JYX JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version
may differ from the original in pagination and typographic details.

Author(s): Pan, Yan; Matilainen, Markus; Taskinen, Sara; Nordhausen, Klaus

Title: A review of second-order blind identification methods

Year: 2022

Version: Published version

Copyright: © 2021 The Authors. WIREs Computational Statistics published by Wiley Periodica

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

OVERVIEW

WIREs COMPUTATIONAL STATISTICS WILEY

A review of second-order blind identification methods

Yan Pan[1] | Markus Matilainen[2] | Sara Taskinen[1] | Klaus Nordhausen[3,1]

[1]Department of Mathematics and Statistics, University of Jyväskylä, Finland

[2]Turku PET Centre, Turku University Hospital and University of Turku, Finland

[3]Institute of Statistics and Mathematical Methods in Economics, TU, Vienna, Austria

Correspondence
Klaus Nordhausen, Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology Wiedner Hauptstr. 7, A-1040 Vienna, Austria.
Email: klaus.nordhausen@tuwien.ac.at

Funding information
Austrian Science Fund, Grant/Award Number: P31881-N32

Edited by James E. Gentle and David W. Scott, Editors-in-Chief.

Abstract

Second-order source separation (SOS) is a data analysis tool which can be used for revealing hidden structures in multivariate time series data or as a tool for dimension reduction. Such methods are nowadays increasingly important as more and more high-dimensional multivariate time series data are measured in numerous fields of applied science. Dimension reduction is crucial, as modeling such high-dimensional data with multivariate time series models is often impractical as the number of parameters describing dependencies between the component time series is usually too high. SOS methods have their roots in the signal processing literature, where they were first used to separate source signals from an observed signal mixture. The SOS model assumes that the observed time series (signals) is a linear mixture of latent time series (sources) with uncorrelated components. The methods make use of the second-order statistics—hence the name "second-order source separation." In this review, we discuss the classical SOS methods and their extensions to more complex settings. An example illustrates how SOS can be performed.

This article is categorized under:
   Statistical Models > Time Series Models
   Statistical and Graphical Methods of Data Analysis > Dimension Reduction
   Data: Types and Structure > Time Series, Stochastic Processes, and Functional Data

KEYWORDS

blind source separation, dimension reduction, joint diagonalization, multivariate time series

1 | INTRODUCTION

Many fields of science (e.g., engineering, finance, and biomedicine) collect nowadays multivariate time series data which consist of correlated component series. As the data are often very high-dimensional, fitting multivariate time series models with large number of parameters is computationally impractical unless the models are noticeably simplified. Blind source separation (BSS) is a widely used multivariate method for extracting latent, uncorrelated source signals from observed multivariate time series data. BSS often provide a deeper inside into the data by revealing hidden structures and the methods can also be used as a preprocessing step to reduce the dimension of the problem at hand. It

is, for example, much easier to model the uncorrelated components individually and transform the results back to the original scale than fitting a multivariate model at the original scale.

BSS methods originate from the signal processing literature (see Comon & Jutten, 2010, and references therein), but have gained recently increasing interest also among statisticians (see Nordhausen & Oja, 2018, for example). In the blind source separation model, under consideration in this review, we assume that an observable p-variate time series $\boldsymbol{x} = (\boldsymbol{x}_t)_{t \in \mathbb{Z}}$ is a linear instantaneous mixture of a p-variate latent vector $\boldsymbol{z} = (\boldsymbol{z}_t)_{t \in \mathbb{Z}}$ with uncorrelated and possibly even independent components. The aim of BSS is then to recover original latent vectors $\boldsymbol{z}$ (sources) given the observed time series data. In this article, we concentrate on the second-order separation (SOS) model, a submodel of BSS, where we assume that the components of $\boldsymbol{z}$ are uncorrelated. Notice that the SOS model does not make any other structural assumptions on the latent vectors (such as a special time series model structure).

The starting point for the development of SOS methods was the algorithm for multiple unknown signals extraction (AMUSE, Tong et al., 1990) and its generalization second-order blind identification (SOBI, Belouchrani et al., 1997), which make use of the temporal dependence of the components via joint diagonalization of one or several autocovariance matrices, respectively. Some recent applications of SOBI include automatic removal of artifacts in EEG data (Joyce et al., 2004), finding interesting signals in brain imaging data (Tang, 2010; Tang, Sutherland, & McKinney, 2005), separating vibrations caused by underground traffic (Popescu & Manolescu, 2007), and forecasting of wind speed (Firat et al., 2010). AMUSE and SOBI are well-established and widely studied methods for the separation of uncorrelated weakly stationary time series (Miettinen et al., 2012, 2014, 2016). However, for other BSS models and different time series settings, these methods need to be modified accordingly. The aim of this review is twofold. First, we review the classical AMUSE and SOBI methods and discuss their statistical properties in detail. Second, we give a comprehensive review of recent generalizations of AMUSE and SOBI methods which are developed to perform well for more complex BSS models.

The structure of this overview is as follows. In Section 2, we review the original AMUSE and SOBI methods and discuss the statistical properties of corresponding signal separation matrices. In Section 3, we list some limitations the original SOBI method has and go through several modifications and generalizations of the method. In Section 4, we demonstrate in an application how SOBI fails to separate electrocardiography (ECG) recordings and why extensions of SOBI are relevant. Section 5 lists available algorithms and software for the methods discussed. Finally, in Section 6, we conclude the article with a brief summary and discuss possible future work in this area.

2 │ AMUSE AND SOBI

Let us first recall the original SOBI algorithm and its special case, the algorithm for multiple unknown signals extraction (AMUSE). These two methods were proposed by Belouchrani et al. (1997) and Tong et al. (1990), respectively, to solve the second-order source separation (SOS) problem. Hence, consider a second-order separation model

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{z}_t, \quad t \in \mathbb{Z}, \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is a location vector and $\boldsymbol{A} \in \mathbb{R}^{p \times p}$ is the mixing matrix. For the p-variate latent time series $\boldsymbol{z} = (\boldsymbol{z}_t)_{t \in \mathbb{Z}}$ we assume that

$$\mathrm{E}(\boldsymbol{z}_t) = \boldsymbol{0} \quad \text{and} \quad \mathrm{Cov}(\boldsymbol{z}_t) = \boldsymbol{I}_p, \tag{2}$$

and that the p processes in $\boldsymbol{z} \in \mathbb{R}^p$ are assumed to be jointly weakly stationary and mutually uncorrelated meaning that

$$\mathrm{E}\left(\boldsymbol{z}_t \boldsymbol{z}'_{t+\tau}\right) = \mathrm{E}\left(\boldsymbol{z}_{t+\tau} \boldsymbol{z}'_t\right) = \boldsymbol{D}_\tau, \tag{3}$$

for all lags $\tau > 0$, where $\boldsymbol{D}_\tau$ is a $p \times p$ diagonal matrix. We also assume that all latent sources have different autocovariance structures. Notice that in this article, we assume a real-valued SOS model. However, many of the presented methods are or can be extended to the case of complex-valued times series. The aim of SOS is to find a signal separation matrix $\boldsymbol{W} \in \mathbb{R}^{p \times p}$ such that the components of $\boldsymbol{W}(\boldsymbol{x}_t - \boldsymbol{\mu}) = \boldsymbol{J}\boldsymbol{P}\boldsymbol{z}_t$ are standardized and mutually uncorrelated, for a sign-change matrix $\boldsymbol{J} \in \mathcal{J}$ and a permutation matrix $\boldsymbol{P} \in \mathcal{P}$. Here $\mathcal{J}$ denotes a set of $p \times p$ sign-change matrices, whereas $\mathcal{P}$ is a set of $p \times p$ permutation matrices.

Both AMUSE and SOBI proceed as follows. Write the standardized time series as $\boldsymbol{x}_t^{st} = \text{Cov}(\boldsymbol{x}_t)^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu})$, where $\text{Cov}(\boldsymbol{x}_t)^{-1/2}$ is the symmetric square root of $\text{Cov}(\boldsymbol{x}_t)$. Then there exists an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{p \times p}$ such that $\boldsymbol{x}_t^{st} = \boldsymbol{U}\boldsymbol{z}_t$ (Cardoso & Souloumiac, 1993; Miettinen et al., 2015). Thus, after standardization, the estimation problem reduces to finding an orthogonal matrix $\boldsymbol{U}$ only. The signal separation matrix that solves the SOS problem is then given by $\boldsymbol{W} = \boldsymbol{U}\,\text{Cov}(\boldsymbol{x}_t)^{-1/2}$.

