumptions on the unknown cdf F based on the sample alone. This is done by firstly replacing F with an estimate F_n, usually F_n is chosen to be the empirical cumulative distribution function (ecdf) defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x).$$

Secondly, the original sample is replaced by a bootstrap sample $X_1^*, \ldots, X_n^*$ which is generated by using the estimator F_n (for simplicity the number or replicates is chosen to be the sample size). Lastly, the distribution of the bootstraped quantity $\hat{\theta}_n^* = g(X_1^*, \ldots, X_n^*; F_n^*)$ denoted as G_n is studied. In practice, as the number of distinct bootstrap samples is given by n^n and the distribution of $\hat{\theta}_n^*$ depends on this additional variation of the bootstrapped sample usually G_n is studied approximately by making use of the Gliwenko-Cantelli theorem. Namely, repeating

the sampling procedure B times leading to $X^*_{1,b}, \ldots, X^*_{n,b}$ and $\hat{\theta}^*_{n,b} = g(X^*_{1,b}, \ldots, X^*_{n,b}; F^*_n)$ for $b = 1, \ldots, B$ then it holds that

$$\frac{1}{B} \sum_{b=1}^{B} I(\hat{\theta}^*_{n,b} \leq x) \approx G_n(x).$$

In this view the distribution of $\hat{\theta}^*_n$ can be approximated with higher precision by simply drawing an increasing number of bootstrap samples. Nordhausen et al. (2017) use this iid bootstrap strategy in the context of the IC model and FOBI.

Matilainen et al. (2018) consider the following strategy for the time series setting. The aim is to investigate the distribution of the test statistic under the null hypothesis H_q which can be used to estimate p-values. Namely, the quantity of interest is a bootstrap estimate of

$$P(t_q \leq x).$$

Naturally, for a time series $\mathbf{x}_t$ following the INM (Definition 12) and in the outline above the test statistic $t_q = t_q(\mathbf{x}_t)$ is a function of $\mathbf{x}_t$. A procedure to generate bootstrap samples $\mathbf{x}^*_t$ under the null hypothesis that the last $p - q$ components of the source are white noise can be formulated in the following three step procedure.

Step 1 The source time series $\hat{\mathbf{z}}_t$ and the unmixing matrix $\hat{\mathbf{W}}$ are estimated by SOBI (or AMUSE) based on the set of lags $\mathcal{L}$ and the corresponding test statistic $t_q(\mathbf{x}_t)$ is computed.

Step 2 The last $p - q$ components of $\hat{\mathbf{z}}_t$ are white noise under the null hypothesis. Hence, they are replaced by a bootstrap sample using one of the four subsequently described strategies.

Step 3 The bootstrap sample $\hat{\mathbf{z}}^*_t$ generated in Step 2 is mixed again by $\hat{\mathbf{x}}^*_t = \hat{\mathbf{W}}^{-1}\hat{\mathbf{z}}^*_t$ which leads to a bootstrap sample $\hat{\mathbf{x}}^*_t$ of the original time series $\mathbf{x}_t$. $\hat{\mathbf{x}}^*_t$ is used to compute the test statistic $t^*_q(\hat{\mathbf{x}}^*_t)$.

In order to estimate the p-value for H_0 Steps 2 and 3 are repeated B times leading to $t^*_{q,b}(\hat{\mathbf{x}}^*_{t,b})$ for $b = 1, \ldots, B$. The p-value is then simply given by the ratio of bootstraped test statistics that are larger as the original one

$$\hat{p} = \frac{\#(t^*_{q,b} \geq t_q) + 1}{B + 1},$$

where the plus one in the numerator and denominator account for possible numerical instabilities ensuring that a p-value of zero is impossible.

Matilainen et al. (2018) introduces four different approaches of generating a bootstrap sample for Step 2 as follows.

Parametric The noise is assumed to be iid standard normal. Therefore, bootstrap samples are drawn from $N(0, 1)$.

Non-parametric 1 The above parametric assumption is relaxed by only assuming that the noise part is an iid realization of some distribution. Hence, the bootstrap sample is drawn from the ecdf of the joint $T(p - q)$ noise vector.

Non-parametric 2 Here the above parametric assumption is relaxed by assuming that each individual noise component is an iid realization of some (not necessarily equal) distribution. Therefore, each noise component is bootstraped from the individual ecdf's.

Non-parametric 3 Here it is assumed that the $p-q$ noise vector is a realization of some multivariate (not serially dependent) distribution. Leading to a the bootstrap sample drawn from the multivariate ecdf of the $p-q$ noise samples.

Note that the above approaches only resample the hypothetical noise part, variation from the time series structure is not accounted. Probably a more suited strategy would be to additionally draw a time series bootstrap sample after the noise part is manipulated.

Equally as in the asymptotic testing strategy the estimation of the signal dimension can be carried out by sequentially testing for different signal dimension, e.g.: with the divide-and-conquer strategy discussed above.

Signal dimension estimation based on the variation of eigenvectors Nordhausen and Virta (2018) estimate the signal dimension q by focusing on the variation of eigenvectors when the corresponding eigenvalues are degenerated. Specifically, the last $p-q$ orthogonal vectors found in the rotation step of the SOBI or AMUSE method cannot be recovered as the last $p-q$ eigenvalues of the symmetrized autocovariance matrices are all zero under the null (as the last $p-q$ components are white noise). Hence, when drawing bootstrap samples the variation of the $p-q$ last orthogonal vectors should be much higher than the variation for the first q ones. This procedure is denoted as ladle strategy and originally formulated by Luo and Li (2016) in an iid context.

For this approach the original iid bootstrap is not suitable as the serial dependence would be completely destroyed when drawing a bootstrap sample from the ecdf of the data. Therefore, Kunsch (1989) introduces the block bootstrap method where the set of bootstrap candidates are not the individual observations for each time stamp but rather sub time series of equal length that either overlap or non-overlap. However, this approach does not necessarily generate bootstrap samples that are second order stationary which is desired as the SOS model is considered. It turns out that relaxing the fixed block length to rather generate sub time series of different lengths where the lengths are samples from the geometric distribution is enough to keep the bootstrapped time series stationary. This procedure is denoted as stationary bootstrap and detailed in Politis and Romano (1994). Nordhausen and Virta (2018) use both bootstrap strategies in the context of the ladle estimate and carry out an extensive simulation study.

3.5.2. PCA and robust whitening for the external noise model

Considering the ENM (Definition 13) with $q < p$ an orthogonal transformation can be found by using PCA and discarding the last $p-q$ components. For that purpose Virta and Nordhausen (2019) replace (ENM 4) by the stricter assumption that $\boldsymbol{\epsilon}_t$ has a spherical distribution centered around zero and that its covariance matrix is of the form $\mathbf{R} = \sigma \mathbf{I}_p$. With this stricter assumption the eigenvalues of the covariance matrix $\mathrm{Cov}(\mathbf{x}_t)$ equal $\lambda_1 + \sigma \geq \cdots \geq \lambda_q + \sigma > \sigma = \cdots = \sigma$ where the values λ_i for $i = 1, \ldots, q$ are the squared singular values of the $p \times q$ mixing matrix $\mathbf{A}$. These eigenvalues can be either used in graphical diagnostic tools such as the scree-plot or Virta and Nordhausen (2019) formulate statistical tests for q based on asymptotic results or bootstrap strategies. The orthogonal transformation given by the first q eigenvectors of

$\mathrm{Cov}(\mathbf{x}_t)$ leads to components that still cover the signal but discard meaningless components of the original time series.

For the case of $q = p$ (this case can also be achieved by firstly using the former strategies) the main aim is to estimate the unmixing matrix by avoiding the influence of the external noise. For this purpose Belouchrani and Cichocki (2000); Georgiev and Cichocki (2001) introduce a robust whitening procedure which accounts for the fact that the regular covariance matrix equals

$$\mathrm{Cov}(\mathbf{x}_t) = \mathbf{A}\,\mathrm{Cov}(\mathbf{z}_t)\mathbf{A}^\top + \mathbf{R},$$

where $\mathrm{Cov}(\mathbf{z}_t)$ is a positive definite diagonal matrix. The above equation shows that whitening with respect to the standard covariance matrix is corrupted by the additional covariance term of the noise $\mathbf{R}$. In contrast, (symmetrized) autocovariance matrices are not influenced by the noise as

$$\mathbf{S}'_\tau(\mathbf{x}_t) = \mathbf{A}\mathbf{S}'_\tau(\mathbf{z}_t)\mathbf{A}^\top \quad \text{for all} \quad \tau > 0,$$

where $\mathbf{S}'_\tau(\mathbf{z}_t)$ is a positive semi-definite diagonal matrix. In that view whitening can be carried out with respect to an autocovariance matrix. However, this is not always possible as it is not ensured that $\mathbf{S}'_\tau(\mathbf{z}_t)$ is positive definite which results in the fact that $\mathbf{S}'_\tau(\mathbf{x}_t)$ cannot be inverted. To overcome this issue a positive definite linear combination of autocovariance matrices

$$\sum_{i=1}^{|\mathcal{L}|} \alpha_i \mathbf{S}'_\tau(\mathbf{x}_t)$$

can be found, where $\mathcal{L}$ is a finite set of positive integers (the set of lags). Belouchrani and Cichocki (2000); Georgiev and Cichocki (2001) outline an iterative algorithm to find the parameters α_i which is ensured to converge in a finite number of steps as outlined in Tong et al. (1991). Note that in the ENM the source time series can never be estimated as $\mathbf{W}\mathbf{x}_t = \mathbf{z}_t + \mathbf{W}\boldsymbol{\epsilon}_t$ (up to the standard ambiguities of BSS) where the (transformed) noise part $\mathbf{W}\boldsymbol{\epsilon}_t$ is still present.

4. Spatial blind source separation

The following part forms the core of the present thesis. It is devoted to combine the geostatistical principles seen in Section 2 with the BSS principles discussed in Section 3 by summarizing the main contributions of Publications I-VI. Specifically, Section 4.1 summarizes the findings of Nordhausen et al. (2015); Bachoc et al. (2020) and Publication I which considers BSS for stationary random fields (SBSS) and a properly adapted INM with concepts of estimating the signal subspace. Section 4.2 discusses implications for the SBSS methodology when a non-constant drift is present (Publication II and III) and when the second order (spatial) dependence is non-stationary (Publication IV). Lastly, Section 4.3 outlines how the unmixing matrix functionals introduced in Section 4.1 and 4.2 can be useful in spatial prediction (Publication V) and how SIR can be adapted to the spatial case (Publication VI).

4.1. Stationary spatial source separation

Sections 4.1.1 and 4.1.2 introduce the statistical model of BSS for stationary random fields based on Nordhausen et al. (2015); Bachoc et al. (2020). In Section 4.1.3 the INM is extended to the spatial stationary BSS model as outlined in Publication I.

4.1.1. Stationary spatial blind source separation model

SBSS is originally introduced by Nordhausen et al. (2015), it is refined and theoretically studied by Bachoc et al. (2020). Both publications assume the following statistical model.

Definition 29 (Stationary spatial blind source separation (SBSS) model)**.** *Consider a p-variate random field $\mathbf{x}(\mathbf{s})$ which is defined on a spatial domain $\mathcal{S} \subseteq \mathbb{R}^d$ of dimension d. $\mathbf{x}(\mathbf{s})$ follows a stationary spatial blind source separation model if it can be written as*

$$\mathbf{x}(\mathbf{s}) = \mathbf{A}\mathbf{z}(\mathbf{s}) + \mathbf{b},$$

where $\mathbf{x}(\mathbf{s})$ is the p-variate observable random field, $\mathbf{z}(\mathbf{s})$ is the p-variate latent random field, $\mathbf{A}$ is the deterministic $p \times p$ mixing matrix of full-rank and $\mathbf{b}$ is a p-variate deterministic location vector. Furthermore, $\mathbf{z}(\mathbf{s})$ fulfills the following two conditions.