AMUSE (Tong et al., 1990) uses the cross-autocovariance matrix

$$\text{Cov}_\tau(\boldsymbol{x}) = \text{E}\left((\boldsymbol{x}_t - \boldsymbol{\mu})(\boldsymbol{x}_{t+\tau} - \boldsymbol{\mu})'\right) \qquad (4)$$

with a given lag $\tau > 0$ to solve the SOS problem. An orthogonal matrix $\boldsymbol{U} = (\boldsymbol{u}_1, ..., \boldsymbol{u}_p)'$ is found simply using the eigendecomposition of the autocovariance matrix as, by assumption, $\text{Cov}_\tau(\boldsymbol{x}_t^{st}) = \boldsymbol{U}\,\text{Cov}_\tau(\boldsymbol{z})\boldsymbol{U}' = \boldsymbol{U}\boldsymbol{D}_\tau\boldsymbol{U}'$ for any lag τ. Equivalently, $\boldsymbol{U}$ is found by maximizing

$$\| \text{diag}\left(\boldsymbol{U}\,\text{Cov}_\tau(\boldsymbol{x}_t^{st})\boldsymbol{U}'\right)\|^2 = \sum_{i=1}^{p}\left(\boldsymbol{u}_i'\,\text{Cov}_\tau(\boldsymbol{x}_t^{st})\boldsymbol{u}_i\right)^2 \qquad (5)$$

for a given lag τ, under the orthogonality constraint $\boldsymbol{U}\boldsymbol{U}' = \boldsymbol{I}_p$. Here $\| \cdot \|$ is the matrix (Frobenius) norm. Notice that the signal separation matrix $\boldsymbol{W} = \boldsymbol{U}\,\text{Cov}(\boldsymbol{x}_t)^{-1/2}$ is uniquely defined only if all the eigenvalues of $\text{Cov}_\tau(\boldsymbol{x}_t^{st})$ are distinct. The signal separation matrix $\boldsymbol{W}$ is affine equivariant in the sense that if $\boldsymbol{x}_t^* = \boldsymbol{B}\boldsymbol{x}_t + \boldsymbol{b}$ for all nonsingular $\boldsymbol{B} \in \mathbb{R}^{p \times p}$ and for all vectors $\boldsymbol{b} \in \mathbb{R}^p$, then the signal separation matrix based on $\boldsymbol{x}_t^*$ is $\boldsymbol{W}^* = \boldsymbol{W}\boldsymbol{B}^{-1}$ up to sign changes of the components (Miettinen et al., 2012). In practice, affine equivariance means that the separation result does not depend on the mixing procedure. This is a desirable property for a BSS method.

The main shortcoming of the AMUSE method is that it uses only one cross-autocovariance matrix for signal separation. Its performance thus depends heavily on the chosen lag τ. There are several extensions to AMUSE found in the literature which aim at better separation results (see, e.g., Chapter 7 of Comon & Jutten, 2010). One extension of AMUSE has become recently very popular. Belouchrani et al. (1997) introduced the SOBI method that jointly diagonalizes several autocovariance matrices simultaneously. Like AMUSE, the SOBI signal separation matrix is also affine equivariant. Notice that as more than two matrices can be exactly jointly diagonalized only if they commute with each other, as is the case here at the population level, approximate joint diagonalization of the autocovariance matrices is needed in practice as the matrices are estimated with an error. The solution for the signal separation matrix $\boldsymbol{W}$ in SOBI is thus achieved by formulating the maximization problem

$$\sum_{\tau \in \mathcal{T}} \| \text{diag}\left(\boldsymbol{U}\,\text{Cov}_\tau(\boldsymbol{x}_t^{st})\boldsymbol{U}'\right)\|^2 = \sum_{\tau \in \mathcal{T}}\sum_{i=1}^{p}\left(\boldsymbol{u}_i'\,\text{Cov}_\tau(\boldsymbol{x}_t^{st})\boldsymbol{u}_i\right)^2, \qquad (6)$$

for an orthogonal matrix $\boldsymbol{U}$, where $\mathcal{T} = \{\tau_1, ..., \tau_K\} \subset \mathbb{Z}_+$ is a set of selected lags. As before, the SOBI signal separation matrix is given by $\boldsymbol{W} = \boldsymbol{U}\,\text{Cov}(\boldsymbol{x}_t)^{-1/2}$. In symmetric SOBI the rows of $\boldsymbol{U}$ are found simultaneously (Miettinen et al., 2016), and the most efficient algorithm for such simultaneous joint diagonalization is based on Jacobi rotations (Clarkson, 1988). In deflation-based SOBI the uncorrelated components are found one by one (Miettinen et al., 2014). These two approaches are discussed and compared thoroughly in Illner et al. (2015) and in Miettinen et al. (2016). In SOBI, we use (6) as a criterion for joint diagonalization. For other proposals for joint diagonalization criteria, see Ziehe et al. (2003); Yeredor (2000); Pham (2001), for example. Yeredor (2002) suggests an algorithm (Alternating Columns—Diagonal Centers, ACDC), where the diagonalizer does not have to be an orthogonal matrix. For a summary of joint diagonalization methods used in blind signal separation, see also Theis and Inouye (2006).

The AMUSE and SOBI estimates for $\boldsymbol{W} \in \mathbb{R}^{p \times p}$ are obtained by replacing $\boldsymbol{\mu}$, $\text{Cov}(\boldsymbol{x}_t)$ and $\text{Cov}_\tau(\boldsymbol{x}_t)$ in the estimation procedure by the sample mean vector $\bar{\boldsymbol{x}}$, the sample covariance matrix $\widehat{\text{Cov}}(\boldsymbol{x}_t)$ and the symmetrized autocovariance matrix

$$\widehat{\text{Cov}}_\tau^S(\boldsymbol{x}_t) = \frac{1}{2}\left(\widehat{\text{Cov}}_\tau(\boldsymbol{x}_t) + \widehat{\text{Cov}}_\tau(\boldsymbol{x}_t)'\right), \qquad (7)$$

where

$$\widehat{\text{Cov}}_\tau(\boldsymbol{x}) = \frac{1}{T-\tau}\sum_{t=1}^{T-\tau}\boldsymbol{x}_t\boldsymbol{x}'_{t-\tau}, \quad \tau = 0,1,2,\ldots,$$

respectively. The estimates of $\widehat{\boldsymbol{U}}$ are then obtained using (5) and (6) for AMUSE and SOBI, respectively, and the signal separation matrix estimates are obtained as $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{U}}\widehat{\text{Cov}}(\boldsymbol{x}_t)^{-1/2}$.

In Miettinen et al. (2012, 2014, 2016) it was shown that, for both AMUSE and SOBI methods, the joint limiting distribution of a signal separation matrix $\widehat{\boldsymbol{W}}$, based on an observed time series $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$, depends on the joint limiting distribution of the sample covariance matrix and the symmetrized autocovariance matrices. Limiting distributions in case of multivariate MA(∞) processes are considered as special case. Consider now the model (1) with $\boldsymbol{\mu} = \boldsymbol{0}$ (wlog). We assume that $\boldsymbol{z}_t$ is a multivariate MA(∞) process that fulfills the assumptions (2) and (3). We also assume that the components of $\boldsymbol{z}_t$ have finite fourth moments and they are exchangeable and marginally symmetric, that is, $\boldsymbol{JPz}_t \sim \boldsymbol{z}_t$, for all sign change matrices $\boldsymbol{J}$ and permutation matrices $\boldsymbol{P}$. As AMUSE and SOBI estimates $\widehat{\boldsymbol{W}}$ are affine equivariant, we can assume that $\boldsymbol{W} = \boldsymbol{I}_p$. Thus $\text{Cov}_\tau^S(\boldsymbol{z}_t) = \boldsymbol{D}_\tau$, in which the diagonal elements are distinct and in decreasing order. The joint limiting distribution of

$$\sqrt{T}\begin{pmatrix} \text{vec}\left(\widehat{\text{Cov}}(\boldsymbol{z}_t) - \boldsymbol{I}_p\right) \\ \text{vec}\left(\widehat{\text{Cov}}_{\tau_1}^S(\boldsymbol{z}_t) - \boldsymbol{D}_{\tau_1}\right) \\ \vdots \\ \text{vec}\left(\widehat{\text{Cov}}_{\tau_K}^S(\boldsymbol{z}_t) - \boldsymbol{D}_{\tau_K}\right) \end{pmatrix}$$

is then a $(K+1)p^2$-variate normal distribution with mean value zero and covariance matrix as given in Miettinen et al. (2016), and the limiting distribution of

$$\sqrt{T}\text{vec}\left(\widehat{\boldsymbol{W}} - \boldsymbol{I}_p\right)$$

is a p^2-variate normal distribution with mean value zero and limiting covariance matrices as given in Miettinen et al. (2012, 2014, 2016). Asymptotic and final sample efficiency studies demonstrate that in most settings the symmetric SOBI estimate outperforms both the AMUSE estimate and the deflation-based SOBI estimate.

As mentioned earlier, the goal of SOBI is to incorporate the information of more lags as compared to AMUSE which is based on a single lag. SOBI does so by generalizing the idea of the diagonalization of two matrices to jointly diagonalizing more matrices. But there are other ways to include the information of more lags. For example, temporal decorrelation source separation or time-delays based separation (TDSEP, Ziehe & Müller, 1998) is a BSS algorithm which generalizes AMUSE in another direction. One version of the algorithm uses the average of several autocovariance matrices, that is,

$$\frac{1}{|\mathcal{T}|}\sum_{\tau\in\mathcal{T}}\text{Cov}_\tau\left(\boldsymbol{x}_t^{st}\right) \tag{8}$$

to find the unknown rotation matrix. Thus, such a TDSEP algorithm can be viewed as an extension of AMUSE, where $\text{Cov}_\tau\left(\boldsymbol{x}_t^{st}\right)$ for a lag τ is replaced by (8). It seems, however, that SOBI prevails in practical applications as compared to TDSEP.

Finally note that Belouchrani and Amin (1998) considered blind source separation in the frequency domain. Instead of joint diagonalizing covariance matrices, their time-frequency separation (TFS) algorithm jointly diagonalizes spatial time-frequency distribution matrices of the form

$$\boldsymbol{D}_{zz}(t,f) = \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi(m,l) \boldsymbol{z}_{t+m+l} \boldsymbol{z}'_{t+m-l} e^{j4\pi fl}, \tag{9}$$

where the kernel $\phi(m,l)$ characterizes the distribution, t refers to time and f to frequency (Cohen, 1995).

3 | GENERALIZATIONS AND OTHER VERSIONS OF SOBI

SOBI has proven to be a simple and powerful method for time series dimension reduction. In some settings its applicability is, however, limited, therefore several generalizations and modifications of the method have been proposed in the literature. The generalizations of SOBI try to answer some of the issues the regular SOBI has:

- There are no tools for selecting a lag set that yields best separation performance.
- Exact diagonalization of the covariance matrix may happen at the expense of poorer diagonalization of the autocovariance matrices.
- The method does not work well when the volatility cannot be considered as fixed throughout the time series.
- The mixing procedure is assumed to stay constant all the time.
- The latent components are assumed to be weakly stationary.
- The method can only be used for vector-valued time series, but not for images or tensor-valued time series.
- The method is not robust to outliers.
- The method does not give a way to directly assess which of the signals are important.
- The method is not applicable if there are more sources than observed signals.

In the following subsections, we introduce some extensions of AMUSE and SOBI which address the restrictions expressed above. Table 1 summarizes these extensions and describes assumptions needed for each method.

3.1 | Using asymptotics for gaining better separation

The signal separation performance of SOBI method depends heavily on the chosen lag set $\mathcal{T} = \{\tau_1, ..., \tau_K\}$. A default lag set in many applications has been $\{1, ..., 12\}$, as these lags are thought to be enough for capturing the correlation structure in data. However, this is just an arbitrary choice, and there are no theoretical foundations for choosing the set. Tang, Liu, and Sutherland (2005), for example, advocate for EEG that the default set is not sufficient.

Miettinen (2015) and Taskinen et al. (2016) have proposed methods that use asymptotic results to obtain more efficient signal separation matrix estimates. Note first that the sum of the limiting variances of the off-diagonal elements of $\sqrt{T}\text{vec}\left(\widehat{\boldsymbol{W}} - \boldsymbol{I}_p\right)$ provides a global measure of the variation of $\widehat{\boldsymbol{W}}$ and thus a tool for comparing asymptotic efficiencies (Ilmonen et al., 2010,b). Miettinen (2015) proposed a method, where the criterion function to be maximized is

$$\sum_{\tau \in \mathcal{T}} \sum_{i=1}^{p} G\left(\boldsymbol{u}'_i \text{Cov}_\tau\left(\boldsymbol{x}^{st}_t\right) \boldsymbol{u}_i\right),$$

where G is an even and continuously differentiable function, for which $G(0) = 0$, $g(x) = G'(x) \geq 0$ for $x > 0$ and $g(x) \leq 0$ for $x < 0$. The choice $G(x) = x^2$ gives the regular SOBI estimator. Also asymptotic properties of this estimate have been carefully investigated. Miettinen (2015) focused especially on aSOBI, a version of the symmetric SOBI method, that uses

$$\sum_{\tau \in \mathcal{T}} \sum_{i=1}^{p} \left|\boldsymbol{u}'_i \text{Cov}_\tau\left(\boldsymbol{x}^{st}_t\right) \boldsymbol{u}_i\right|^a, \tag{10}$$

TABLE 1 Overview of SOBI and its variants considered in this review and assumptions needed for each method

Method	Data assumptions	Reference(s)
AMUSE, SOBI, TDSEP	Uncorrelated weakly stationary processes where most information is in the second moments. Constant mixing matrix	Tong et al. (1990), Belouchrani et al. (1997), Ziehe and Müller (1998)
aSOBI, eSOBI	Sources are uncorrelated stationary $MA(\infty)$-processes. Constant mixing matrix	Miettinen (2015), Taskinen et al. (2016)
WASOBI	Uncorrelated weakly stationary processes where most information is in the second moments with some additional information on the sources available. Constant mixing matrix	Yeredor (2000)
vSOBI, FixNA, FixNA2	Independent stationary processes where most information is in the higher order moments, for example, stochastic volatility models	Matilainen et al. (2017), Shi et al. (2009), Hyvärinen (2001)
gSOBI	Convex linear combination of SOBI and vSOBI assuming independent stationary processes with information in second and/or higher order moments	Miettinen, Matilainen, et al. (2020)
TV-SOBI	Uncorrelated weakly stationary processes where most information is in the second moments. Mixing matrix changes over time	Yeredor (2003), Weisman and Yeredor (2006)
NSS-TD-JD	L independent p-variate time series with uncorrelated weakly stationary sources and all having the same mixing matrix or block-stationary p-variate time series where the sources have constant mean but variances and covariance functions of them are block-stationary	Choi and Cichocki (2000)
mdSOBI	Uncorrelated multidimensional data such as images which are weakly stationary	Theis et al. (2004)
TSOBI	Tensorial time series with uncorrelated weakly stationary sources	Virta and Nordhausen (2017b)
SOBIUM	Uncorrelated weakly stationary processes. The number of sources exceeds the number of observed components	Lathauwer and Castaing (2008)
SAM-SOBI, eSAM-SOBI	Robust variants of SOBI which make the additional assumption of symmetric sources	Theis et al. (2010), Ilmonen et al. (2015)
RmdSOBI	Robust variant of mdSOBI which makes the additional assumption of symmetric sources	Lietzén et al. (2017)

where $a \in (1, \infty)$, as a criterion function. A large value of a means that more emphasis is put on the matrices with large autocovariances. An adaptive estimator based on aSOBI is suggested to improve efficiency of signal separation. First the regular SOBI is used to obtain estimates for source components. Then the asymptotic results are used to find the value a that minimizes the sum of the limiting variances. Finally, the signal separation matrix is estimated using the chosen a (Miettinen, 2015).

In Taskinen et al. (2016), similar approach as in aSOBI is used, but instead of using different G-functions, the most efficient combination of lags is chosen. In efficient SOBI (eSOBI) one compiles a list of different lag combinations, and the lag set that leads to the lowest measure of variation is then chosen for the problem at hand. Notice that for computing limiting variances, both aSOBI and eSOBI assume that the signals are $MA(\infty)$ processes.

3.2 | Reformulating SOBI as weighted least squares problem

Due to the prewhitening step in SOBI, the covariance matrix is exactly diagonalized while the autocovariance matrices are approximately jointly diagonalized. This exact diagonalization of the covariance matrix may, however, happen at the expense of poorer diagonalization of the autocovariance matrices (Cardoso, 1994). The intuition is that the global maximal diagonality of autocovariance matrices are most likely achieved when the covariance matrix is not exactly

diagonal, but the diagonal covariance is mandatory for SOBI's joint diagonalization in (6). Also, when estimating the correlations, the errors are highly correlated. The use of a weighted least squares criterion can thus be beneficial when compared to the ordinary least squares optimization criterion that is used in classical SOBI.

Yeredor (2000) has suggested weights-adjusted SOBI (WASOBI) algorithm, where instead of the approximate joint diagonalization, the problem is reformulated as a weighted least squares problem. Observe that

$$\widehat{\boldsymbol{y}} = G_0(\boldsymbol{A})\mathrm{diag}(\boldsymbol{D}), \tag{11}$$

which is a stacked form over all $\tau \in \mathcal{T}$ of the following equation,

$$\widehat{\boldsymbol{y}}_\tau = G(\boldsymbol{A})\mathrm{diag}(\boldsymbol{D}_\tau), \tag{12}$$

where $\widehat{\boldsymbol{y}}_\tau = \mathrm{vec}\left(\widehat{\mathrm{Cov}}_\tau(\boldsymbol{x}_t^{st})\right)$, $G(\boldsymbol{A}) = \boldsymbol{A} \odot \boldsymbol{A}$ with $\odot$ denoting Khatri–Rao product, and $G_0(\boldsymbol{A}) = \boldsymbol{I}_p \otimes G(\boldsymbol{A})$. Since the symmetry property cannot be ensured in the vectorized least squared estimation, in practice $\widehat{\boldsymbol{y}}_\tau$ is more commonly assigned as $\widehat{\boldsymbol{y}}_\tau = \mathrm{svec}\left(\widehat{\mathrm{Cov}}_\tau(\boldsymbol{x}_t^{st})\right)$, which is achieved by stacking only the lower triangle (including diagonal positions) of $\frac{1}{2}\left(\widehat{\mathrm{Cov}}_\tau(\boldsymbol{x}_t^{st}) + \widehat{\mathrm{Cov}}'_\tau(\boldsymbol{x}_t^{st})\right)$. Meanwhile, the design matrix $\boldsymbol{H}$ containing 0, 1 and 1/2 is introduced for $G_0(\boldsymbol{A}) = \boldsymbol{I}_p \otimes (\boldsymbol{H}(\boldsymbol{A} \odot \boldsymbol{A}))$ to conform matrix multiplication (Tichavský, Doron, et al., 2006). Finally, weighted least squares theory suggests that using $\widehat{\boldsymbol{y}}$ as in Equation (11), the following criterion should be minimized

$$C_{WLS}(\boldsymbol{A}, \mathrm{diag}(\boldsymbol{D})) = (\widehat{\boldsymbol{y}} - G_0(\boldsymbol{A})\mathrm{diag}(\boldsymbol{D}))' \boldsymbol{Q} (\widehat{\boldsymbol{y}} - G_0(\boldsymbol{A})\mathrm{diag}(\boldsymbol{D})), \tag{13}$$

where the appropriate weight $\boldsymbol{Q}$ is calculated based on the correlations between the (i, j)th element of $\widehat{\mathrm{Cov}}_\tau(\boldsymbol{x}_t^{st})$ and the (k, l)th element of $\widehat{\mathrm{Cov}}_\tau(\boldsymbol{x}_t^{st})$. The weighted least square problem is solved by an iterative manner. For convenience, write $\boldsymbol{f}(\theta) = G_0(\boldsymbol{A})\mathrm{diag}(\boldsymbol{D})$, where θ is a vector containing unknown parameters in $\boldsymbol{A}$ and $\mathrm{diag}(\boldsymbol{D})$. The minimizing criteria (13) can then be written as $C_{WLS}(\theta) = (\widehat{\boldsymbol{y}} - \boldsymbol{f}(\theta))' \boldsymbol{Q} (\widehat{\boldsymbol{y}} - \boldsymbol{f}(\theta))$. The iteration step is thus defined by

$$\theta^{[i+1]} = \theta^{[i]} + \left(\boldsymbol{F}'_i \boldsymbol{Q} \boldsymbol{F}_i\right)^{-1} \boldsymbol{Q} \boldsymbol{F}_i \left(\widehat{\boldsymbol{y}} - \boldsymbol{f}\left(\theta^{[i]}\right)\right), \tag{14}$$

where $\boldsymbol{F}_i = \partial \boldsymbol{f}(\theta)/\partial\theta|_{\theta = \theta^{[i]}}$ (Tichavský, Doron, et al., 2006).