(SBSS 1) $\mathbb{E}(\mathbf{z}(\mathbf{s})) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{z}(\mathbf{s})) = \mathbb{E}\left(\mathbf{z}(\mathbf{s})\mathbf{z}(\mathbf{s})^\top\right) = \mathbf{I}_p$ *for all* $\mathbf{s} \in \mathcal{S}$ *and*

(SBSS 2) $\mathrm{Cov}(\mathbf{z}(\mathbf{s}), \mathbf{z}(\mathbf{s}')) = \mathbb{E}\left(\mathbf{z}(\mathbf{s})\mathbf{z}(\mathbf{s}')^\top\right) = \mathbf{D}_\mathbf{h}$, *where* $\mathbf{h} = \mathbf{s} - \mathbf{s}'$, $\forall\ \mathbf{s}, \mathbf{s}' \in \mathcal{S},\ \mathbf{s} \neq \mathbf{s}'$ *and* $\mathbf{D}_\mathbf{h}$ *is a diagonal matrix containing the (univariate) covariance functions for each entry of $\mathbf{z}(\mathbf{s})$ on its diagonal elements which depend only on $\mathbf{h}$.*

The SBSS model is based on the location-scatter model as seen in Equation (3.1) it can be viewed as a direct extension of the SOS model (Definition 8) for stationary time series to stationary random fields. As usual (SBSS 1) states on-site moment assumptions which fix the mean to be zero and the scale to be identity. (SBSS 2) states that the source vector $\mathbf{z}(\mathbf{s})$

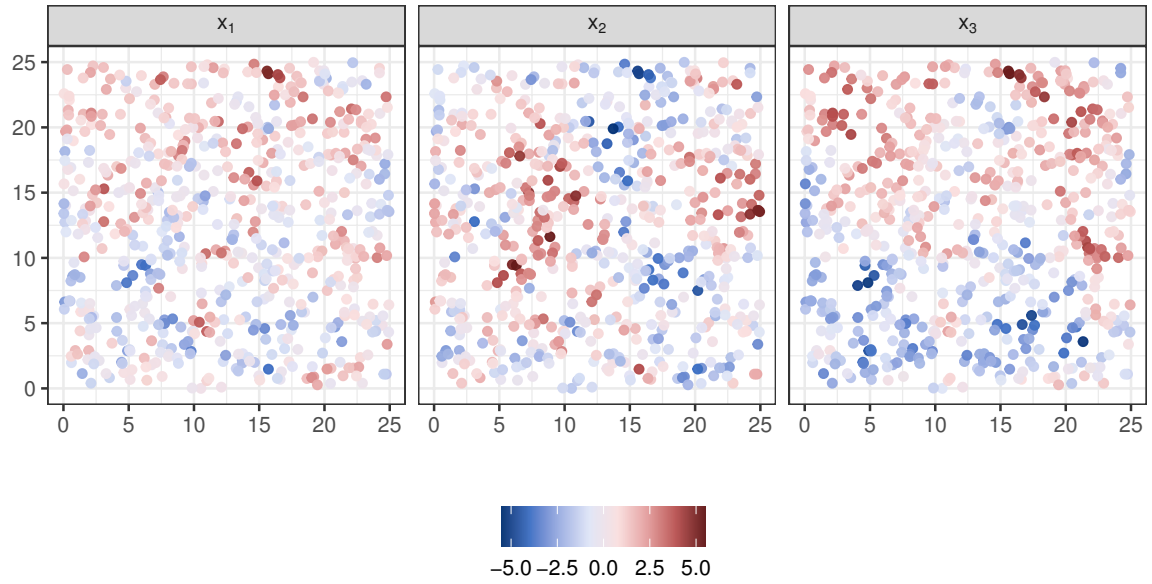

Figure 4.1.: Example for a SBSS model which is formed by mixing the random fields seen in Figure 2.5. The entries of the mixing matrix are drawn iid from $N(0,1)$ and the location vector is set to $\mathbf{b} = \mathbf{0}$.

contains stationary random fields, leading to the fact that also $\mathbf{x}(\mathbf{s})$ is stationary as the location vector $\mathbf{b}$ is globally constant. In total, (SBSS 1) and (SBSS 2) state that $\mathbf{z}(\mathbf{s})$ if formed by uncorrelated, stationary random fields, the assumption of uncorrelatedness might be replaced by statistical independence. Figure 4.1 depicts an example of a SBSS model.

Let $C_i(\mathbf{h})$ denote the stationary covariance functions of $(\mathbf{z}(\mathbf{s}))_i$ for $i = 1, \ldots, p$, then the covariance matrix of $\mathbf{x}(\mathbf{s})$ from the SBSS model writes as

$$\operatorname{Cov}(\mathbf{x}(\mathbf{0}), \mathbf{x}(\mathbf{h})) = \sum_{i=1}^{p} \mathbf{a}_i \mathbf{a}_i^\top C_i(\mathbf{h}).$$

Here, $\mathbf{a}_i$ is the i-th row vector of the mixing matrix $\mathbf{A}$. In that view the SBSS model is a special case of the LMC (Equation (2.3)) model which is symmetric and the coregionalization matrices are of rank one. The key advantage to the usual LMC methodology is that the SBSS methods find these coregionalization matrices (or equivalently the unmixing matrix) without any model assumption or estimation of the p univariate correlation functions of the latent field. In that view the complex fitting of an LMC or a multivariate Matérn model can be avoided by using the powerful BSS methodology which leads to the following unmixing matrix functionals.

4.1.2. Unmixing matrix functionals

Similarly as in the SOS model, (SBSS 3) states that the spatial covariance matrix of the latent field is diagonal for any lag vector. Therefore, it might be a good strategy to diagonalize scatter matrices that measure spatial second order dependence evaluated on the whitened version of the observable in the spirit of AMUSE and SOBI. For that purpose Bachoc et al. (2020) introduce local covariance matrices which are defined as

$$\mathbf{LCov}_f(\mathbf{x}(\mathbf{s})) = n^{-1} \sum_{i,j=1}^{n} f(\mathbf{s}_i - \mathbf{s}_j) \mathbb{E}\left((\mathbf{x}(\mathbf{s}_i) - \mathbb{E}\left(\mathbf{x}(\mathbf{s}_i)\right)) (\mathbf{x}(\mathbf{s}_j) - \mathbb{E}\left(\mathbf{x}(\mathbf{s}_j)\right))^\top \right). \tag{4.1}$$

The sample version of the former quantity is given by

$$\hat{\mathbf{LCov}}_f(\mathbf{x}(\mathbf{s})) = n^{-1} \sum_{i,j=1}^{n} f(\mathbf{s}_i - \mathbf{s}_j)(\mathbf{x}(\mathbf{s}_i) - \bar{\mathbf{x}})(\mathbf{x}(\mathbf{s}_j) - \bar{\mathbf{x}})^\top,$$

where $\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^{n} \mathbf{x}(\mathbf{s}_i)$. Given n sample locations these matrices compute a weighted average of covariance matrices between all n^2 possible pairs of sample locations. The weights are determined by functions $f : \mathbb{R}^d \to \mathbb{R}$ which are denoted as spatial kernel functions. The following three options are considered by Bachoc et al. (2020).

- Ball kernel: $f_b(\mathbf{s}; r) = I(\|\mathbf{s}\| \leq r)$ where $r \geq 0$.
- Ring kernel: $f_r(\mathbf{s}; r_i, r_o) = I(r_i < \|\mathbf{s}\| \leq r_o)$ where $r_o > r_i \geq 0$.
- Gauss kernel: $f_g(\mathbf{s}; r) = \exp(-0.5(\Phi^{-1}(0.95)\|\mathbf{s}\|/r)^2)$ where $\Phi^{-1}(0.95)$ is the 95% quantile of the standard Normal distribution and $r > 0$.

$I(\cdot)$ denotes the indicator function. All kernel functions above are designed for isotropic random fields as they exclusively rely on the norm of the distance as argument. It is possible to design kernel function that specifically account for the direction of the separation vector. As an example a $\mathbf{LCov}_f$ matrix in conjunction with a ring kernel function measures the average second order spatial dependence of observations separated at least by a distance of r_i and at most by a distance of r_o. This might be viewed as an extension of autocovariance matrices where the ring kernel forms a higher dimensional extension of lags in time series. Note that local covariance matrices are scatter matrices. For the whitening step it is convenient to write the covariance matrix as a local covariance matrix by using a ball kernel function with a parameter of $r = 0$ namely $f = f_0 = f_b(\cdot; 0)$ which yields

$$\mathbf{LCov}_{f_0}(\mathbf{x}(\mathbf{s})) = n^{-1} \sum_{i=1}^{n} \mathbb{E}\left((\mathbf{x}(\mathbf{s}_i) - \mathbb{E}\left(\mathbf{x}(\mathbf{s}_i)\right)) (\mathbf{x}(\mathbf{s}_i) - \mathbb{E}\left(\mathbf{x}(\mathbf{s}_i)\right))^\top \right).$$

The sample version of $\mathbf{LCov}_{f_0}$ follows from the above sample version of $\mathbf{LCov}_f$.

As $\mathbf{LCov}_{f_0}(\mathbf{x}(\mathbf{s}))$ is a proper choice for whitening and the $\mathbf{LCov}_f(\mathbf{x}(\mathbf{s}))$ matrices are diagonal for the latent random field in a SBSS model SBSS unmixing matrix functionals can be defined based on simultaneous/joint diagonalization diagonalization as typically used in BSS (Section 3.3).

Definition 30 (SBSS simultaneous diagonalization)**.** *Let* $\mathbf{x}(\mathbf{s})$ *be a* p*-variate random field following the SBSS model (Definition 29) and let* f *be a spatial kernel function. The SBSS unmixing matrix functional* $\mathbf{W} = \mathbf{W}(\mathbf{x}(\mathbf{s}))$ *based on simultaneous diagonalization satisfies*

$$\mathbf{W}\,\mathbf{LCov}_{f_0}(\mathbf{x}(\mathbf{s}))\mathbf{W}^\top = \mathbf{I}_p \quad \textit{and} \quad \mathbf{W}\,\mathbf{LCov}_f(\mathbf{x}(\mathbf{s}))\mathbf{W}^\top = \mathbf{D}_f,$$

where $\mathbf{D}_f$ *is a diagonal matrix with decreasingly ordered diagonal elements.*

The above unmixing matrix functional can be seen as a direct extension of AMUSE (Definition 22) as the covariance and one local covariance matrices are simultaneously diagonalized. The above method is affine equivariant and as stated by Bachoc et al. (2020, Proposition 1) if the diagonal elements of $\mathbf{LCov}_f(\mathbf{z}(\mathbf{s}))$ are pairwise distinct then the solution is unique up to sign (the order is fixed by the ordering of the diagonal elements of $\mathbf{D}_f$ and the scale is fixed by the SBSS model). This identifiability condition is a joint property of the actual covariance functions for each entry of the latent field as well as the used spatial kernel function which leads to the fact that the choice of the kernel function is critical for the performance of this method. In practical applications this might lead to problems. Similarly, in the time series case (SOS model) it was found to be beneficial to jointly diagonalize several autocovariance matrices (SOBI) over diagonalizing only one (AMUSE) in order to reduce the influence of the chosen lag for the latter approach. Hence, joint diagonalization is applied in the SBSS context as a direct extension of the SOBI method (Definition 23).

Definition 31 (SBSS joint diagonalization)**.** *Let* $\mathbf{x}(\mathbf{s})$ *be a* p*-variate random field following the SBSS model (Definition 29) and* $\{f_1, \ldots, f_L\}$ *be a set of spatial kernel functions. Define the whitened version of* $\mathbf{x}(\mathbf{s})$ *by* $\mathbf{x}^{wh}(\mathbf{s}) = \mathbf{LCov}_{f_0}^{-1/2}(\mathbf{x}(\mathbf{s}))(\mathbf{x}(\mathbf{s}) - \mathbf{b})$*. Let* $\mathbf{U} = \mathbf{U}(\mathbf{x}(\mathbf{s}))$ *be the orthogonal* $p \times p$ *joint diagonalization matrix of* $\mathbf{LCov}_{f_l}(\mathbf{x}^{wh}(\mathbf{s}))$ *for* $l = 1, \ldots, L$*, maximizing*

$$\sum_{l=1}^{L} \|\mathrm{diag}(\mathbf{U}^\top \mathbf{LCov}_{f_l}(\mathbf{x}^{wh}(\mathbf{s}))\mathbf{U})\|_F^2.$$

The SBSS unmixing matrix functional $\mathbf{W} = \mathbf{W}(\mathbf{x}(\mathbf{s}))$ *based on joint diagonalization equals* $\mathbf{W} = \mathbf{U}^\top \mathbf{LCov}_{f_0}^{-1/2}(\mathbf{x}(\mathbf{s}))$.

The former method is affine equivariant and as formulated by Bachoc et al. (2020, Proposition 5) a unique solution (up to sign as the order is fixed by the arrangement of the pseudo-eigenvalues and the scale is fixed by the SBSS model) can be found if for every pair $i, j = 1, \ldots, p$ with $i \neq j$ there exists a number $l = 1, \ldots, L$ for which it holds that $(\mathbf{LCov}_{f_l}(\mathbf{z}(\mathbf{s})))_{ii} \neq (\mathbf{LCov}_{f_l}(\mathbf{z}(\mathbf{s})))_{jj}$. This identifiability condition is far less restrictive than the one stated for the method outlined in Definition 30 which is the great advantage of joint over simultaneous diagonalization methods. Note that for the case of $L = 1$ Definition 31 reduces to Definition 30. The optimization equation for the orthogonal matrix can be equivalently rewritten as

$$\sum_{i=1}^{p} \lambda_i^2,\ \lambda_i^2 = \sum_{l=1}^{L} \left(\mathbf{u}_i^\top \mathbf{LCov}_{f_l}(\mathbf{x}^{wh}(\mathbf{s}))\mathbf{u}_i\right)^2,$$

where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$ is the orthogonal joint diagonalization matrix and λ_i^2 are the pseudo-eigenvalues which hint the spatial dependence of the found latent field entries. For a given sample the population versions of the above quantities are replaced by their sample counterparts and approximate joint digaonalization algorithms have to be utilized as outlined in Section 3.4. Figure 4.2 depicts the SBSS solution using two ring kernel functions for the data seen in Figure 4.1. In summary, methods relying on joint diagonalization are generally less sensitive to the parameter choice but come at the cost of using complex approximate joint diagonalization algorithms.