The computational complexity of WASOBI can be relaxed when the sources are assumed to follow a known time series model. Optimal weighting for moving average sources was presented in Yeredor (2000), whereas AR-WASOBI is a fast algorithm for autoregressive sources (Tichavský, Doron, et al., 2006; Tichavský & Yeredor, 2009). Tichavský, Koldovský, et al. (2006) have also proposed EFWS (EFica-WaSobi) and COMBI (COMBInation), which are both combinations of the efficient variant of fastICA (EFICA) (Koldovský et al., 2006) and WASOBI. Other methods, which do not require exact diagonalization of the covariance matrix are proposed in Pham (2001); Yeredor (2002), for example.

3.3 | Separating time series with volatility clustering

Regular SOBI methods do not work if the time series exhibit volatility clustering, that is, in cases where the volatility cannot be considered constant over time. Examples of such times series models include stochastic volatility (SV) models and generalized autoregressive conditional heteroskedasticity (GARCH) models (Bollerslev, 1986; Taylor, 1982). See also Matteson and Tsay (2011) for a recent review on GARCH models. For such models the autocovariances (3) equal to zero. However, zero autocovariances do not imply the absence of temporal dependence.

In Matilainen et al. (2017) the SOBI method is extended to the case of volatility clustering. Notice first that the maximization problem in SOBI (6) can alternatively be written as

$$\sum_{\tau \in \mathcal{T}} \sum_{i=1}^{p} \left(\mathrm{E}\left(\left(\boldsymbol{u}'_i \boldsymbol{x}_t^{st}\right) \left(\boldsymbol{u}'_i \boldsymbol{x}_{t+\tau}^{st}\right) \right) \right)^2. \tag{15}$$

A set of methods suitable for times series exhibiting volatility clustering is then obtained by using nonlinearity functions in (15). In a variant of SOBI (vSOBI) proposed by Matilainen et al. (2017), the maximization problem for an orthogonal matrix $\boldsymbol{U}$ is

$$\sum_{\tau \in \mathcal{T}} \sum_{i=1}^{p} \left(\mathrm{E}\left(G\left(\boldsymbol{u}_i' \boldsymbol{x}_{t+\tau}^{st}\right) G\left(\boldsymbol{u}_i' \boldsymbol{x}_t^{st}\right)\right) - \mathrm{E}\left(G\left(\boldsymbol{u}_i' \boldsymbol{x}_t^{st}\right)\right) \mathrm{E}\left(G\left(\boldsymbol{u}_i' \boldsymbol{x}_{t+\tau}^{st}\right)\right) \right)^2, \tag{16}$$

where G is any twice continuously differentiable function. Some commonly used choices are $G(x) = x^2$ and $G(x) = \log(\cosh(x))$. Notice that the vSOBI method needs stronger assumptions than SOBI, as the components of $\boldsymbol{z}$ are required to be mutually independent, not just uncorrelated. The methods FixNA (Shi et al., 2009) and FixNA2 (Hyvärinen, 2001) are closely related to vSOBI. For the statistical properties and comparisons on these methods, see Matilainen et al. (2017). Dynamic orthogonal analysis (DOC), as suggested in Matteson and Tsay (2011), is very similar to vSOBI. Generalized fourth order blind identification (gFOBI) and generalized joint approximation of eigenmatrices (gJADE), as suggested in Matilainen et al. (2015), are designed for time series considered in DOC.

SOBI is thus applicable when variances of the source time series can be regarded as constant over time and vSOBI works for time series exhibiting volatility clustering. Miettinen, Matilainen, et al. (2020) combined SOBI and vSOBI with $G(z) = z^2$ and proposed a generalized SOBI (gSOBI) method, which maximizes

$$w \sum_{\tau \in \mathcal{T}_1} \sum_{i=1}^{p} \left(\mathrm{E}\left(\boldsymbol{u}_i' \boldsymbol{x}_t^{st} \boldsymbol{u}_i' \boldsymbol{x}_{t+\tau}^{st}\right) \right)^2 + (1-w) \sum_{\tau \in \mathcal{T}_2} \sum_{i=1}^{p} \left(\mathrm{E}\left(\left(\boldsymbol{u}_i' \boldsymbol{x}_t^{st}\right)^2 \left(\boldsymbol{u}_i' \boldsymbol{x}_{t+\tau}^{st}\right)^2 - 1 \right) \right)^2, \tag{17}$$

under the constraint $\boldsymbol{U}\boldsymbol{U}' = \boldsymbol{I}_p$. Here $\mathcal{T}_1$ and $\mathcal{T}_2$ are the sets of lags for the linear and quadratic parts, respectively, and $0 \le w \le 1$ is a parameter which gives the weight for the two parts.

In Miettinen, Matilainen, et al. (2020) the limiting distribution of the signal separation matrix is derived under general conditions, and the asymptotic variances are derived in the case of ARMA-GARCH model. Asymptotic and finite sample efficiency studies are used to compare different choices of a weight coefficient. As it is often of interest to identify all those components which exhibit stochastic volatility features, a method for ordering the time series according to their "volatility" is proposed along with the test statistics for testing linear autocorrelation and volatility clustering. Visual analytic tools which might help to choose w, $\mathcal{T}_1$ and $\mathcal{T}_2$ are suggested in Piccolotto et al. (2020).

3.4 | Time-dependent mixing matrix

When using SOBI, we assume a SOS model (1), where the mixing matrix $\boldsymbol{A}$ is assumed to be constant through time. However, we can also allow the mixing matrix to change over time. This enables the use of SOBI type methods also in cases where the observed signals are nonstationary, as the sources are still assumed to be stationary. Yeredor (2003) proposed time-variable SOBI (TV-SOBI) for such situations. The time-variable BSS (TV-BSS) model is written as

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{A}_t \boldsymbol{z}_t, \quad t \in \mathbb{Z},$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is again a location vector and $\boldsymbol{A}_t \in \mathbb{R}^{p \times p}$ is the mixing matrix at time t. The assumptions on $\boldsymbol{z}$ are the same as for SOBI. Yeredor (2003) concentrated on a case where the variation is assumed to be linear in time, that is,

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \left(\boldsymbol{I}_p + t\boldsymbol{R}\right)\boldsymbol{A}_0 \boldsymbol{z}_t, \quad t \in \mathbb{Z}, \tag{18}$$

where $\boldsymbol{A}_0 \in \mathbb{R}^{p \times p}$ is a full-rank mixing matrix at time zero and $\boldsymbol{R} \in \mathbb{R}^{p \times p}$ is a "relative rate" matrix, which generates the time effect. In general, the values in matrix $\boldsymbol{R}$ are assumed to be very small, that is, the mixing matrix changes only a little at a time. Weisman and Yeredor (2006) considered models for periodical time variation, that is,

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \big(I_p + \boldsymbol{R}_c \cos(\omega t) + \boldsymbol{R}_s \sin(\omega t)\big)\boldsymbol{A}_m \boldsymbol{z}_t, \quad t \in \mathbb{Z}, \tag{19}$$

where ω is an angular frequency, $\boldsymbol{A}_m \in \mathbb{R}^{p \times p}$ is a full-rank matrix representing the "mean" mixing matrix, and matrices $\boldsymbol{R}_c, \boldsymbol{R}_s \in \mathbb{R}^{p \times p}$ describe "relative rates" of harmonics generating periodical variation.

By definition, time variation increases complexity of the signal separation matrix estimation due to the increased number of parameters in models. The developed estimation methods involve standard SOBI procedures on top of applicable matrix algebra (Yeredor, 2003). In return, these approaches can outperform SOBI especially if the time series length T is very large and/or observation indicates clear periodic pattern. For more details on time-varying mixtures in SOS, see, for example, Chapter 7 of Comon and Jutten (2010) and the references therein.

3.5 | Multisubject SOS and block-stationary SOS

So far it is assumed that the latent sources are stationary. This is a convenient but strong assumption, hence there are approaches which try to relax it. One motivation for such approaches are for example multisubject studies. Assume that we observe for L subjects p-variate centered time series $\boldsymbol{x}_t^i$ of length T_i, $i = 1, ..., L$. A popular model for example in neurosciences states that for all L subjects a SOS model holds with the additional assumption that the mixing matrix $\boldsymbol{A}$ is the same for all subjects. The latent components $\boldsymbol{z}_t^i$ might, however, follow different processes. This is often referred to as group ICA, see, for example, Cong et al. (2013) and references therein.