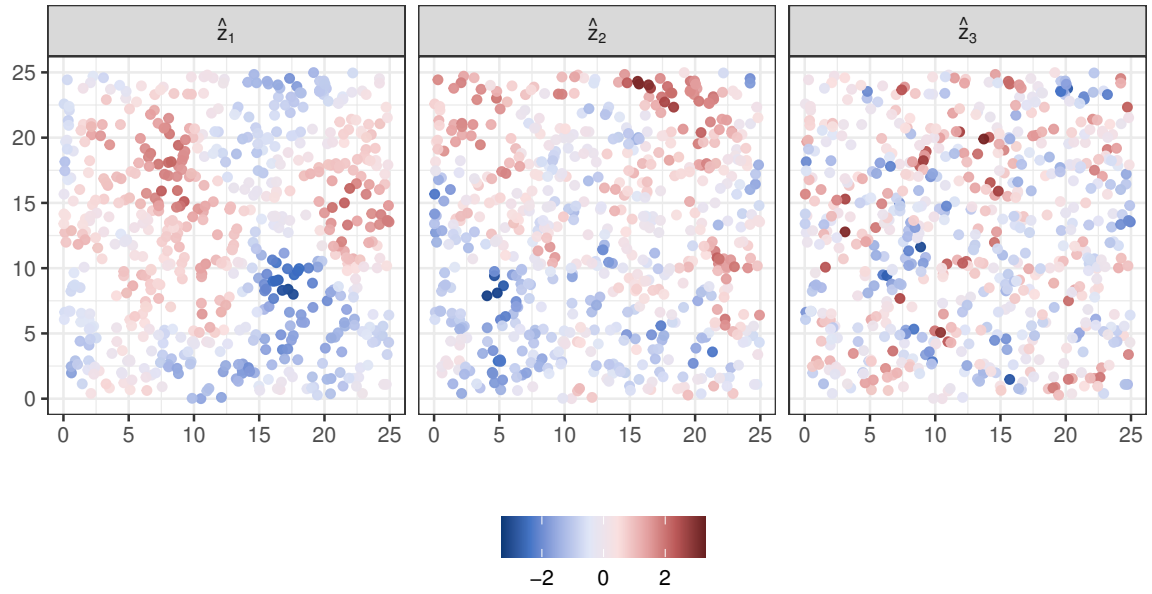

Figure 4.2.: Latent random fields which are found by applying SBSS with two ring kernel functions where the parameters equal $(0, 1)$ and $(1, 2)$ applied on the data from Figure 4.1.

4.1.3. Determining the signal subspace

In the same fashion as in the INM for stationary time series in the BSS model (Definition 12) not all components of the latent field might be of interest. For that purpose Publication I (Muehlmann et al., 2020a) introduces the INM for the SBSS model as follows.

Definition 32 (Spatial internal noise model (SINM))**.** *Let* $\mathbf{x}(\mathbf{s})$ *be a* p*-variate observable random field defined on a spatial domain* $\mathcal{S} \subseteq \mathbb{R}^d$*.* $\mathbf{x}(\mathbf{s})$ *follows a spatial internal noise model if it can be written as*

$$\mathbf{x}(\mathbf{s}) = \mathbf{A} \begin{pmatrix} \mathbf{z}_{sg}(\mathbf{s}) \\ \mathbf{z}_{ns}(\mathbf{s}) \end{pmatrix} + \mathbf{b},$$

where $\mathbf{A}$ *is the deterministic* $p \times p$ *mixing matrix of full-rank and* $\mathbf{b}$ *is a* p*-variate deterministic location vector. The latent random field* $\mathbf{z}(\mathbf{s}) = \left(\mathbf{z}_{sg}(\mathbf{s})^\top, \mathbf{z}_{ns}(\mathbf{s})^\top\right)^\top$ *can be divided into a* q*-variate signal random field* $\mathbf{z}_{sg}(\mathbf{s})$ *and* $(p-q)$*-variate noise random field* $\mathbf{z}_{ns}(\mathbf{s})$ *satisfying the following assumptions.*

(SINM 1) $\mathbb{E}\left(\mathbf{z}_{sg}(\mathbf{s})\right) = \mathbf{0}$ *for* $\forall\ \mathbf{s} \in \mathcal{S}$*,*

(SINM 2) $\mathrm{Cov}(\mathbf{z}_{sg}(\mathbf{s})) = \mathbb{E}\left(\mathbf{z}_{sg}(\mathbf{s})\mathbf{z}_{sg}(\mathbf{s})^\top\right) = \mathbf{I}_p$ *for* $\forall\ \mathbf{s} \in \mathcal{S}$ *and*

(SINM 3) $\mathrm{Cov}(\mathbf{z}_{sg}(\mathbf{s}), \mathbf{z}_{sg}(\mathbf{s}')) = \mathbb{E}\left(\mathbf{z}_{sg}(\mathbf{s})\mathbf{z}_{sg}(\mathbf{s}')^\top\right) = \mathbf{D}_\mathbf{h}$*, where* $\mathbf{h} = \mathbf{s} - \mathbf{s}'$*, for* $\forall\ \mathbf{s}, \mathbf{s}' \in \mathcal{S}$*,* $\mathbf{s} \neq \mathbf{s}'$ *and* $\mathbf{D}_\mathbf{h}$ *is a diagonal matrix containing the (univariate) covariance functions for each entry of* $\mathbf{z}(\mathbf{s})$ *on its diagonal elements which depend only on* $\mathbf{h}$*.*

(SINM 4) *The noise latent field* $\mathbf{z}_{ns}(\mathbf{s})$ *is white noise with identity covariance matrix and it is statistically independent of* $\mathbf{z}_{sg}(\mathbf{s})$*.*

Typical for an INM is that only a few components are of interest and the remaining ones are discarded as noise. Specifically, (SINM 1) - (SINM 3) state that the q signal parts are uncorrelated stationary random fields exhibiting second order spatial dependence, namely the signal part follows a SBSS model. In contrast (SINM 4) states that the remaining $p-q$ components of the latent field are statistically independent of the signal part and are white noise (Definition 4).

The aim of Publication I is to infer the noise dimension q based on asymptotic results and bootstrap strategies. For that purpose the publication introduces an adapted version of local covariance matrices (Equation (4.1)) to obtain elegant asymptotic results. Namely, given n sample locations the normalization factor n^{-1} from Bachoc et al. (2020) is replaced by $n^{-1}F_{f,n}^{-1/2}$ where

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

which might be interpreted as the average (squared) number of neighboring sample locations for a given kernel function f. This adaptation leads to

$$\mathbf{LCov}'_f(\mathbf{x}(\mathbf{s})) = \frac{1}{n\sqrt{F_{f,n}}} \sum_{i,j=1}^{n} f(\mathbf{s}_i - \mathbf{s}_j)\mathbb{E}\left((\mathbf{x}(\mathbf{s}_i) - \mathbb{E}\left(\mathbf{x}(\mathbf{s}_i)\right))(\mathbf{x}(\mathbf{s}_j) - \mathbb{E}\left(\mathbf{x}(\mathbf{s}_j)\right))^\top\right),$$

and its sample version is given by

$$\hat{\mathbf{LCov}}'_f(\mathbf{x}(\mathbf{s})) = \frac{1}{n\sqrt{F_{f,n}}} \sum_{i,j=1}^{n} f(\mathbf{s}_i - \mathbf{s}_j)(\mathbf{x}(\mathbf{s}_i) - \bar{\mathbf{x}})(\mathbf{x}(\mathbf{s}_j) - \bar{\mathbf{x}})^\top.$$

Consequently, the former two SBSS methods (Definition 30 and 31) are adapted by replacing the $\mathbf{LCov}_f(\mathbf{x}(\mathbf{s}))$ with the corresponding $\mathbf{LCov}'_f(\mathbf{x}(\mathbf{s}))$ matrices. The general outline of the methods and the corresponding optimization strategies keep unchanged and the procedure for testing the signal dimension q is equal to the one outlined for the INM in Section 3.5.1. The considered test statistic is given by

$$t_{q,n} = \frac{n}{2} \sum_{l=1}^{L} \|\hat{\mathbf{D}}_{f_l}^{00}\|_F^2, \tag{4.2}$$

where the approximately jointly diagonalized local covariance matrices

$$\hat{\mathbf{D}}_{f_l} = \hat{\mathbf{W}}\hat{\mathbf{LCov}}'_{f_l}(\mathbf{x}(\mathbf{s}))\hat{\mathbf{W}}^\top \text{ for } l = 1, \ldots, L,$$

are obtained by one of the SBSS methods before. Again they are partitioned into the noise and signal sub-matrices by

$$\hat{\mathbf{D}}_{f_l} = \begin{pmatrix} \hat{\mathbf{D}}_{f_l}^{qq} & \hat{\mathbf{D}}_{f_l}^{q0} \\ \hat{\mathbf{D}}_{f_l}^{0q} & \hat{\mathbf{D}}_{f_l}^{00} \end{pmatrix} \text{ for } l = 1, \ldots, L.$$

$t_{q,n}$ is expected to diverge with n when there is a signal component in the hypothetical noise part and bounded when the noise part consists only of white noise. The distribution of the null hypothesis can be either approximated by asymptotic results or bootstrap strategies described below.

Asymptotic results As outlined in Publication I under sufficient conditions the limiting distribution of the test statistic seen in Equation (4.2) can be derived yielding the following proposition.

Proposition 2. *Under sufficient conditions, detailed in Publication I, it holds that*

$$t_{q,n} \xrightarrow{\mathcal{L}} \chi^2_{L(p-q)(p-q+1)/2} \text{ as } n \to \infty,$$

where χ^2_ν is the chi-squared distribution with ν degrees of freedom.

Roughly, these conditions are that the latent field consists of Gaussian components, that the domain increases as the number of sample locations increases (increasing domain asymptotics) and more technical conditions which can be found in the publication. The former result is based on the outline of Virta and Nordhausen (2021).

The nice closed form asymptotic distribution of the test statistic seen in Proposition 2 can be easily used to carry out an (approximate) test for the signal dimension q. Additionally, the publication also provides results that this test strategy can consistently estimate q but in practical considerations estimation of the signal dimension are rather carried out by applying sequential testing strategies such as the ones discussed in Section 3.5.

Bootstrap strategies Publication I considers the iid bootstrap similar as Matilainen et al. (2018) (see Section 3.5) to infer a signal dimension. Additionally, an optional spatial bootstrap step is added to the testing procedure. Namely, the quantity of interest is a bootstrap estimate of

$$P(t_{q,n} \leq x),$$

where $t_{q,n} = t_{q,n}(\mathbf{x}(\mathbf{s}))$ is a function of $\mathbf{x}(\mathbf{s})$. The bootstrap is based on the four following steps.

Step 1 The latent random field $\hat{\mathbf{z}}(\mathbf{s})$ and the unmixing matrix $\hat{\mathbf{W}}$ are estimated by some SBSS method and the corresponding test statistic $t_{q,n}(\mathbf{x}(\mathbf{s}))$ is computed.

Step 2 The last $p-q$ components of $\hat{\mathbf{z}}(\mathbf{s})$ are white noise under the null hypothesis. They are either replaced by an iid sample of $N(0,1)$ (parametric strategy) or an iid bootsrap sample of the ecdf of the $n(p-q)$-variate vector formed by joining all noise vectors (non-parametric 1 strategy). This leads to $\hat{\mathbf{z}}^*(\mathbf{s})$.

Step 3 Step 2 only accounts for the hypothetical noise part but does not account for the spatial dependence of the data at hand. Therefore, as an optional step 3 a full spatial bootstrap sample from $\hat{\mathbf{z}}^*(\mathbf{s})$ is drawn.

Step 4 The bootstrap sample $\hat{\mathbf{z}}^*(\mathbf{s})$ generated in Step 2 (and optionally in Step 3) is mixed again by $\hat{\mathbf{x}}^*(\mathbf{s}) = \hat{\mathbf{W}}^{-1}\hat{\mathbf{z}}^*(\mathbf{s})$ which leads to a bootstrap sample $\hat{\mathbf{x}}^*(\mathbf{s})$ of the original random field $\mathbf{x}(\mathbf{s})$. With that bootstrap sample the test statistic $t^*_{q,n}(\hat{\mathbf{x}}^*(\mathbf{s}))$ can be computed.

In the same fashion as described in Section 3.5 the computation of the p-value is based on B repetitions of Step 2 - 4. Equally as in the asymptotic testing strategy the estimation of the signal dimension can be carried out by sequentially testing for different signal dimension, e.g.: with the divide-and-conquer strategy discussed above.

The spatial bootstrap variant used as an optional Step 3 is based on sampling from a suitable set of blocks rather then sampling from the individual observations. It is in detail discussed in Lahiri (2003, Chapter 12) and recalled as follows. Consider a univariate random field $x(\mathbf{s})$ that is observed at n sample locations $\mathcal{C} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathcal{S}$ which are located inside the spatial domain $\mathcal{S} \subseteq \mathbb{R}^d$. For a real number $\alpha > 0$ the set of non-overlapping blocks (partly) contained in $\mathcal{S}$ can be defined by

$$\mathcal{B}_{nol} = \{\boldsymbol{\mu} \in \mathbb{Z}^d : \alpha \left(\boldsymbol{\mu} + (0,1]^d\right) \cap \mathcal{S} \neq \varnothing\}$$

which leads to $\mathcal{S} = \cup_{\boldsymbol{\mu} \in \mathcal{B}_{nol}} \mathcal{S}_{\boldsymbol{\mu}}$ with $\mathcal{S}_{\boldsymbol{\mu}} = \alpha \left(\boldsymbol{\mu} + (0,1]^d\right) \cap \mathcal{S}$ for all $\boldsymbol{\mu} \in \mathcal{B}_{nol}$. Loosely formulated $\mathcal{S}_{\boldsymbol{\mu}}$ are blocks with a side length of α that are cut to the shape of the domain border if the block does not fully lie inside the domain. Similarly, the set of overlapping blocks that lie completely inside the domain is defined by

$$\mathcal{B}_{ol} = \{\boldsymbol{\nu} \in \mathbb{Z}^d : \boldsymbol{\nu} + (0,1]^d \subset \mathcal{S}\}.$$

For each non-overlapping block $\boldsymbol{\mu} \in \mathcal{B}_{nol}$ a block from the set of overlapping blocks can be selected and trimmed to the shape of $\mathcal{S}_{\boldsymbol{\mu}}$ by

$$\mathcal{S}_{\boldsymbol{\mu},\boldsymbol{\nu}} = (\mathcal{S}_{\boldsymbol{\mu}} - \alpha\boldsymbol{\mu} + \boldsymbol{\nu}) \cap \left(\boldsymbol{\nu} + (0,1]^d\right).$$

With this definition $\mathcal{S}_{\boldsymbol{\mu},\boldsymbol{\nu}}$ has the same shape as $\mathcal{S}_{\boldsymbol{\mu}}$ but lies at the location given by $\boldsymbol{\nu}$. The bootstrap sample is defined by replacing each $\mathcal{S}_{\boldsymbol{\mu}}$ by some block $\mathcal{S}_{\boldsymbol{\mu},\boldsymbol{\nu}}$ where the index $\boldsymbol{\nu} \in \mathcal{B}_{ol}$ is drawn randomly with replacement. Specifically, for each $\boldsymbol{\mu} \in \mathcal{B}_{nol}$ a random variable $Y_{\boldsymbol{\mu}}$ can be defined that puts equal weight on each possible index of overlapping blocks $\boldsymbol{\nu} \in \mathcal{B}_{ol}$ by

$$P(Y_{\boldsymbol{\mu}} = \boldsymbol{\nu}) = \frac{1}{|\mathcal{B}_{ol}|} \text{ for all } \boldsymbol{\nu} \in \mathcal{B}_{ol}.$$

Finally, the bootstrap version of the random field $Z(\mathbf{s})$ is formed by concatenated all bootstraped blocks by

$$Z^*(\mathbf{s}) = \cup_{\boldsymbol{\mu} \in \mathcal{B}_{nol}} \{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{S}_{\boldsymbol{\mu},Y_{\boldsymbol{\mu}}} \cap \mathcal{C}\}.$$

Similar as in the iid case the distribution of quantities that are based on $Z^*(\mathbf{s})$ is estimated by drawing B samples of the random variables $Y_{\boldsymbol{\mu}}$ for all $\boldsymbol{\mu} \in \mathcal{B}_{nol}$ leading to $Z_b^*(\mathbf{s})$ for $b = 1, \ldots, B$. Lahiri (2003, Chapter 12) provides also a detailed discussion about the allowed random processes that generate the set of sample locations $\mathcal{C}$, discussion about allowed shapes of the domain $\mathcal{S}$ as well as discussion about the large sample behavior of the above outline.