One way to analyze multisubject data is to concatenate the time series into $\boldsymbol{x}_{t'} = \big(\boldsymbol{x}_t^1, ..., \boldsymbol{x}_t^L\big)$ with $t' = 1, ..., \sum_{l=1}^{L} T_l$, and to apply a BSS method to $\boldsymbol{x}_{t'}$. Naturally $\boldsymbol{x}_{t'}$ is not anymore stationary, but rather block-stationary, where the observations coming from the same subject form blocks. This information is exploited in a block-stationary variant of SOBI known as nonstationary source separation with temporal delayed correlation matrices using joint diagonalization (NSS-TD-JD, Choi and Cichocki (2000)). For that purpose, define the block autocovariance matrix as

$$\widehat{\mathrm{Cov}}_\tau^l(\boldsymbol{x}_{t'}) = \frac{1}{T_l - \tau} \sum_{t' \in t_l} \boldsymbol{x}_{t'+\tau} \boldsymbol{x}_{t'}',$$

where t_l indicates the index in t' which identifies block l, $l = 1, ..., L$. In NSS-TD-JD the lag set $\mathcal{T} = \{0, \tau_1, ..., \tau_K\}$ contains usually the zero lag, and the unmixing matrix is obtained using the following steps. Write the standardized times series as $\boldsymbol{x}_{t'}^{st} = \widehat{\mathrm{Cov}}(\boldsymbol{x}_{t'})^{-1/2}(\boldsymbol{x}_{t'} - \bar{\boldsymbol{x}}_{t'})$ and find the orthogonal matrix $\mathbf{U}$ maximizing

$$\sum_{l \in L} \sum_{\tau \in \mathcal{T}} \left\| \boldsymbol{U} \widehat{\mathrm{Cov}}_\tau^l(\boldsymbol{x}_{t'}) \boldsymbol{U}' \right\|^2.$$

The signal separating matrix is then $\boldsymbol{W} = \boldsymbol{U} \widehat{\mathrm{Cov}}(\boldsymbol{x}_{t'})^{-1/2}$. Notice that there are in total $L(K+1)$ matrices to be jointly diagonalized. If $L = 1$, the method just reduces to SOBI.

In a group ICA context the number of blocks and their ranges are naturally given. In nonstationary source separation (NSS) it is assumed that $\boldsymbol{z}_{t'}$ is first-order stationary and that $\mathrm{Cov}_\tau(\boldsymbol{z}_{t'}) = \boldsymbol{D}_{\tau, t'}$ is a diagonal matrix dependent on τ and t'. NSS-TD-JD was originally suggested in this setting and requires dividing the process in L blocks. Hence L and K are parameters which have impact on the performance of the method.
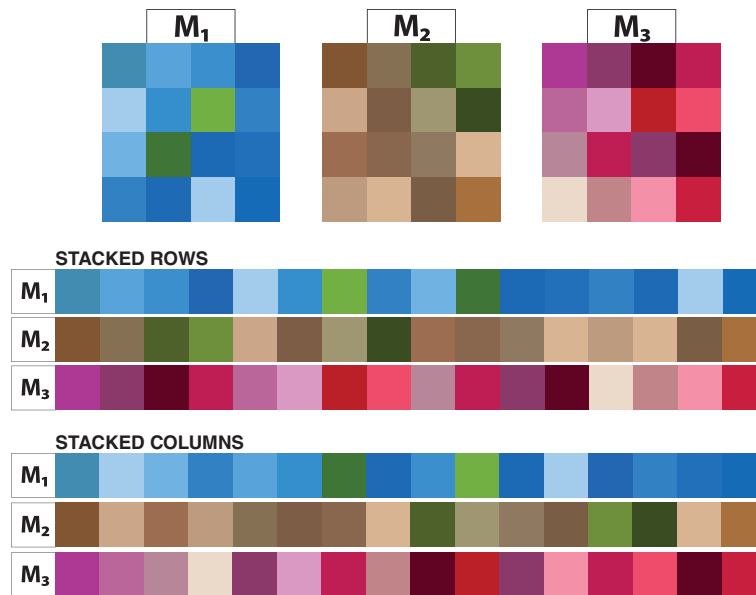
FIGURE 1 Schematic representation of the difference when stacking rows and columns respectively

3.6 | Generalization to *m*-dimensional data

Demixing p images is a frequent application area of BSS. It is obvious that in image data there is serial dependence present within columns and rows of the images although one usually assumes more general local spatial dependencies. A common, though naive, way to approach this problem is to vectorize the images and use these vectors as input in standard BSS methods such as AMUSE or SOBI. The separation result then depends of course on whether the rows or columns have been stacked as for example visualized in Figure 1. Also, an image may actually be represented not as a 2-dimensional object, but for example as a 3-dimensional object as, besides width and height, also the intensity of the colors is measured. Thus, images can be thought of as p m-dimensional objects of the same size or as p-dimensional vectors with a m-dimensional index. Using the later representation Theis et al. (2004) suggested the concept of multidimensional autocovariance, which is defined for a centered p-variate vector x_t with an m-dimensional index $t = (t_1, ..., t_m)$ as

$$\text{Cov}_\tau(x_t) = E(x_t x'_{t+\tau}) = E\left(x_{t_1,...,t_m} x'_{t_1+\tau_1,...,t_m+\tau_m}\right),$$

where $\tau = (\tau_1, ..., \tau_m)$ specifies the lags for the corresponding dimensions.

Replacing regular symmetrized autocovariance matrices (7) with multidimensional autocovariance matrices and their symmetrized versions in the regular AMUSE and SOBI yields what Theis et al. (2004) refer to as multidimensional AMUSE (mdAMUSE) and multidimensional SOBI (mdSOBI), respectively, which they then apply for example to fMRI data.

In mdAMUSE and mdSOBI we assume that p multidimensional objects are observed once. However, one can also observe a series of multidimension objects, so-called tensorial time series. In tensorial time series at each time point t a m-dimensional tensor $\mathbb{X}_t \in \mathbb{R}^{p_1 \times \cdots \times p_m}$ is observed. Videos serve as examples of typical tensorial time series, where for each frame a tensor is observed, see Figure 2 for a visualization. In case of a black and white video $m = 2$ and in case of a colored video $m = 3$ as indicated above. Tensorial AMUSE (TAMUSE) and tensorial SOBI (TSOBI) were introduced in Virta and Nordhausen (2017b) and are defined for any positive integer m, where with $m = 1$ the standard methods for vector-valued data are obtained. Consider next the case $m = 2$, that is, at each time point a matrix X_t is observed.

The corresponding model for tensorial, that is, matrix-variate SOS is
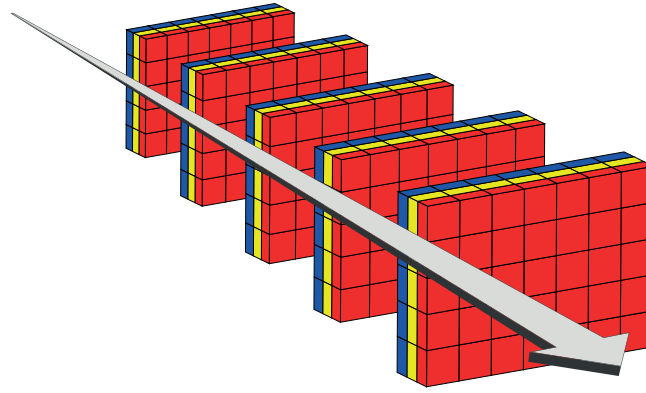
$$X_t = A_L Z_t A'_R + \mu,$$

FIGURE 2 Representation of a time series of mode three tensors

where $\boldsymbol{\mu} \in \mathbb{R}^{p_1 \times p_2}$ specifies the location, $\boldsymbol{A}_L \in \mathbb{R}^{p_1 \times p_1}$ and $\boldsymbol{A}_R \in \mathbb{R}^{p_2 \times p_2}$ are the nonsingular left and right mixing matrices describing the induced column and row dependencies on the latent process $\boldsymbol{Z}_t \in \mathbb{R}^{p_1 \times p_2}$ which fulfills the assumptions

$$E(\text{vec}(\boldsymbol{Z}_t)) = \boldsymbol{0}, \quad \text{Cov}(\text{vec}(\boldsymbol{Z}_t)) = \boldsymbol{I}_{p_1 p_2},$$
$$E(\boldsymbol{Z}_t \boldsymbol{Z}'_{t+\tau}) = \boldsymbol{D}^L_\tau \quad \text{and} \quad E(\boldsymbol{Z}'_t \boldsymbol{Z}_{t+\tau}) = \boldsymbol{D}^R_\tau,$$

where $\boldsymbol{D}^L_\tau$ and $\boldsymbol{D}^R_\tau$ are diagonal for all $\tau = 1, 2, \ldots$. Thus, the assumptions are a straightforward generalizations of the assumptions (2) and (3) listed in the vector-valued case.