4.2. Non-stationary spatial source separation

It is well-studied that the second order stationary assumptions might be too restricting in practical considerations. On the one hand the random field at hand might shows a non-constant drift function leading to the considerations of Section 4.2.1 but on the other hand the variance or generally the second order (spatial) dependence varies across the domain and is not invariant under shifts, this case is discussed in Section 4.2.2.

4.2.1. Non-stationary drift function

Publication II and III (Muehlmann et al., 2020b, 2021c) evolve the idea of using differences over the usual local covariances for random fields with a non-constant drift function. The use of differences leads to many advantages in the context of ICA (Section 3.4.1) and is common practice in geostatistics where often the variogram is favored over the covariance matrix (Section 2.1). Local difference matrices are defined by

$$\mathbf{LDiff}_f(\mathbf{x}(\mathbf{s})) = \frac{1}{n}\sum_{i,j=1}^{n} f(\mathbf{s}_i - \mathbf{s}_j)\mathbb{E}\left((\mathbf{x}(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_j))(\mathbf{x}(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_j))^\top\right),$$

where f is a spatial kernel functions as discussed above. For a multivariate random field exhibiting the location scatter-model with a non-constant drift function $\mathbf{x}(\mathbf{s}) = \mathbf{A}\mathbf{z}(\mathbf{s}) + \mathbf{m}(\mathbf{s})$ local difference matrices are of the form

$$\mathbf{LDiff}_f(\mathbf{x}(\mathbf{s})) = \mathbf{LDiff}_f(\mathbf{A}\mathbf{z}(\mathbf{s})) + \frac{1}{n}\sum_{i,j=1}^{n} f(\mathbf{s}_i - \mathbf{s}_j)(\mathbf{m}(\mathbf{s}_i) - \mathbf{m}(\mathbf{s}_j))(\mathbf{m}(\mathbf{s}_i) - \mathbf{m}(\mathbf{s}_j))^\top.$$

The last term is supposed to be negligible if the drift is a function with a low-valued absolute value of the gradient vector. In contrast local covariance matrices are highly corrupted as the mean needs to be estimated which is usually done by the sample mean. Due to this reasons Publication II and III replace local covariance matrices by local difference matrices in the SBSS methods (Definition 30 and 31). For this adaptation it is advisable to use ball kernels with a small parameter in order to keep the influence of the drift small.

Still the adapted methods utilize whitening with respect to the standard covariance matrix. Therefore, Publication III replaces also the covariance in the whitening step by a local difference matrix which is an example of the robust whitening procedure described in Section 3.3.1 and 3.5.2. However, local difference matrices always capture the on-site covariance which leads to the fact that they are always invertible under the SBSS model. This adaptation yields the following definition.

Definition 33. *Consider two spatial kernel functions f_1 and f_2. For a random field $\mathbf{x}(\mathbf{s})$ following the SBSS model in Definition 29. The unmixing matrix functional $\mathbf{W} = \mathbf{W}(\mathbf{x}(\mathbf{s}))$ simultaneously diagonalizes the corresponding two local difference matrices such that*

$$\mathbf{W}\,\mathbf{LDiff}_{f_1}(\mathbf{x}(\mathbf{s}))\mathbf{W}^\top = \mathbf{I}_p \text{ and } \mathbf{W}\,\mathbf{LDiff}_{f_2}(\mathbf{x}(\mathbf{s}))\mathbf{W}^\top = \mathbf{D}_{f_1 f_2}.$$

Where $\mathbf{D}_{f_1 f_2}$ is a diagonal matrix with increasingly ordered diagonal elements.

Note that all the adapted methods still rely on the strict stationarity assumptions of the SBSS model (Definition 29). Only for the method seen in Definition 33 the assumptions (SBSS 1) and (SBSS 2) can be relaxed to the following ones.

(SBSS 1') $\mathbb{E}(\mathbf{z}(\mathbf{s})) = \mathbf{0}$, $\mathbf{LDiff}_{f_1}(\mathbf{z}(\mathbf{s})) = \mathbf{I}_p$ for all $\mathbf{s} \in \mathcal{S}$ and

(SBSS 2') $\mathbb{E}\left((\mathbf{z}(\mathbf{s}) - \mathbf{z}(\mathbf{s}'))(\mathbf{z}(\mathbf{s}) - \mathbf{z}(\mathbf{s}'))^\top\right) = \mathbf{D}_\mathbf{h}$ where $\mathbf{h} = \mathbf{s} - \mathbf{s}'$, $\forall\ \mathbf{s}, \mathbf{s}' \in \mathcal{S},\ \mathbf{s} \neq \mathbf{s}'$ and $\mathbf{D}_\mathbf{h}$ is a diagonal matrix containing the (univariate) variogram functions for each entry of $\mathbf{z}(\mathbf{s})$ on its diagonal elements which depend only on $\mathbf{h}$.

Publication III shows the usefulness of this methods in an extensive simulation study where stationary random fields with non-constant drift functions and intrinsic stationary random fields are considered as well as a geochemical data example.

4.2.2. Non-stationary second order spatial dependence

Publication IV (Muehlmann et al., 2021a) adapts the original SBSS methodology by keeping the constant drift assumptions but allowing non-stationary second order (spatial) dependence. This publications considers the following model which can be seen as a direct extension of the NSS model (Definition 9).

Definition 34 (Spatial non-stationary source separation model (SNSS))**.** *A p-variate random field $\mathbf{x}(\mathbf{s})$ defined on a d-dimensional spatial domain $\mathcal{S} \subseteq \mathbb{R}^d$ follows a spatial non-stationary source separation model (SNSS) if it can be formulated as*

$$\mathbf{x}(\mathbf{s}) = \mathbf{A}\mathbf{z}(\mathbf{s}) + \mathbf{b},$$

where $\mathbf{A}$ is a deterministic invertible $p \times p$ mixing matrix, $\mathbf{b}$ is a p-variate deterministic location vector and $\mathbf{z}(\mathbf{s})$ is a p-variate latent random field which fulfills the following assumptions

(SNSS 1) $\mathbb{E}(\mathbf{z}(\mathbf{s})) = \mathbf{0}$ *for all* $\mathbf{s} \in \mathcal{S}$,

(SNSS 2) $\operatorname{Cov}(\mathbf{z}(\mathbf{s})) = \mathbb{E}\left(\mathbf{z}(\mathbf{s})\mathbf{z}(\mathbf{s})^\top\right) = \mathbf{D}_{\mathbf{s}}$ *for all* $\mathbf{s} \in \mathcal{S}$ *where* $\mathbf{D}_{\mathbf{s}}$ *is a positive definite diagonal matrix and*

(SNSS 3) $\operatorname{Cov}(\mathbf{z}(\mathbf{s}), \mathbf{z}(\mathbf{s}')) = \mathbb{E}\left(\mathbf{z}(\mathbf{s})\mathbf{z}(\mathbf{s}')^\top\right) = \mathbf{D}_{\mathbf{s}\mathbf{s}'}$, *for all* $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$ *with* $\mathbf{s} \neq \mathbf{s}'$ *where* $\mathbf{D}_{\mathbf{s}\mathbf{s}'}$ *is a diagonal matrix depending on* $\mathbf{s}$ *and* $\mathbf{s}'$.

Similarly as for the NSS methods (Definition 24, 25 and 26) local covariance matrices are adapted to only consider sub-domains which do not necessarily need to cover all sample locations yielding

$$\mathbf{LCov}_{\mathcal{S},f}(\mathbf{x}(\mathbf{s})) = \frac{1}{|\mathcal{S} \cap \mathcal{C}|} \sum_{\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S} \cap \mathcal{C}} f(\mathbf{s}_i - \mathbf{s}_j) \mathbb{E}\left((\mathbf{x}(\mathbf{s}_i) - \mathbb{E}(\mathbf{x}(\mathbf{s}_i)))(\mathbf{x}(\mathbf{s}_j) - \mathbb{E}(\mathbf{x}(\mathbf{s}_j)))^\top\right).$$

As the covariance function is not defined for the SNSS model the whitening step can be carried out with respect to an adapted local covariance matrix with the kernel choice $f = f_0 = f_b(\cdot; 0)$ which yields

$$\mathbf{LCov}_{\mathcal{S},f_0}(\mathbf{x}(\mathbf{s})) = \frac{1}{|\mathcal{S} \cap \mathcal{C}|} \sum_{\mathbf{s} \in \mathcal{S} \cap \mathcal{C}} \mathbb{E}\left((\mathbf{x}(\mathbf{s}) - \mathbb{E}(\mathbf{x}(\mathbf{s})))(\mathbf{x}(\mathbf{s}) - \mathbb{E}(\mathbf{x}(\mathbf{s})))^\top\right).$$

In this outline three SNSS methods are introduced which are a direct extension of the three NSS methods seen in Definition 24, 25 and 26. The first one is based on simultaneous diagonalization of covariance matrices for two sub-domains.

Definition 35 (Spatial non-stationary source separation - simultaneous diagonalization (SNSS-SD))**.** *Consider a random field $\mathbf{x}(\mathbf{s})$ following the SNSS model (Definition 34) and a partition of the spatial domain $\mathcal{S}$ into $\mathcal{S}_1, \mathcal{S}_2$ where $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$. The SNSS-SD functional $\mathbf{W} = \mathbf{W}(\mathbf{x}(\mathbf{s}))$ is defined as the simultaneous diagonalizer satisfying*

$$\mathbf{W}\,\mathbf{LCov}_{\mathcal{S}_1,f_0}(\mathbf{x}(\mathbf{s}))\mathbf{W}^\top = \mathbf{I}_p \quad \text{and} \quad \mathbf{W}\,\mathbf{LCov}_{\mathcal{S}_2,f_0}(\mathbf{x}(\mathbf{s}))\mathbf{W}^\top = \mathbf{D}_{\mathcal{S}_1\mathcal{S}_2},$$

where $\mathbf{D}_{\mathcal{S}_1\mathcal{S}_2}$ is a diagonal matrix with decreasingly ordered diagonal elements.

The second method extends the former one by jointly diagonalizing covariance matrices for more than two sub-domains.

Definition 36 (Spatial non-stationary source separation - joint diagonalization (SNSS-JD))**.** *Consider a random field* $\mathbf{x}(\mathbf{s})$ *following the SNSS model (seen in Definition 34). Whiten* $\mathbf{x}(\mathbf{s})$ *by* $\mathbf{x}^{wh}(\mathbf{s}) = \mathbf{M}_{\mathcal{S},f_0}^{-1/2}(\mathbf{x}(\mathbf{s}))(\mathbf{x}(\mathbf{s}) - \mathbf{b})$ *and partition the spatial domain* $\mathcal{S}$ *into* $\mathcal{S}_1, \ldots, \mathcal{S}_K$ *where* $\mathcal{S}_m \cap \mathcal{S}_n = \emptyset$ *for* $m, n = 1, \ldots, K$ *and* $m \neq n$*. Let* $\mathbf{U} = \mathbf{U}(\mathbf{x}(\mathbf{s}))$ *be the orthogonal* $p \times p$ *joint diagonalizer of the matrices* $\mathbf{LCov}_{\mathcal{S}_k,f_0}(\mathbf{x}^{wh}(\mathbf{s}))$ *for* $k = 1, \ldots, K$*, which maximizes*

$$\sum_{k=1}^{K} \|\mathrm{diag}(\mathbf{U}\,\mathbf{LCov}_{\mathcal{S}_k,f_0}(\mathbf{x}^{wh}(\mathbf{s}))\mathbf{U}^\top)\|_F^2.$$

Then, the SNSS-JD functional equals $\mathbf{W}(\mathbf{x}(\mathbf{s})) = \mathbf{U}\,\mathbf{LCov}_{\mathcal{S},f_0}^{-1/2}(\mathbf{x}(\mathbf{s}))$.

The last method is the most general one as for the sub-domains not only covariance matrices but also local covariance matrices are used.

Definition 37 (Spatial non-stationary source separation - spatial joint diagonalization (SNSS-SJD))**.** *Consider a random field* $\mathbf{x}(\mathbf{s})$ *following the SNSS model (Definition 34). Whiten* $\mathbf{x}(\mathbf{s})$ *by* $\mathbf{x}^{wh}(\mathbf{s}) = \mathbf{LCov}_{\mathcal{S},f_0}^{-1/2}(\mathbf{x}(\mathbf{s}))(\mathbf{x}(\mathbf{s}) - \mathbf{b})$ *and partition the spatial domain* $\mathcal{S}$ *into* $\mathcal{S}_1, \ldots, \mathcal{S}_K$ *where* $\mathcal{S}_m \cap \mathcal{S}_n = \emptyset$ *for* $m, n = 1, \ldots, K$ *and* $m \neq n$*. For a set of spatial kernel functions* $\{f_1, \ldots, f_L\}$*,* $\mathbf{U} = \mathbf{U}(\mathbf{x}(\mathbf{s}))$ *is an orthogonal* $p \times p$ *joint diagonalization matrix of the matrices* $\mathbf{LCov}_{\mathcal{S}_k,f_l}(\mathbf{x}^{st}(\mathbf{s}))$ *for all* $k = 1, \ldots, K$ *and* $l = 1, \ldots, L$*, which maximizes*

$$\sum_{k=1}^{K}\sum_{l=1}^{L} \|\mathrm{diag}(\mathbf{U}\,\mathbf{LCov}_{\mathcal{S}_k,f_l}(\mathbf{x}^{wh}(\mathbf{s}))\mathbf{U}^\top)\|_F^2.$$

Then, the SNSS-SJD functional is given by $\mathbf{W}(\mathbf{x}(\mathbf{s})) = \mathbf{U}\,\mathbf{LCov}_{\mathcal{S},f_0}^{-1/2}(\mathbf{x}(\mathbf{s}))$.

Publication IV proofs the affine equivariance property for all the above methods and gives sufficient conditions for the identifiability of the model parameters. Additionally, the use of these methods are proven experimentally in an extensive simulation study and on a geochemical dataset.

4.3. Spatial source separation in spatial prediction

BSS in the context of spatial data analysis is also useful in spatial prediction. Section 4.3.1 discusses the use of SBSS in the context of spatial prediction and Section 4.3.2 extends TSIR to stationary random fields which has great advantages in spatial regression tasks.

4.3.1. Spatial prediction

For all above discussed SBSS methods it holds that the estimated latent random field consists of uncorrelated components. This comes with great advantage in spatial prediction (Section 2.3) as multivariate predictions of the original field can be based on p univariate ones in the following three step procedure.

Step 1 The latent random field $\hat{\mathbf{z}}(\mathbf{s})$ and the unmixing matrix $\hat{\mathbf{W}}$ are estimated by some SBSS variant outlined above.

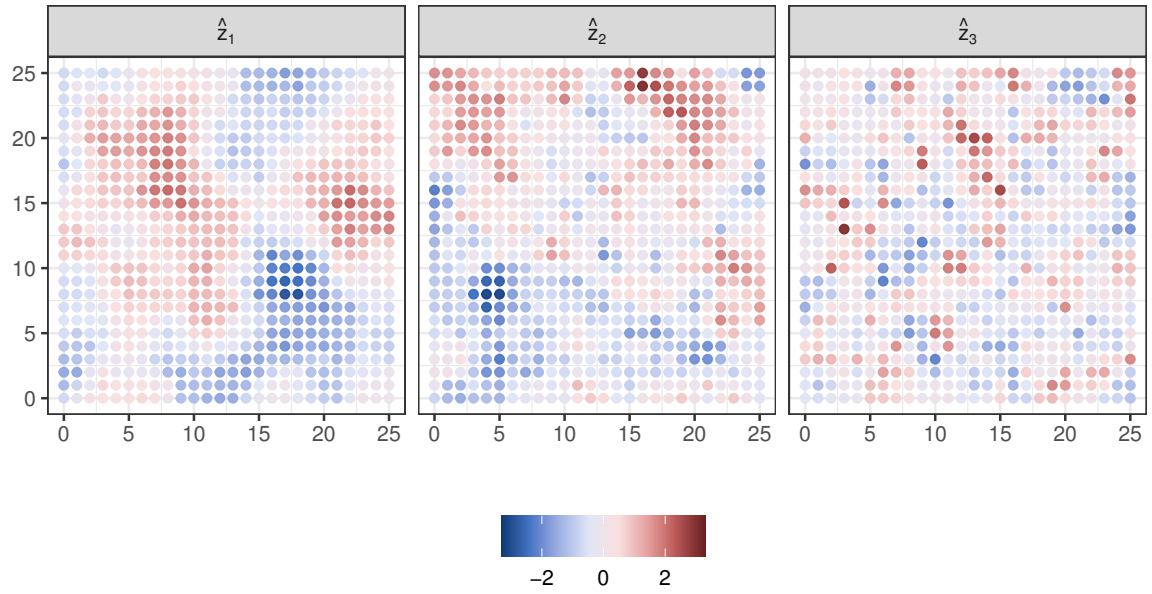

Figure 4.3.: IDW predictions (see Section 2.3) using the parameter $r = 5$ of the latent fields seen in Figure 4.2 on the integer grid ($\mathcal{S} \cap \mathbb{Z}^2$).

Step 2 A univariate spatial prediction model can be used for each entry of the estimated latent field $\hat{z}_i(\mathbf{s})$ for $i = 1, \ldots, p$ individually to predict the latent field components at some unobserved location $\mathbf{s}'$. The individual predictions are again collected in a vector by $\hat{\mathbf{z}}(\mathbf{s}') = (\hat{z}_1(\mathbf{s}'), \ldots, \hat{z}_p(\mathbf{s}'))^\top$.

Step 3 The prediction of the original field is obtained by mixing the predictions from Step 2. Namely, $\hat{\mathbf{x}}(\mathbf{s}') = \hat{\mathbf{W}}^{-1}\hat{\mathbf{z}}(\mathbf{s}') + \hat{\mathbf{T}}(\mathbf{x}(\mathbf{s}))$ where $\hat{\mathbf{T}}(\mathbf{x}(\mathbf{s}))$ is some location functional.

For Step 2 any spatial prediction method can be utilized. For example if the method of Definition 33 is used in Step 1 and it is assumed that the observable has a non-constant drift then a universal Kriging estimator would be suitable as still some transformed drift is present on the latent field (see also Publication III). For the case of the SINM model (Definition 32) in conjunction with the methods of Definition 30 or 31 Step 2 can be simplified even further by estimating the signal subspace as outlined in Section 4.1.3. If $q < p$ then only q predictions need to be carried out as the white noise is simply predicted with zero.

The above procedure is illustrated by considering the following example. Figure 4.2 depicts the SBSS solution of the data seen in Figure 4.1 using two ring kernel functions which corresponds to Step 1. Figure 4.3 depicts IDW predictions on a grid of the latent field illustrating Step 2 of the above procedure. Finally, Step 3 forms multivariate predictions of the original data on a grid by mixing the predictions again as seen in Figure 4.4.

Publication V (Muehlmann et al., 2020c) studies the usefulness of this procedure on a geochemical dataset and by carrying out an extensive simulation study. The contender methods are multivariate predictions using the gold standard Cokriging estimator in conjunction with the LMC (Equation (2.3)) and different forms of neural networks. Pre-processing the data with SBSS and then using univariate predictions with an ordinary Kriging estimator using

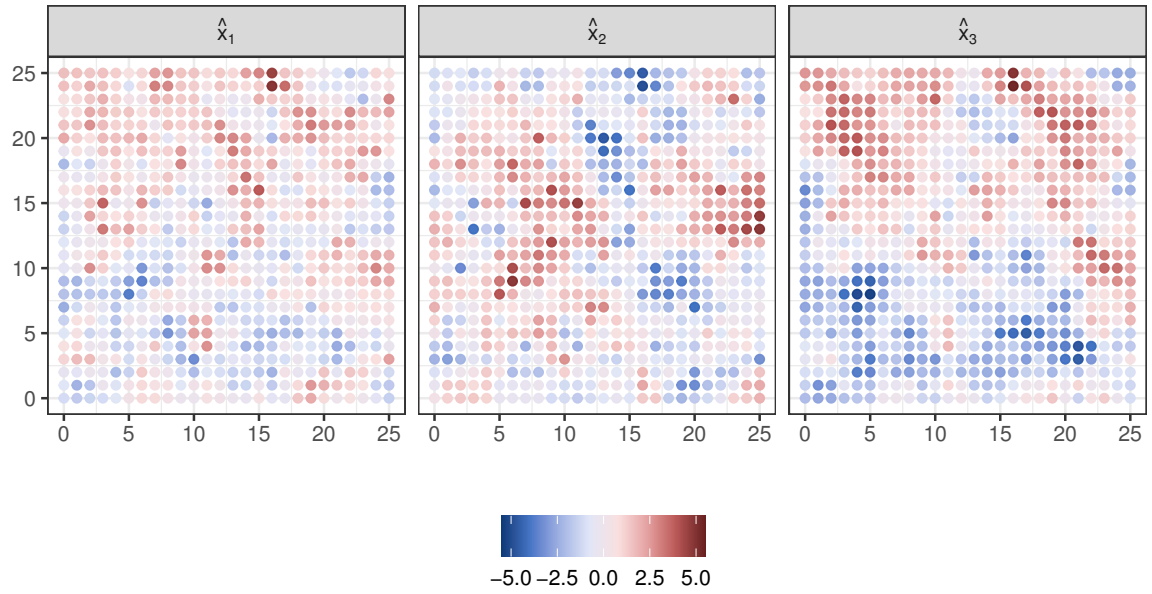

Figure 4.4.: Multivariate predictions of the random field seen in Figure 4.1 based on using SBSS in conjunction with IDW on the integer grid ($\mathcal{S} \cap \mathbb{Z}^2$) (see also Figure 4.2 and 4.3).

the univariate Matérn covariance function (Equation (2.2)) showed in almost all simulation settings better results compared to the contender methods.

4.3.2. Spatial regression

Publication VI (Muehlmann et al., 2021e) extends the time series version of SIR (Definition 28) to the spatial case by considering a univariate response random field $(y(\mathbf{s}))_{\mathbf{s}\in\mathbb{Z}^2}$ which is regressed on the predicting random field $(\mathbf{x}(\mathbf{s}))_{\mathbf{s}\in\mathbb{Z}^2}$ assumed to be second order stationary. The following model is considered.

Definition 38 (Spatial sufficient dimension reduction model in a BSS context (SSDR))**.** *Consider a $(p+1)$-variate random field $(y(\mathbf{s}), \mathbf{x}(\mathbf{s})^\top)^\top$ with $\mathbf{s} \in \mathbb{Z}^2$ where $y(\mathbf{s})$ is the (univariate) response and $\mathbf{x}(\mathbf{s})$ is its explaining random field which has the representation*

$$\mathbf{x}(\mathbf{s}) = \mathbf{A}\mathbf{z}(\mathbf{s}) + \mathbf{b} = \mathbf{A}\begin{pmatrix}\mathbf{z}_1(\mathbf{s})\\ \mathbf{z}_2(\mathbf{s})\end{pmatrix} + \mathbf{b},$$

where $\mathbf{A}$ is the invertible $p \times p$ mixing matrix, $\mathbf{b}$ is the deterministic p-dimensional location vector and $\mathbf{z}(\mathbf{s}) = (\mathbf{z}_1(\mathbf{s})^\top, \mathbf{z}_2(\mathbf{s})^\top)^\top$ is the p-variate stationary latent random field. $\mathbf{z}(\mathbf{s})$ is partitioned into a r-variate random field $\mathbf{z}_1(\mathbf{s})$ and a $(p-r)$-variate random field $\mathbf{z}_2(\mathbf{s})$ which satisfy:

(SSDR 1) $\mathbb{E}(\mathbf{z}(\mathbf{s})) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{z}(\mathbf{s})) = \mathbb{E}\left(\mathbf{z}(\mathbf{s})\mathbf{z}(\mathbf{s})^\top\right) = \mathbf{I}_p$ *for all* $\mathbf{s} \in \mathbb{Z}^2$ *and*

(SSDR 2) $\left(y(\mathbf{s}''), \mathbf{z}_1(\mathbf{s}')^\top\right)^\top \perp\!\!\!\perp \mathbf{z}_2(\mathbf{s})$ *for all* $\mathbf{s}, \mathbf{s}', \mathbf{s}'' \in \mathbb{Z}^2$.

The dimension r and the partitioning are minimal as in Definition 10.

For the extension of TSIR to the spatial case the variation of the lagged inverse regression curve can be defined by

$$\operatorname{Cov}(\mathbb{E}\left(\mathbf{x}(\mathbf{s})|y(\mathbf{s}+\mathbf{h})\right) \quad \text{for all} \quad \mathbf{h} \in \mathbb{Z}^2.$$

Note that this extension can be easily done as all considerations in this chapter are concerned with random fields which are defined on a gird-shaped domain ($\mathcal{S} = \mathbb{Z}^2$). Based on the former model it holds that

$$\operatorname{Cov}(\mathbb{E}\left(\mathbf{z}(\mathbf{s})|y(\mathbf{s}+\mathbf{h})\right) = \begin{pmatrix} \operatorname{Cov}(\mathbb{E}\left(\mathbf{z}_1(\mathbf{s})|y(\mathbf{s}+\mathbf{h})\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{for all} \quad \mathbf{h} \in \mathbb{Z}^2,$$

which motivates to jointly diagonalize several of this quantities as utilized in the following extension of SIR.