Virta and Nordhausen (2017b) showed that the unmixing can be done mode-wise, where the left and right covariance matrices needed for whitening are

$$\text{Cov}_L(\boldsymbol{X}_t) = \frac{1}{p_2} E((\boldsymbol{X}_t - \bar{\boldsymbol{X}}_t)(\boldsymbol{X}_t - \bar{\boldsymbol{X}}_t)')$$

and

$$\text{Cov}_R(\boldsymbol{X}_t) = \frac{1}{p_1} E((\boldsymbol{X}_t - \bar{\boldsymbol{X}}_t)'(\boldsymbol{X}_t - \bar{\boldsymbol{X}}_t)),$$

yielding the standardized matrix-valued time series

$$\boldsymbol{X}_t^{st} = \text{Cov}_L(\boldsymbol{X}_t)^{-1/2}(\boldsymbol{X}_t - \bar{\boldsymbol{X}}_t)\text{Cov}_R(\boldsymbol{X}_t)^{-1/2}.$$

Accordingly, the mode-wise autocovariance matrices with lag τ are defined as

$$\text{Cov}_{L,\tau}(\boldsymbol{X}_t) = \frac{1}{p_2} E((\boldsymbol{X}_t - \bar{\boldsymbol{X}}_t)(\boldsymbol{X}_{t+\tau} - \bar{\boldsymbol{X}}_t)')$$

and

$$\text{Cov}_{R,\tau}(\boldsymbol{X}_t) = \frac{1}{p_1} E((\boldsymbol{X}_t - \bar{\boldsymbol{X}}_t)'(\boldsymbol{X}_{t+\tau} - \bar{\boldsymbol{X}}_t)),$$

which can then be used to find the left and right rotation matrices $\boldsymbol{U}_L$ and $\boldsymbol{U}_R$ as the maximizers of

$$\sum_{\tau\in\mathcal{T}_L}\|\operatorname{diag}\left(\boldsymbol{U}\operatorname{Cov}_{L,\tau}\left(\boldsymbol{X}_t^{st}\right)\boldsymbol{U}'\right)\|^2 \quad \text{and} \quad \sum_{\tau\in\mathcal{T}_R}\|\operatorname{diag}\left(\boldsymbol{U}\operatorname{Cov}_{R,\tau}\left(\boldsymbol{X}_t^{st}\right)\boldsymbol{U}'\right)\|^2,$$

respectively, yielding $\boldsymbol{W}_L=\boldsymbol{U}_L\operatorname{Cov}_L(\boldsymbol{X}_t)^{-1/2}$ and $\boldsymbol{W}_R=\boldsymbol{U}_R\operatorname{Cov}_R(\boldsymbol{X}_t)^{-1/2}$. Depending on the number of lags used in $\mathcal{T}_L$ and $\mathcal{T}_R$ one distinguishes between tensorial AMUSE or tensorial SOBI.

The extension to higher order tensors follows the same ideas as in the matrix case and the key is still to unmix mode-wise. Actually, when considering the mth mode of the tensor $\mathbb{X}\in\mathbb{R}^{p_1\times\cdots\times p_r}$, the tensor is flattened or matricized to its m-flattening matrix $\boldsymbol{X}^{(m)}\in\mathbb{R}^{p_m\times\rho_m}$ horizontally by stacking all its m-mode vectors where accordingly $\rho_m=\prod_{i\neq m}^r p_i$. To simplify in the following the notation we define the multiplication operator $\odot_m$ over a tensor $\mathbb{X}\in\mathbb{R}^{p_1\times\cdots\times p_r}$ and a matrix $\boldsymbol{A}\in\mathbb{R}^{p_m\times p_m}$ in an element-wise manner as,

$$(\mathbb{X}\odot_m\boldsymbol{A})_{i_1\cdots i_r}=\sum_{j_m}x_{i_1\cdots i_{m-1}j_m i_{m+1}\cdots i_r}a_{i_m j_m},$$

where $m=1,2,\ldots,r$ and the product $(\mathbb{X}\odot_m\boldsymbol{A})$ has the same dimensionality as $\mathbb{X}$. The tensor blind source separation model for a tensor-valued time series $\mathbb{X}_t$ can then be written as

$$\mathbb{X}_t=\mathbb{Z}_t\odot_1\boldsymbol{A}_1\cdots\odot_r\boldsymbol{A}_r,$$

where $\boldsymbol{A}_m\in\mathbb{R}^{p_m\times p_m}$ are m-mode mixing matrices, and the unobserved tensor-valued $\mathbb{Z}_t$ fulfills the assumptions

$$\mathrm{E}\left[\operatorname{vec}(\mathbb{Z}_t)\right]=\boldsymbol{0},\quad \operatorname{Cov}\left[\operatorname{vec}(\mathbb{Z}_t)\right]=\mathbf{I}\quad\text{and}\quad \mathrm{E}\left[\mathbb{Z}_t^{(m)}\mathbb{Z}_{t+\tau}^{(m)'}\right]=\mathrm{E}\left[\mathbb{Z}_{t+\tau}^{(m)}\mathbb{Z}_t^{(m)'}\right]=\boldsymbol{D}_\tau^m,$$

which means that again, as in the vector and matrix case, the matrices $\boldsymbol{D}_\tau^m$ are required to be diagonal for all modes $m=1,\ldots,r$ and all lags τ. Assume for simplicity that in the following $\mathbb{X}_t$ is centered, then the m-mode (auto) covariance matrix is defined as

$$\operatorname{Cov}_\tau^m(\mathbb{X}_t)=\frac{1}{\rho_m}\mathrm{E}\left(\boldsymbol{X}_t^{(m)}\left(\boldsymbol{X}_{t+\tau}^{(m)}\right)'\right)$$

and accordingly the tensor is standardized from each mode

$$\mathbb{X}_t^{st}=\mathbb{X}_t\odot_1\left(\operatorname{Cov}_0^1(\mathbb{X}_t)\right)^{-1/2}\cdots\odot_r\left(\operatorname{Cov}_0^r(\mathbb{X}_t)\right)^{-1/2}.$$

Virta and Nordhausen (2017b) then show that subsequently the mth mode unmixing matrix is $\boldsymbol{W}^m(\mathbb{X}_t)=\boldsymbol{U}_m'\left(\operatorname{Cov}_0^m(\mathbb{X}_t)\right)^{-1/2}$ where $\boldsymbol{U}_m$ is the orthogonal matrix which maximizes

$$\sum_{\tau\in\mathcal{T}^m}\|\operatorname{diag}(\boldsymbol{U}\operatorname{Cov}_\tau^m(\mathbb{X}_t^{st})\boldsymbol{U}'\|^2.$$

Extensions to tensorial nonstationary time series following the same ideas are discussed in Virta and Nordhausen (2017a).

3.7 | Robust SOBI methods

The sample statistics used in SOBI and most of the related methods, that is, the sample mean, the sample covariance matrix and the sample cross-autocovariance matrix are highly sensitive to outlying observations and inefficient under heavy-tailed distributions. To overcome this issue some robust counterparts for classical SOBI have been derived in the literature. Belouchrani and Cichocki (2000) proposed a robust standardization step for the SOBI procedure by applying in standardization a linear combination of selected autocovariance matrices, that is $\left(\sum_{\tau\in\mathcal{T}}\boldsymbol{\alpha}_\tau\operatorname{Cov}_\tau(\boldsymbol{x}_t)\right)^{-1/2}$, where the

parameters $\boldsymbol{\alpha}_\tau$ are determined in an iterative method that ensure the positive-definite property. The existence of $\boldsymbol{\alpha}_\tau$ is proved by Tong et al. (1992). Notice, however, that for deriving genuine robust unmixing matrix estimates, we need robust estimates of location, covariance matrix and autocovariance matrix which measure the same quantity as their nonrobust counterparts and satisfy assumptions (2) and (3), that is, the covariance matrices are diagonal under the SOS model. This usually requires assuming symmetric components Nordhausen and Tyler (2015).

Theis et al. (2010) proposed a sign autocovariance SOBI (SAM-SOBI) algorithm, where data are first centered as

$$\boldsymbol{x}_t^{st} = \mathrm{SCov}(\boldsymbol{x}_t)^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu}_S), \tag{20}$$

where $\boldsymbol{\mu}_S$ is the spatial median (see, e.g., Haldane, 1948) and $\mathrm{SCov}(\boldsymbol{x}) = E(\|\boldsymbol{x}\|^{-2}\boldsymbol{x}\boldsymbol{x}')$, where $\|\boldsymbol{x}\| = (\boldsymbol{x}'\boldsymbol{x})^{1/2}$, is the spatial sign covariance matrix (Visuri et al., 2000). The spatial sign autocovariance matrices

$$\mathrm{SCov}_\tau = E\left(\frac{\boldsymbol{x}_t}{\|\boldsymbol{x}_t\|}\frac{\boldsymbol{x}'_{t+\tau}}{\|\boldsymbol{x}_{t+\tau}\|}\right)$$

are then calculated for all lags $\tau \in \mathcal{T}$ and symmetrized as in (7). Finally, a joint diagonalization of spatial sign autocovariance matrices is used to obtain an estimate for the signal separation matrix. The proposed method is highly robust. However, the obtained signal separation matrix estimate is not affine equivariant, as the spatial median and the spatial sign covariance matrix are only orthogonally equivariant (Visuri et al., 2000).

Ilmonen et al. (2015) proposed an eSAM-SOBI method, which is an affine equivariant version of SAM-SOBI. The spatial sign autocovariance matrices are still used in joint diagonalization, but the sample mean and the sample covariance matrix in (20) are replaced by the Hettmansperger–Randles estimates of location and scatter (Hettmansperger & Randles, 2002). These estimates are not only robust, but also affine equivariant. Simulation studies demonstrate that eSAM-SOBI outperforms both SOBI and SAM-SOBI under contaminated data. In Lietzén et al. (2017) RmdSOBI, a robust version of mdSOBI Theis et al. (2004), is given. Here whitening of the series is done using the robust Hettmansperger–Randles estimates of location and scatter, and multidimensional autocovariance matrices are replaced with the multidimensional spatial sign autocovariance matrices. Similarly, robust nonstationary blind source separation using spatial signs is discussed in Nordhausen (2014).