Definition 39 (Spatial SIR in a BSS framework (SSIR))**.** *Let* $(y(\mathbf{s}), \mathbf{x}(\mathbf{s})^\top)^\top$ *be a* $(p+1)$*-variate random field following the SSDR model (Definition 38). Define the whitened version of* $\mathbf{x}(\mathbf{s})$ *by* $\mathbf{x}^{wh}(\mathbf{s}) = \operatorname{Cov}^{-1/2}(\mathbf{x}(\mathbf{s}))(\mathbf{x}(\mathbf{s}) - \mathbf{b})$*. Let* $\mathcal{L} \subset \mathbb{Z}^2$ *be a finite set of spatial lags. Let* $\mathbf{U} = \mathbf{U}((y(\mathbf{s}), \mathbf{x}(\mathbf{s})^\top)^\top)$ *be the orthogonal* $p \times p$ *joint diagonalization matrix of* $\operatorname{Cov}\left(\mathbb{E}\left(\mathbf{x}^{wh}(\mathbf{s})|y(\mathbf{s}+\mathbf{h})\right)\right)$ *for all* $\mathbf{h} \in \mathcal{L}$*, maximizing*

$$\sum_{\mathbf{h}\in\mathcal{L}} \left\| \operatorname{diag}\left(\mathbf{U}^\top \operatorname{Cov}\left(\mathbb{E}\left(\mathbf{x}^{wh}(\mathbf{s})|y(\mathbf{s}+\mathbf{h}\right)\right)\mathbf{U}\right)\right\|_F^2.$$

The spatial SIR (SSIR) unmixing matrix functional $\mathbf{W} = \mathbf{W}((y(\mathbf{s}), \mathbf{x}(\mathbf{s})^\top)^\top)$ *is given by* $\mathbf{W} = \mathbf{U}^\top \operatorname{Cov}^{-1/2}(\mathbf{x}(\mathbf{s}))$.

Publication VI studies the performance of the above method in an extensive simulation study and also discusses the identification of the signal dimension.

5. Conclusion

BSS for spatial data is a relatively new field in the statistical literature. Formulating SBSS by considering the location-scatter model with latent fields that are second order stationary is the starting point of this thesis. The core quantity for the estimation of the unmixing matrix for such a model is given by local covariance matrices. These matrices are formed by a weighted average of spatial covariance matrices where the weights are given by spatial kernel functions. In this thesis the original SBSS methodology is extended in three ways. Namely, allowing some latent components to be white noise, a mild drift or non-stationary second order spatial dependence. These three extensions are treated with asymptotic theory and bootstrap based tests, by replacing local covariance with differences and by computing local covariance matrices for sub-domains respectively. Furthermore, this thesis investigates SBSS as a pre-processing tool in the context of multivariate spatial prediction and formulates the popular SIR method in a BSS context for spatial data on a grid.

Naturally, many more extensions of the original SBSS methodology can be considered in future research. The BSS problem solved by simultaneous diagonalization of two scatter matrices can act as a way of robustifing SBSS. For example, a robust covariance estimator and local covariance matrices evaluated on signs might be a good choice that is motivated by robust time series BSS. Another direction applied in time series BSS is the use of higher order moments in the estimation of the unmixing matrix. This is useful when the latent time series follow models where all information lies in higher order time dependence such as ARCH or GARCH models. Similar, SBSS might be reformulated to account for higher order spatial dependence found in spatial ARCH and GARCH models (Otto et al., 2018; Otto and Schmid, 2020). For the spatial SIR adaptation it is currently restrictive that the allowed sample location patterns are lattices. Extensions for irregular sample locations could be achieved by using spatial kernel functions as in the original SBSS. Furthermore, SAVE might be also extended for spatial data similar as SIR. Generally, all extensions presented in this thesis lack results for the case when the sample size is increasing, thus, asymptotic results might be derived in future research.

Although almost all methods presented in this thesis consider d-dimensional spatial domains the focus lies on $d = 2$ which is the usual case for spatial data. However, measurements might also be taken for example on transects or in different depths of drill holes leading to spatial data where the domain is naturally one-dimensional. The practical usefulness of SBSS for such data might be explored in future work. In similar fashion, BSS for spatial data might be reformulated by considering functional data with spatial dependence as it is recently introduced for PCA by Kuenzer et al. (2021). Another direction is given by considering not only measurements taken at different spatial locations but also in time. Orthogonal diagonalization for such spatio-temporal data has already been considered in the literature by Iaco et al. (2013) but a full embedding of spatio-temporal data in the BSS framework is still an open research question.

References

J. Aitchison. *The Statistical Analysis of Compositional Data*. Blackburn Press, Caldwell, 2003.

E. B. Anderes and M. L. Stein. Local likelihood estimation for nonstationary random fields. *Journal of Multivariate Analysis*, 102(3):506–520, 2011. doi: 10.1016/j.jmva.2010.10.010.

T. V. Apanasovich, M. G. Genton, and Y. Sun. A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193, 2012. doi: 10.1080/01621459.2011.643197.

F. Bachoc, M. G. Genton, K. Nordhausen, A. Ruiz-Gazen, and J. Virta. Spatial blind source separation. *Biometrika*, 107(3):627–646, 2020. doi: 10.1093/biomet/asz079.

B. Bauer-Marschallinger, W. A. Dorigo, W. Wagner, and A. I. J. M. van Dijk. How oceanic oscillation drives soil moisture variations over mainland Australia: An analysis of 32 years of satellite observations. *Journal of Climate*, 26(24):10159–10173, 2013. doi: 10.1175/JCLI-D-13-00149.1.

A. Belouchrani and A. Cichocki. A robust whitening procedure in blind source separation context. *Electronics Letters*, 36(24):2050–2051, 2000. doi: 10.1049/el:20001436.

A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997. doi: 10.1109/78.554307.

E. Bura and R. D. Cook. Extending sliced inverse regression: The weighted chi-squared test. *Journal of the American Statistical Association*, 96(455):996–1003, 2001. doi: 10.1198/016214501753208979.

J.-F. Cardoso. Source separation using higher order moments. In *International Conference on Acoustics, Speech, and Signal Processing Proceedings*, pages 2109–2112, 1989. doi: 10.1109/ICASSP.1989.266878.

J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. In *IEEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 362–370, 1993. doi: 10.1049/ip-f-2.1993.0054.

J.-P. Chilés and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York, 2012. doi: 10.1002/9781118136188.

S. Choi and A. Cichocki. Blind separation of nonstationary sources in noisy mixtures. *Electronic Letters*, 36(9):848–849, 2000a. doi: 10.1049/el:20000623.

S. Choi and A. Cichocki. Blind separation of nonstationary and temporally correlated sources from noisy mixtures. In *Neural Networks for Signal Processing X. Proceedings of the 2000*

IEEE Signal Processing Society Workshop, volume 1, pages 405–414, 2000b. doi: 10.1109/NNSP.2000.889432.

S. Choi, A. Cichocki, and A. Belouchrani. Blind separation of second-order nonstationary and temporally colored sources. In *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, pages 444–447, 2001. doi: 10.1109/SSP.2001.955318.

S. Choi, A. Cichocki, and A. Beloucharni. Second order nonstationary source separation. *Journal of VLSI Signal Processing Systems*, 32(1-2):93–104, 2002. doi: 10.1023/A:1016319502849.

A. Cichocki and S.-I. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, New York, 2002. doi: 10.1002/0470845899.

D. B. Clarkson. Remark AS R74: A least squares version of algorithm AS 211: The F-G diagonalization algorithm. *Journal of the Royal Statistical Society: Series C*, 37(2):317–321, 1988. doi: 10.2307/2347359.

P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, Oxford, 2010. doi: 10.1016/C2009-0-19334-0.

F. Cong, Z. He, J. Hämäläinen, P. H. T. Leppänen, H. Lyytinen, A. Cichocki, and T. Ristaniemi. Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection. *Journal of Neuroscience Methods*, 212(1):165–172, 2013. doi: 10.1016/j.jneumeth.2012.09.029.

R. Cook and S. Weisberg. Sliced inverse regression for dimension reduction: Comment. *Journal of the American Statistical Association*, 86:328–332, 1991. doi: 10.2307/2290564.

R. D. Cook. Save: A method for dimension reduction and graphics in regression. *Communications in Statistics - Theory and Methods*, 29(9-10):2109–2121, 2000. doi: 10.1080/03610920008832598.

N. A. C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, New York, 1993. doi: 10.1002/9781119115151.

V. Dobrić and F. M. Ojeda. Fractional Brownian fields, duality, and martingales. In *High Dimensional Probability: Proceedings of the Fourth International Conference*, volume 51, pages 77–95, 2006. doi: 10.1214/074921706000000770.

X. Emery. Iterative algorithms for fitting a linear model of coregionalization. *Computers & Geosciences*, 36(9):1150–1160, 2010. doi: 10.1016/j.cageo.2009.10.007.

J. Eriksson and V. Koivunen. Identifiability, separability, and uniqueness of linear ICA models. *IEEE Signal Processing Letters*, 11(7):601–604, 2004. doi: 10.1109/LSP.2004.830118.

A. E. Gelfand, A. M. Schmidt, S. Banerjee, and C. F. Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, 2004. doi: 10.1007/BF02595775.

M. G. Genton and W. Kleiber. Cross-covariance functions for multivariate geostatistics. *Statistical Science*, 30(2):147–163, 2015. doi: 10.1214/14-STS487.

P. Georgiev and A. Cichocki. Blind source separation via symmetric eigenvalue decomposition. In *Proceedings of the Sixth International Symposium on Signal Processing and its Applications*, volume 1, pages 17–20, 2001. doi: 10.1109/ISSPA.2001.949764.

S. G. Ghurye and I. Olkin. A characterization of the multivariate normal distribution. *The Annals of Mathematical Statistics*, 33(2):533–541, 1962.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010. doi: 10.1198/jasa.2010.tm09420.

P. Goovaerts. Factorial kriging analysis: A useful tool for exploring the structure of multivariate spatial soil information. *Journal of Soil Science*, 43(4):597–619, 1992. doi: 10.1111/j.1365-2389.1992.tb00163.x.

P. Goovaerts. Ordinary cokriging revisited. *Mathematical Geology*, 30(1):21–42, 1998. doi: 10.1023/A:1021757104135.

M. Goulard and M. Voltz. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3):269–286, 1992. doi: 10.1007/BF00893750.

P. Guttorp and T. Gneiting. Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, 93(4):989–995, 2006. doi: 10.1093/biomet/93.4.989.

D. A. Harville. *Matrix Algebra From a Statistician's Perspective*. Springer, New York, 1997. doi: 10.1007/b98818.

J. Helterbrand and N. Cressie. Universal cokriging under intrinsic coregionalization. *Mathematical Geosciences*, 26(2):205–226, 1994. doi: 10.1007/BF02082764.

T. P. Hettmansperger and R. H. Randles. A practical affine equivariant multivariate median. *Biometrika*, 89(4):851–860, 2002. doi: 10.1093/biomet/89.4.851.

P. J. Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435–475, 1985. doi: 10.1214/aos/1176349519.

M. Hubert, M. Debruyne, and P. J. Rousseeuw. Minimum covariance determinant and extensions. *WIREs Computational Statistics*, 10(3):e1421, 2018. doi: 10.1002/wics.1421.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997. doi: 10.1162/neco.1997.9.7.1483.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999a. doi: 10.1109/72.761722.

A. Hyvärinen. Gaussian moments for noisy independent component analysis. *IEEE Signal Processing Letters*, 6(6):145–147, 1999b. doi: 10.1109/97.763148.

A. Hyvärinen. Blind source separation by nonstationarity of variance: A cumulant-based approach. *IEEE Transactions on Neural Networks*, 12(6):1471–1474, 2001. doi: 10.1109/72.963782.

S. D. Iaco, D. Myers, M. Palma, and D. Posa. Using simultaneous diagonalization to identify a space–time linear coregionalization model. *Mathematical Geosciences*, 45:69–86, 2013. doi: 10.1007/s11004-012-9408-3.

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

P. Ilmonen, J. Nevalainen, and H. Oja. Characteristics of multivariate distributions and the invariant coordinate system. *Statistics & Probability Letters*, 80(23):1844–1853, 2010a. doi: 10.1016/j.spl.2010.08.010.

P. Ilmonen, K. Nordhausen, H. Oja, and E. Ollila. A new performance index for ICA: Properties, computation and asymptotic analysis. In V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, editors, *Latent Variable Analysis and Signal Separation. LVA/ICA 2010. Lecture Notes in Computer Science*, volume 6365, pages 229–236, Berlin, Heidelberg, 2010b. Springer. doi: 10.1007/978-3-642-15995-4_29.

P. Ilmonen, H. Oja, and R. Serfling. On invariant coordinate system (ICS) functionals. *International Statistical Review*, 80(1):93–110, 2012. doi: 10.1111/j.1751-5823.2011.00163.x.

P. Ilmonen, K. Nordhausen, H. Oja, and F. Theis. An affine equivariant robust second-order BSS method. In E. Vincent, A. Yeredor, Z. Koldovský, and P. Tichavský, editors, *Latent Variable Analysis and Signal Separation. LVA/ICA 2015. Lecture Notes in Computer Science*, volume 9237, pages 328–335, Cham, 2015. Springer. doi: 10.1007/978-3-319-22482-4_38.

Z. Jiang. A survey on spatial prediction methods. *IEEE Transactions on Knowledge and Data Engineering*, 31(9):1645–1664, 2019. doi: 10.1109/TKDE.2018.2866809.

W. Kleiber and D. Nychka. Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis*, 112:76–91, 2012. doi: 10.1016/j.jmva.2012.05.011.

T. Kuenzer, S. Hörmann, and P. Kokoszka. Principal component analysis of spatially indexed functions. *Journal of the American Statistical Association*, 116(535):1444–1456, 2021. doi: 10.1080/01621459.2020.1732395.