3.8 | Choosing the dimension of the signal subspace

Most users of BSS methods assume that actually only a few of the latent components are of interest and the others are noise. For the differentiation of noise and signal "noisy" BSS models are needed. The two widely used models are the following. The external noise model (ENM) is given as

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{z}_t + \boldsymbol{\varepsilon}_t,$$

where now $\boldsymbol{A}$ is a $p \times q$ matrix and the p-variate noise ε_t is independent from the q-variate signal $\boldsymbol{z}_t$. Notice that in the econometric literature such model is known as dynamic factor model (see, e.g., Forni et al., 2000; Bai & Ng, 2002) which is, however, in contrast to the SOS approach usually considered in the case where p and T are increasing. An alternative noise model is the internal noise model (INM) given as

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{z}_t = \boldsymbol{\mu} + \boldsymbol{A}_s\boldsymbol{z}_{s,t} + \boldsymbol{A}_n\boldsymbol{z}_{n,t},$$

which basically says that the p-variate latent component $\boldsymbol{z}_t = (\boldsymbol{z}'_{s,t}\boldsymbol{z}'_{n,t})'$ consists of a q-variate signal $\boldsymbol{z}_{s,t}$ which is independent form the $(p - q)$-variate noise $\boldsymbol{z}_{n,t}$. Here the mixing matrix has the corresponding partition into $\boldsymbol{A} = (\boldsymbol{A}_s\boldsymbol{A}_n)$.

In both models q is unknown and the goal is to estimate q together with the latent signals, for which in the context of this article the classical SOBI assumptions from Section 2 are made. In the ENM the noise is usually considered to be Gaussian white noise, that is, $\boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_p)$ while in the INM also alternatives like white noise with a $(p - q)$-variate spherical distribution or $p - q$ independent components without second-order serial dependence were considered.

The main difference between INM and ENM is that consistent unmixing matrices for the signals can always be found, but in the ENM the recovered signals will always be contaminated by noise, which is not a problem in the INM.

3.8.1 | Estimation of q in ENM

The main idea to obtain q in the ENM is based on the eigenvector-eigenvalue decomposition of Cov $(x_t) = UDU'$, where U is an orthogonal matrix and D is a diagonal matrix with eigenvalues $d_1, ..., d_p$ ordered in decreasing order. Then, under ENM, these eigenvalues should be $d_1 \geq \cdots \geq d_q > \sigma^2 = \cdots = \sigma^2$, that is, the $p - q$ smallest eigenvalues should be equal to the noise variance. The decision about the number of informative components (q) is so far done mainly graphically using for example a scree plot. The estimate for the noise variance is then the average of the $p - q$ smallest eigenvalues. Inference relating the eigenvalues makes either iid assumptions on the signal or requires the signals to be Gaussian in which case information theoretic criteria are available. For details see for example Wax and Kailath (1985); Zhao et al. (1986); Virta and Nordhausen (2019). In the dynamic factor model framework inference on q is studied in settings where both p and T are growing, that is, $p \rightarrow \infty$ and $T \rightarrow \infty$ (see, e.g., Forni et al., 2000; Bai & Ng, 2002; Onatski, 2009).

3.8.2 | Estimation of q in INM

In the INM all components which do not have second-order serial dependence are considered as noise. For the purpose of this review we assume the most common case, that is, the noise is Gaussian white noise. Therefore, the main approach is to consider autocovariance matrices which should then have q nonzero eigenvalues. Usually the decision is then approached using successive tests for the null hypothesis

$$H_{0k} : z_t \text{ contains a } (p-k)-\text{variate subvector of white noise.}$$

Again, as it is risky to base the decision here on one lag τ only, that is, on AMUSE alone, pooling information is again a safer choice. The pooled "SOBI" eigenvalues are defined as

$$\sum_{\tau \in \mathcal{T}} \text{diag}\left(\widehat{U} \widehat{\text{Cov}}_\tau^S (x_t^{st}) \widehat{U}' \right)$$

and are ordered in decreasing order according to their squared values as the closest one to zero form the noise. Thus $\widehat{U}$ has the partitioning $\left(\widehat{U}_s' \widehat{U}_n' \right)'$ identifying the signal and noise spaces.

The test statistic suggested in Matilainen et al. (2018); Virta and Nordhausen (2021) for this purpose is

$$t_k = \frac{1}{|\mathcal{T}| (p-k)^2} \sum_{\tau \in \mathcal{T}} \|D_{\tau,k}\|^2,$$

where $D_{\tau,k} = \widehat{U}_n \widehat{\text{Cov}}_\tau^S (x_t^{st}) \widehat{U}_n'$. Matilainen et al. (2018) suggested bootstrapping strategies to get the distribution of t_k under H_{0k}, and Virta and Nordhausen (2021) showed that under quite broad assumptions for a time series of length T, $T |\mathcal{T}| (p-q)^2 t_q$ converges to a $\chi^2_{|\mathcal{T}|(p-q)(p-q-1)/2}$ distribution under H_{0q}. Virta and Nordhausen (2021) also showed that successive testing can be used to get a consistent estimate for q.

Using stationary bootstrapping strategies for the pseudoeigenvalues of SOBI Nordhausen and Virta (2018) provided a ladle estimator for q. However, Virta and Nordhausen (2021) argued that successive application of the asymptotic test with a moderate set of lags is the preferable way to estimate q.

3.9 | More sources than observed signals

In our second-order separation model (1), the mixing matrix A is assumed to be a $p \times p$ square matrix, which implies the equality of the number of sources and observations. However, as described by Belouchrani et al. (1997), the SOBI method can still be applied if the mixing matrix A is a $p \times q$ matrix with $p > q$, that is, the dimension of sources is lower than the dimension of observations. If $p < q$, the SOBI method is not applicable since the diagonality of a nonsquare

matrix is not well-defined and thus joint diagonalization is no longer applicable. Lathauwer and Castaing (2008) has extended SOBI to a case where the number of unobserved signals is larger than the number of observed signals. This method is called as second-order blind identification of underdetermined mixtures (SOBIUM) and it proceeds as follows.

In SOBIUM autocovariance matrices are first calculated and stacked into a third-order tensor $\mathbb{C} \in \mathbb{R}^{p \times q \times k}$, where k is the number of lags selected. Then, SOBIUM performs parallel factor analysis (PARAFAC) (Harshman & Lundy, 1994) for $\mathbb{C}$, that is, it finds vectors $\boldsymbol{a}_1, ..., \boldsymbol{a}_q, \boldsymbol{d}_1, ..., \boldsymbol{d}_q$ that satisfy

$$\mathbb{C} = \sum_{i=1}^{q} \boldsymbol{a}_i \circ \boldsymbol{a}_i' \circ \boldsymbol{d}_i, \tag{21}$$

where $\boldsymbol{a}_i$ represents each column of mixing matrix $\boldsymbol{A}$, $\boldsymbol{d}_i$ is the ith column of the stacked $\boldsymbol{D}_\tau$ as defined in (3), and $\circ$ stands for element-wise product (Hardmard product). The PARAFAC is usually performed using iterative algorithm such as alternating least squares (ALS) algorithm that minimizes the cost function

$$f(\boldsymbol{a}_i, \boldsymbol{d}_i) = \| \mathbb{C} - \sum_{i=1}^{q} \boldsymbol{a}_i \circ \boldsymbol{a}_i' \circ \boldsymbol{d}_i \|^2 \tag{22}$$

(Kroonenberg & De Leeuw, 1980). Enhanced line search (ELS) proposed by Rajih et al. (2008) provides improved computational efficiency when compared to PARAFAC.

The SOBIUM algorithm is shown to be identifiable when $p \leq q \leq 2p - 2$ and the number of selected lags $k \geq q$. There also exists an enhanced version of SOBIUM that is specifically designed for $k \geq q$ (Lathauwer & Castaing, 2008).

4 | AN APPLICATION OF SOBI AND RELATED METHODS

As mentioned earlier SOBI and related methods are used in many applied fields. In the following we consider electrocardiography (ECG) recordings of a pregnant women. The goal of the analysis is to separate the heartbeats of the fetus. In an ECG recording sensors are placed on the skin of the mother measuring the electrical potential generated by the heart muscles. Figure 3 gives a schematic representation of a recording with two sensors. Here the two sources illustrate heartbeats of the mother and her fetus. The fetal heart rate is expected to be higher than that of the mother. The two sensors measure mixtures of the two heartbeats, and the weights can be seen as a measure of proximity to the
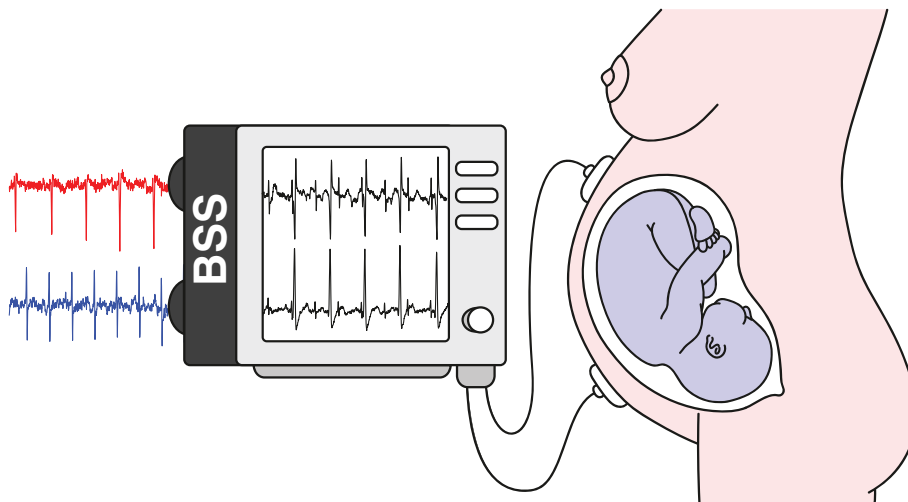