H. R. Kunsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989. doi: 10.1214/aos/1176347265.

S. N. Lahiri. *Resampling Methods for Dependent Data*. Springer, New York, 2003. doi: 10.1007/978-1-4757-3803-2.

B. Li. *Sufficient Dimension Reduction*. Chapman and Hall/CRC, Boca Raton, 2017. doi: 10.1201/9781315119427.

J. Li and A. D. Heap. Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53:173–189, 2014. doi: 10.1016/j.envsoft.2013.12.008.

K.-C. Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991. doi: 10.2307/2290563.

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

E. Liski, K. Nordhausen, and H. Oja. Supervised invariant coordinate selection. *Statistics*, 48 (4):711–731, 2014. doi: 10.1080/02331888.2013.800067.

W. Luo and B. Li. Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika*, 103(4):875–887, 2016. doi: 10.1093/biomet/asw051.

W. Luo and B. Li. On order determination by predictor augmentation. *Biometrika*, 108(3): 557–574, 2021. doi: 10.1093/biomet/asaa077.

Y. Ma and L. Zhu. A review on dimension reduction. *International Statistical Review*, 81(1): 134–150, 2013. doi: 10.1111/j.1751-5823.2012.00182.x.

Y. Ma, J.-J. Royer, H. Wang, Y. Wang, and T. Zhang. Factorial kriging for multiscale modelling. *Journal of the Southern African Institute of Mining and Metallurgy*, 114(8):651–659, 2014.

M. Matilainen, K. Nordhausen, and H. Oja. New independent component analysis tools for time series. *Statistics & Probability Letters*, 105:80–87, 2015. doi: 10.1016/j.spl.2015.04.033.

M. Matilainen, C. Croux, K. Nordhausen, and H. Oja. Supervised dimension reduction for multivariate time series. *Econometrics and Statistics*, 4:57–69, 2017a. doi: 10.1016/j.ecosta.2017.04.002.

M. Matilainen, J. Miettinen, K. Nordhausen, H. Oja, and S. Taskinen. On independent component analysis with stochastic volatility models. *Austrian Journal of Statistics*, 46(3-4): 57–66, 2017b. doi: 10.17713/ajs.v46i3-4.671.

M. Matilainen, K. Nordhausen, and J. Virta. On the number of signals in multivariate time series. In Y. Deville, S. Gannot, R. Mason, M. Plumbley, and D. Ward, editors, *Latent Variable Analysis and Signal Separation. LVA/ICA 2018. Lecture Notes in Computer Science*, volume 10891, pages 248–258, Cham, 2018. Springer. doi: 10.1007/978-3-319-22482-4_38.

M. Matilainen, C. Croux, K. Nordhausen, and H. Oja. Sliced average variance estimation for multivariate time series. *Statistics*, 53(3):630–655, 2019. doi: 10.1080/02331888.2019.1605515.

K. Matsuoka, M. Ohoya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995. doi: 10.1016/0893-6080(94)00083-X.

D. S. Matteson and D. Ruppert. Time-series models of dynamic volatility and correlation. *IEEE Signal Processing Magazine*, 28(5):72–82, 2011. doi: 10.1109/MSP.2011.941553.

J. Miettinen, K. Nordhausen, H. Oja, and S. Taskinen. Statistical properties of a blind source separation estimator for stationary time series. *Statistics & Probability Letters*, 82(11):1865–1873, 2012. doi: 10.1016/j.spl.2012.06.025.

J. Miettinen, K. Nordhausen, H. Oja, and S. Taskinen. Fast equivariant JADE. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6153–6157, 2013. doi: 10.1109/ICASSP.2013.6638847.

J. Miettinen, K. Nordhausen, H. Oja, and S. Taskinen. Deflation-based FastICA with adaptive choices of nonlinearities. *IEEE Transactions on Signal Processing*, 62(21):5716–5724, 2014. doi: 10.1109/TSP.2014.2356442.

J. Miettinen, S. Taskinen, K. Nordhausen, and H. Oja. Fourth moments and independent component analysis. *Statistical Science*, 30(3):372–390, 2015. doi: 10.1214/15-STS520.

J. Miettinen, K. Illner, K. Nordhausen, H. Oja, S. Taskinen, and F. J. Theis. Separation of uncorrelated stationary time series using autocovariance matrices. *Journal of Time Series Analysis*, 37(3):337–354, 2016. doi: 10.1111/jtsa.12159.

J. Miettinen, K. Nordhausen, H. Oja, S. Taskinen, and J. Virta. The squared symmetric FastICA estimator. *Signal Processing*, 131(C):402–411, 2017. doi: 10.1016/j.sigpro.2016.08.028.

J. Miettinen, M. Matilainen, K. Nordhausen, and S. Taskinen. Extracting conditionally heteroskedastic components using independent component analysis. *Journal of Time Series Analysis*, 41(2):293–311, 2020. doi: 10.1111/jtsa.12505.

C. Muehlmann, F. Bachoc, K. Nordhausen, and M. Yi. Test of the latent dimension of a spatial blind source separation model. Submitted. Preprint at arXiv:2011.01711, 2020a.

C. Muehlmann, P. Filzmoser, and K. Nordhausen. Local difference matrices for spatial blind source separation. To appear in Proceedings of the 3rd Conference of the Arabian Journal of Geosciences, 2020b.

C. Muehlmann, K. Nordhausen, and M. Yi. On cokriging, neural networks, and spatial blind source separation for multivariate spatial prediction. *IEEE Geoscience and Remote Sensing Letters*, page n/a, 2020c. doi: 10.1109/LGRS.2020.3011549.

C. Muehlmann, F. Bachoc, and K. Nordhausen. Blind source separation for non-stationary random fields. Submitted. Preprint at arXiv:2107.01916, 2021a.

C. Muehlmann, K. Fačevicová, A. Gardlo, H. Janečková, and K. Nordhausen. Independent component analysis for compositional data. In A. Daouia and A. Ruiz-Gazen, editors, *Advances in Contemporary Statistics and Econometrics: Festschrift in Honor of Christine Thomas-Agnan*, pages 525–545. Springer, Cham, 2021b. doi: 10.1007/978-3-030-73249-3_27.

C. Muehlmann, P. Filzmoser, and K. Nordhausen. Spatial blind source separation in the presence of a drift. Submitted. Preprint at arXiv:2108.13813, 2021c.

C. Muehlmann, K. Nordhausen, and J. Virta. *SpatialBSS: Blind Source Separation for Multivariate Spatial Data*, 2021d. URL `https://CRAN.R-project.org/package=SpatialBSS`. R package version 0.11-0.

C. Muehlmann, H. Oja, and K. Nordhausen. Sliced inverse regression for spatial data. In E. Bura and B. Li, editors, *Festschrift in Honor of R. Dennis Cook: Fifty Years of Contribution to Statistical Science*, pages 87–107. Springer, Cham, 2021e. doi: 10.1007/978-3-030-69009-0_5.

K. Nordhausen. On robustifying some second order blind source separation methods for nonstationary time series. *Statistical Papers*, 55(1):141–156, 2014. doi: 10.1007/s00362-012-0487-5.

K. Nordhausen and H. Oja. Independent component analysis: A statistical perspective. *WIREs Computational Statistics*, 10(5):e1440, 2018. doi: 10.1002/wics.1440.

K. Nordhausen and D. E. Tyler. A cautionary note on robust covariance plug-in methods. *Biometrika*, 102(3):573–588, 2015. doi: 10.1093/biomet/asv022.

K. Nordhausen and J. Virta. Ladle estimator for time series signal dimension. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pages 428–432, 2018. doi: 10.1109/SSP.2018.8450695.

K. Nordhausen and J. Virta. An overview of properties and extensions of FOBI. *Knowledge-Based Systems*, 173:113–116, 2019. doi: 10.1016/j.knosys.2019.02.026.

K. Nordhausen, P. Ilmonen, A. Mandal, H. Oja, and E. Ollila. Deflation-based FastICA reloaded. In *Proceedings of the 19th European Signal Processing Conference*, pages 1854–1858, 2011.

K. Nordhausen, H. W. Gutch, H. Oja, and F. J. Theis. Joint diagonalization of several scatter matrices for ica. In F. J. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, editors, *Latent Variable Analysis and Signal Separation*, pages 172–179, Berlin, Heidelberg, 2012. Springer. doi: https://doi.org/10.1007/978-3-642-28551-6_22.

K. Nordhausen, H. Oja, P. Filzmoser, and C. Reimann. Blind source separation for spatial compositional data. *Mathematical Geosciences*, 47(7):753–770, 2015. doi: 10.1007/s11004-014-9559-5.

K. Nordhausen, H. Oja, and E. Ollila. Robust independent component analysis based on two scatter matrices. *Austrian Journal of Statistics*, 37(1):91–100, 2016. doi: 10.17713/ajs.v37i1.290.

K. Nordhausen, H. Oja, D. E. Tyler, and J. Virta. Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace. *IEEE Signal Processing Letters*, 24(6):887–891, 2017. doi: 10.1109/LSP.2017.2696880.

K. Nordhausen, G. Fischer, and P. Filzmoser. Blind source separation for compositional time series. *Mathematical Geosciences*, 53:905–924, 2021a. doi: 10.1007/s11004-020-09869-y.

K. Nordhausen, H. Oja, and D. E. Tyler. Asymptotic and bootstrap tests for subspace dimension. *Journal of Multivariate Analysis*, online first:104830, 2021b. doi: 10.1016/j.jmva.2021.104830.

H. Oja, S. Sirkiä, and J. Eriksson. Scatter matrices and independent component analysis. *Austrian Journal of Statistics*, 35(2-3):175–189, 2006. doi: 10.17713/ajs.v35i2\&3.364.

E. Ollila. The deflation-based FastICA estimator: Statistical analysis revisited. *IEEE Transactions on Signal Processing*, 58(3):1527–1541, 2010. doi: 10.1109/TSP.2009.2036072.

P. Otto and W. Schmid. Spatial and spatiotemporal GARCH models – A unified approach. Preprint at arXiv:1908.08320, 2020.

P. Otto, W. Schmid, and R. Garthoff. Generalised spatial and spatiotemporal autoregressive conditional heteroscedasticity. *Spatial Statistics*, 26:125–145, 2018. doi: 10.1016/j.spasta.2018.07.005.

Y. Pan, M. Matilainen, S. Taskinen, and K. Nordhausen. A review of second-order blind identification methods. *WIREs Computational Statistics*, n/a(n/a):e1550, 2021. doi: 10.1002/wics.1550.

N. Pfister, S. Weichwald, P. Bühlmann, and B. Schölkopf. Robustifying independent component analysis by adjusting for group-wise stationary noise. *Journal of Machine Learning Research*, 20(147):1–50, 2019. doi: 10.3929/ethz-b-000374036.

D. Pham. Exploiting source non stationary and coloration in blind source separation. In *14th International Conference on Digital Signal Processing Proceedings. DSP 2002*, volume 1, pages 151–154, 2002. doi: 10.1109/ICDSP.2002.1027859.

D.-T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Transactions on Signal Processing*, 49(9):1837–1848, 2001. doi: 10.1109/78.942614.

N. Piccolotto, M. Bögl, T. Gschwandtner, C. Muehlmann, K. Nordhausen, P. Filzmoser, and S. Miksch. TBSSvis: Visual analytics for temporal blind source separation. Submitted. Preprint at arXiv:2011.09896, 2020.

D. N. Politis and J. P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994. doi: 10.2307/2290993.

C. Reimann, M. Äyräs, and V. Chekushin. *Environmental Geochemical Atlas of the Central Barents Region*. Schweizerbart Science Publishers, Stuttgart, 2010.

C. Reimann, M. Birke, A. Demetriades, P. Filzmoser, and P. O'Connor. *Chemistry of Europe's Agricultural Soils, Part A*. Schweizerbart Science Publishers, Stuttgart, 2014.

P. D. Sampson. Constructions for nonstationary spatial processes. In A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, editors, *Handbook of Spatial Statistics*, pages 119–130. CRC Press, Boca Raton, 2010. doi: 10.1201/9781420072884-c9.

Z. Shi, Z. Jiang, F. Zhou, and J. Yin. Blind source separation with nonlinear autocorrelation and non-Gaussianity. *Journal of Computational and Applied Mathematics*, 229(1):240–247, 2009. doi: 10.1016/j.cam.2008.10.031.

R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications. With R Examples*. Springer, Cham, 2011. doi: 10.1007/978-3-319-52452-8.

A. Stein and L. C. A. Corsten. Universal Kriging and cokriging as a regression procedure. *Biometrics*, 47(2):575–587, 1991. doi: 10.2307/2532147.

M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999. doi: 10.1007/978-1-4612-1494-6.

A. Tanaka, H. Imai, and M. Miyakoshi. Theoretical foundations of second-order-statistics-based blind source separation for non-stationary sources. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 3, pages III–III, 2006. doi: 10.1109/ICASSP.2006.1660725.

A. C. Tang, J.-Y. Liu, and M. T. Sutherland. Recovery of correlated neuronal sources from EEG: The good and bad ways of using SOBI. *NeuroImage*, 28(2):507–519, 2005. doi: 10.1016/j.neuroimage.2005.06.062.

S. Taskinen, S. Sirkiä, and H. Oja. Independent component analysis based on symmetrised scatter matrices. *Computational Statistics & Data Analysis*, 51(10):5103–5111, 2007. doi: 10.1016/j.csda.2006.07.010.

R. C. Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

F. Theis and Y. Inouye. On the use of joint diagonalization in blind signal processing. In *2006 IEEE International Symposium on Circuits and Systems*, pages 3586–3589, 2006. doi: 10.1109/ISCAS.2006.1693402.

F. J. Theis, N. S. Müller, C. Plant, and C. Böhm. Robust second-order source separation identifies experimental responses in biomedical imaging. In V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, editors, *Latent Variable Analysis and Signal Separation*, pages 466–473, Berlin, Heidelberg, 2010. Springer. doi: 10.1007/978-3-642-15995-4_58.

L. Tong, V. Soon, Y. Huang, and R. Liu. AMUSE: A new blind identification algorithm. In *IEEE International Symposium on Circuits and Systems*, volume 3, pages 1784–1787, 1990. doi: 10.1109/ISCAS.1990.111981.

L. Tong, Y. Inouye, and R. Liu. A finite-step global convergence algorithm for the cumulant-based parameter estimation of multichannel moving average processes. In *ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing Proceedings*, volume 5, pages 3445–3448, 1991. doi: 10.1109/ICASSP.1991.150195.

L. Tong, R. Liu, V. C. Soon, and Y. Huang. Indeterminacy and identifiability of blind identification. *IEEE Transactions on Circuits and Systems*, 38(5):499–509, 1991. doi: 10.1109/31.76486.

D. E. Tyler, F. Critchley, L. Dümbgen, and H. Oja. Invariant co-ordinate selection. *Journal of the Royal Statistical Society: Series B*, 71(3):549–592, 2009. doi: 10.1111/j.1467-9868.2009.00706.x.

J. Vargas-Guzmán, A. Warrick, and D. Myers. Coregionalization by linear combination of nonorthogonal components. *Mathematical Geology*, 34(4):405–419, 2002. doi: 10.1023/A:1015078911063.

J. Virta. *Independent component analysis for non-standard data structures*. PhD thesis, University of Turku – Annales Universitatis Turkuensis, 2018.

J. Virta and K. Nordhausen. Estimating the number of signals using principal component analysis. *Stat*, 8(1):e231, 2019. doi: 10.1002/sta4.231.

J. Virta and K. Nordhausen. Determining the signal dimension in second order source separation. *Statistica Sinica*, 31(1):135–156, 2021. doi: 10.5705/ss.202018.0347.

J. Virta, K. Nordhausen, and H. Oja. Projection pursuit for non-Gaussian independent components. Preprint at arXiv:1612.05445, 2016.

H. Wackernagel. Cokriging versus Kriging in regionalized multivariate data analysis. *Geoderma*, 62(1):83–92, 1994. doi: 10.1016/0016-7061(94)90029-9.

H. Wackernagel. *Multivariate Geostatistics*. Springer, Berlin, Heidelberg, 2003. doi: 10.1007/978-3-662-05294-5.

H. Wang, Y. Guan, and B. Reich. Nearest-neighbor neural networks for geostatistics. In *Proceedings of the 2019 International Conference on Data Mining Workshops (ICDMW)*, pages 196–205, 2019. doi: 10.1109/ICDMW.2019.00038.

T. Wei. A convergence and asymptotic analysis of the generalized symmetric FastICA algorithm. *IEEE Transactions on Signal Processing*, 63(24):6445–6458, 2015. doi: 10.1109/TSP.2015.2468686.

A. Yeredor. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Transactions on Signal Processing*, 50(7):1545–1553, 2002. doi: 10.1109/TSP.2002.1011195.

Curriculum vitae

Personal data

Name	**Christoph Mühlmann**
Date of Birth	10.07.1992
Nationality	Austria
E-Mail	christoph.muehlmann@tuwien.ac.at

Work experience

04/2019 - present	FWF project assistant at the CSTAT group, Institute of Statistics and Mathematical Methods in Economics, TU Wien, Austria
02/2018 - 03/2019	Data analyst at the CRESST experiment searching for dark matter, Institute of High Energy Physics (HEPHY) of the Austrian Academy of Sciences (ÖAW), Austria
03/2017 - 09/2017	Project student at the qBounce experiment, Atominstitut, TU Wien, Austria
07/2010 - 08/2010	Construction design and implementation of an electrical motor test bench, KS Engineers, Graz, Austria

Education

04/2019 - present	**Doctoral program** in Natural Sciences Technical Mathematics, TU Wien, Austria
10/2016 - 03/2019	**Dipl.-Ing.**, Master program in Technical Physics, TU Wien, Austria
10/2012 - 09/2016	**BSc**, Bachelor program in Technical Physics, TU Graz, Austria
09/2006 - 06/2011	**General Qualification for University Entrance**, Automation Engineering program, HTBLA Kaindorf, Austria

List of presentations

Talk, Blind source separation for non-stationary random fields (Data Science, Statistics & Visualisation, DSSV 2021)

Talk, Blind source separation for non-stationary random fields (The 13th Workshop on Spatial Statistics and Image Analysis in Biology, SSIAB13, 2021)

Talk, Local difference matrices for spatial blind source separation (CAJG 2020)

Talk, Spatial supervised dimension reduction in a blind source separation framework (Machine Learning Seminar, University Paul Sabatier, Toulouse, France)

Talk, Supervised dimension reduction for spatial data (CMStatistics 2019)

Talk, Supervised dimension reduction for spatial data (Statistiktage 2019 der Österreichischen Statistischen Gesellschaft)

Talk, Pulse-Shape discrimination with deep learning in CRESST (CRESST Collaboration Meeting November 2018)

Talk, Pulse-Shape discrimination with deep learning in CRESST (Colloquium on Machine Learning in Particle Physics, ESA Munich, 2018)

Talk, Pulse-Shape discrimination with deep learning in CRESST (CRESST Collaboration Meeting June 2018)

List of software

C. Muehlmann, K. Nordhausen, and J. Virta. SpatialBSS: Blind source separation for multivariate spatial data, 2021. `https://CRAN.R-project.org/package=SpatialBSS`. R package.

List of articles

C. Muehlmann, P. Filzmoser, and K. Nordhausen. Local difference matrices for spatial blind source separation. To appear in proceedings of the 3rd conference of the Arabian journal of geosciences, 2020.

C. Muehlmann, K. Nordhausen, and M. Yi. On Cokriging, neural networks, and spatial blind source separation for multivariate spatial prediction. *IEEE Geoscience and Remote Sensing Letters*, pages 1–5, 2020. doi: 10.1109/LGRS.2020.3011549.

C. Muehlmann, H. Oja, and K. Nordhausen. Sliced inverse regression for spatial data. In E. Bura and B. Li, editors, *Festschrift in Honor of R. Dennis Cook: Fifty Years of Contribution to Statistical Science*, pages 87–107. Springer, Cham, 2021. doi: 10.1007/978-3-030-69009-0_5.

C. Muehlmann, K. Fačevicová, A. Gardlo, H. Janečková and K. Nordhausen. Independent component analysis for compositional data. In A. Daouia and A. Ruiz-Gazen, editors, *Advances in Contemporary Statistics and Econometrics: Festschrift in Honor of Christine Thomas-Agnan*, pages 525–545. Springer, Cham, 2021. doi: 10.1007/978-3-030-73249-3_27.

C. Muehlmann, F. Bachoc, and K. Nordhausen. Blind source separation for non-stationary random fields. Submitted. Preprint at arXiv:2107.01916, 2021.

C. Muehlmann, P. Filzmoser, and K. Nordhausen. Spatial blind source separation in the presence of a drift. Submitted. Preprint at arXiv:2108.13813, 2021.

N. Piccolotto, M. Bögl, T. Gschwandtner, **C. Muehlmann**, K. Nordhausen, P. Filzmoser and S. Miksch. TBSSvis: Visual analytics for temporal blind source separation. Submitted. Preprint at arXiv:2011.09896, 2020.

C. Muehlmann, F. Bachoc, K. Nordhausen, and M. Yi. Test of the latent dimension of a spatial blind source separation model. Submitted. Preprint at arXiv:2011.01711, 2020.

Vienna, October 13, 2021

Christoph Mühlmann

Part II.

Publications

Test of the latent dimension of a spatial blind source separation model

Summary

Publication I considers the spatial blind source separation model introduced by Nordhausen et al. (2015); Bachoc et al. (2020) but assumes that some components of the latent field are white noise. The goal is to determine the signal subspace of the latent field. For this purpose bootstrap test strategies are introduced and a test based on asymptotic results is derived where both approaches are found to be useful in an extensive simulation study and on a real data example.

Bibliographic information

C. Muehlmann, F. Bachoc, K. Nordhausen, and M. Yi. Test of the latent dimension of a spatial blind source separation model. Submitted. Preprint at arXiv:2011.01711, 2020.

Author's contribution

C. Muehlmann participated in several discussions with the coauthors to develop the idea and the methodology. Furthermore, C. Muehlmann implemented all methods in R, carried out all simulations as well as the real data example. C. Muehlmann wrote the bootstrap, simulation and real data part of the draft and finalized the draft based on joint discussions with the coauthors.

Local difference matrices for spatial blind source separation

Summary

Publication II considers the spatial blind source separation model as introduced by Nordhausen et al. (2015); Bachoc et al. (2020) for the case when the location vector might depend on the actual sample location. Hence, for the case when a drift is present. As local covariance matrices rely on the estimation of a global location the drift significantly worsens the performance of the methods. To overcome this issue local covariance matrices are adapted to consider difference processes in favor of the original one. This avoids the estimation of a global mean. A geochemical dataset is analyzed with this approach.

Bibliographic information

C. Muehlmann, P. Filzmoser, and K. Nordhausen. Local difference matrices for spatial blind source separation. To appear in Proceedings of the 3rd Conference of the Arabian Journal of Geosciences, 2020.

Author's contribution

C. Muehlmann participated in several discussions with the coauthors to develop the idea and the methodology. Furthermore, C. Muehlmann implemented all methods in R and carried out the real data example. C. Muehlmann wrote the first version of the draft and finalized the draft based on joint discussions with the coauthors.

Spatial blind source separation in the presence of a drift

Summary

Publication III is an extended version of Publication II. In addition to Publication II this article adapts all spatial blind source separation methods introduced by Nordhausen et al. (2015); Bachoc et al. (2020) to use local difference matrices in contrast to local covariance matrices. Furthermore, the whitening procedure is also adapted to be robust against a drift present in the data. This is achieved by replacing the covariance matrix by a local difference matrix. The statistical model is also adapted to relax the stationarity assumptions to intrinsic stationarity. The new methods are shown to be useful in an extensive simulation study and on a real data example.

Bibliographic information

C. Muehlmann, P. Filzmoser, and K. Nordhausen. Spatial blind source separation in the presence of a drift. Submitted. Preprint at arXiv:2108:13813, 2021.

Author's contribution

C. Muehlmann participated in several discussions with the coauthors to develop the idea and the methodology. Furthermore, C. Muehlmann implemented all methods in R and carried out all simulations as well as the real data example. C. Muehlmann wrote the first version of the draft and finalized the draft based on joint discussions with the coauthors.

Blind source separation for non-stationary random fields

Summary

Publication IV introduces spatial blind source separation for the case when the components of the latent field exhibit non-stationary second order spatial dependence. A new spatial non-stationary source separation model is formulated and three methods to recover the latent field are introduced. The affine equivariance property and parameter identifiability conditions are derived for all three methods. These methods are an adaptation of the methods described by Nordhausen (2014) for the spatial setting. The new estimators are evaluated in an extensive simulation study and on a geochemical application.

Bibliographic information

C. Muehlmann, F. Bachoc, and K. Nordhausen. Blind source separation for non-stationary random fields. Submitted. Preprint at arXiv:2107.01916, 2021.

Author's contribution

C. Muehlmann participated in several discussions with the coauthors to develop the idea and the methodology. Furthermore, C. Muehlmann implemented all methods in R and carried out the simulations and the real data example. C. Muehlmann wrote the first version of the draft and finalized the draft based on joint discussions with the coauthors.

On Cokriging, neural networks, and spatial blind source separation for multivariate spatial prediction

Summary

Publication V considers the spatial blind source separation methods introduced by Nordhausen et al. (2015); Bachoc et al. (2020) in the context of spatial prediction. Specifically, predictions of the original multivariate random field can be formed by predicting each component of the latent field individually with any univariate spatial prediction tool and then mixing these predictions again. This approach is motivated by the fact that the components of the latent field are uncorrelated or even independent. Predictions with this procedure are compared with predictions from Cokriging and neural networks in an extensive simulation study and on a data example.

Bibliographic information

C. Muehlmann, K. Nordhausen, and M. Yi. On Cokriging, neural networks, and spatial blind source separation for multivariate spatial prediction. *IEEE Geoscience and Remote Sensing Letters*, pages 1–5, 2020. doi: 10.1109/LGRS.2020.3011549.

Author's contribution

C. Muehlmann participated in several discussions with the coauthors to develop the idea and the methodology. Furthermore, C. Muehlmann implemented all methods in R and carried out the simulations as well as the real data example. C. Muehlmann wrote the first version of the draft and finalized the draft based on joint discussions with the coauthors.

Sliced inverse regression for spatial data

Summary

Certain sufficient dimension reduction methods such as sliced inverse regression or sliced average variance estimation can be embedded in the blind source separation methodology. This is done by Matilainen et al. (2017a, 2019) for time series data. Publication VI formulates sliced inverse regression for stationary spatial data observed on a grid and evaluates the method in an extensive simulation study.

Bibliographic information

C. Muehlmann, H. Oja, and K. Nordhausen. Sliced inverse regression for spatial data. In E. Bura and B. Li, editors, *Festschrift in Honor of R. Dennis Cook: Fifty Years of Contribution to Statistical Science*, pages 87–107. Springer, Cham, 2021. doi: 10.1007/978-3-030-69009-0_5.

Author's contribution

C. Muehlmann participated in several discussions with the coauthors to develop the idea and the methodology. Furthermore, C. Muehlmann implemented all methods in R and carried out all simulations. C. Muehlmann wrote the first version of the draft and finalized the draft based on joint discussions with the coauthors.