FIGURE 3 Schematic representation of an electrocardiography recording with two sensors for a pregnant women. BSS, blind source separation
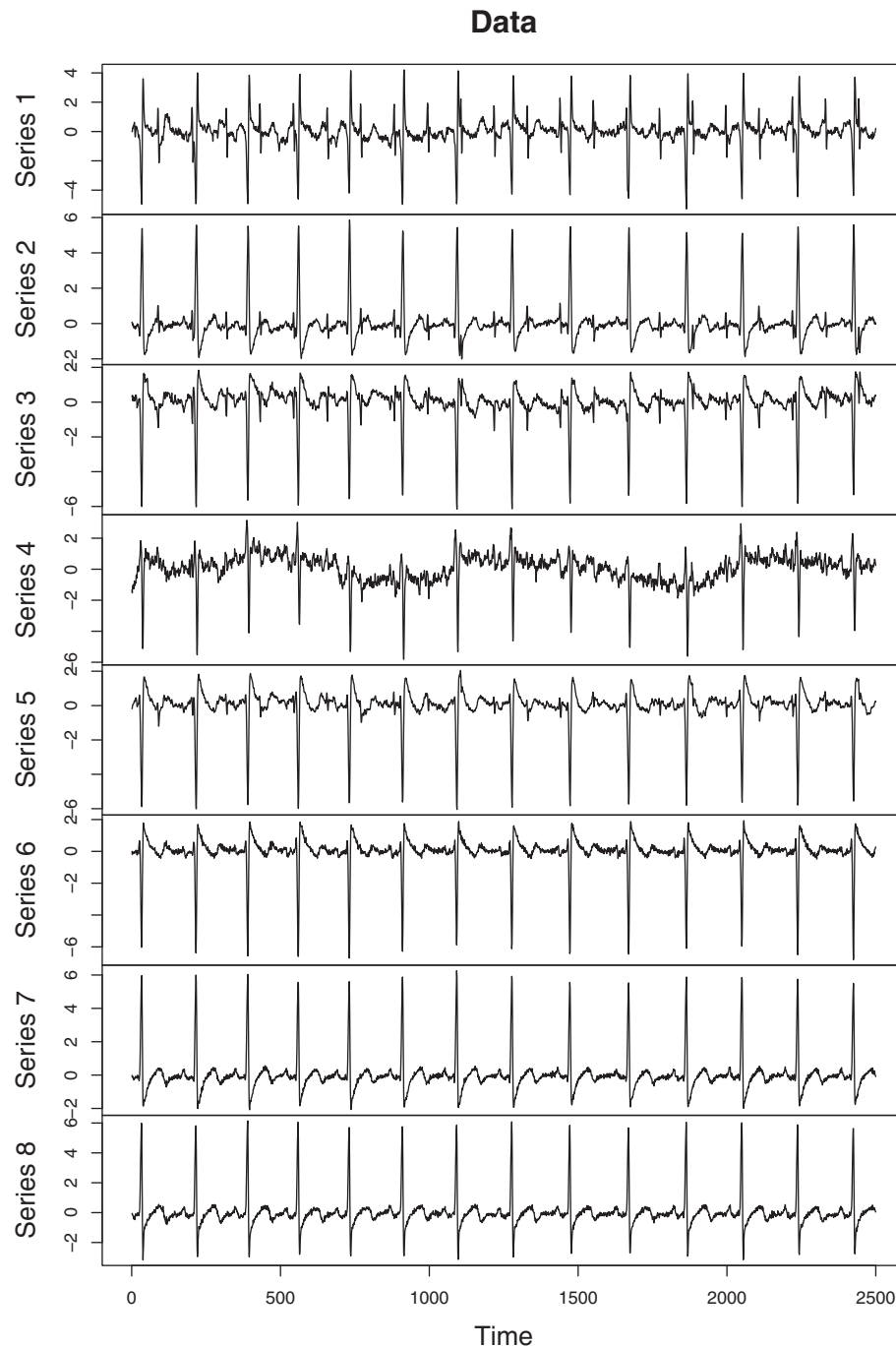
Data

FIGURE 4   Electrocardiography recordings of a pregnant women using eight sensors

corresponding heart. The goal of SOS is to separate the heartbeats of the fetus from those of the mother's based on the observed mixtures. As also other muscles in the body create electrical potential, which can be considered as artifacts, more than two sensors will in practice be used in actual ECG recordings. De Lathauwer et al. (1995) consider a recording that uses eight sensors, five in the stomach area and three in the chest area. The data are available for example in the supplement of Miettinen et al. (2017). The observed ECG measurements are given in Figure 4. Figure 5 depicts latent sources based on SOBI with $\mathcal{T} = \{1, \dots, 12\}$. As seen in the figure, one clear heartbeat pattern is seen, but not two. Thus, we explore the results based on gSOBI using $w = 0.9$, $\mathcal{T}_1 = \{1, \dots, 12\}$ and $\mathcal{T}_2 = \{1, 2, 3\}$. Estimated sources based on gSOBI are visualized in Figure 6. Series 3 seems to correspond to the heartbeat of the fetus being twice as fast as the mother's heartbeat in Series 2. This example illustrates the case where also the higher moments include crucial information on the time series.
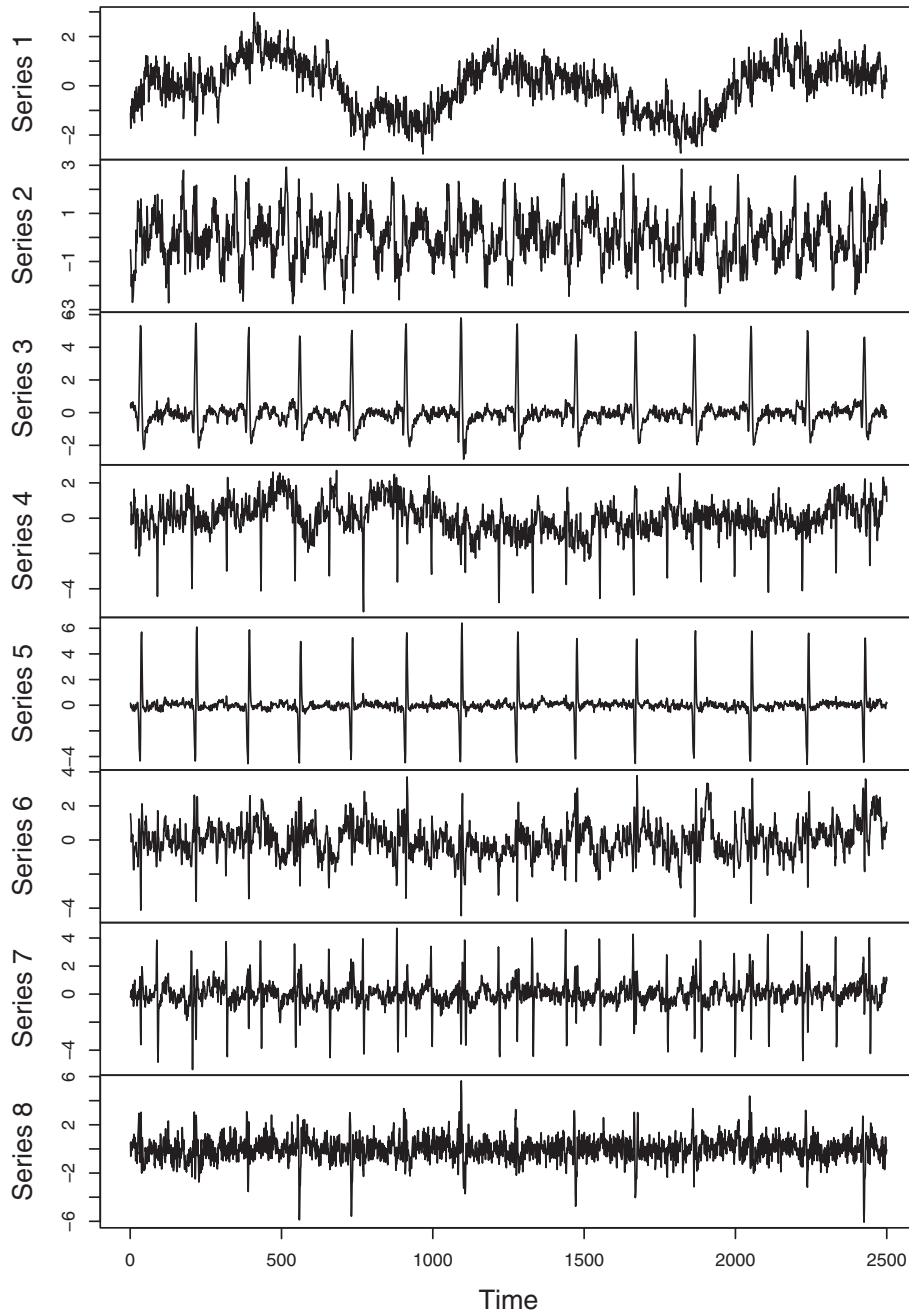
SOBI

FIGURE 5 Electrocardiography latent source components using second-order blind identification (SOBI) with $\mathcal{T} = \{1, ..., 12\}$

5 | ALGORITHMS AND IMPLEMENTATIONS OF SOBI AND RELATED METHODS

The estimation of SOBI unmixing matrix is simple when data are moderate in size. However, as typical applications, such as those encountered in neurosciences, usually generate huge datasets, computational issues have got more attention in recent years. The impact of the used (approximate) joint diagonalization algorithm on the estimation has been considered in Illner et al. (2015); Kalogiannis et al. (2017). Possibilities for an online variant of the SOBI algorithm, where the observations arrive successively, have been considered in Zhang et al. (2015). Parallel implementations of SOBI are studied in Li et al. (2019).
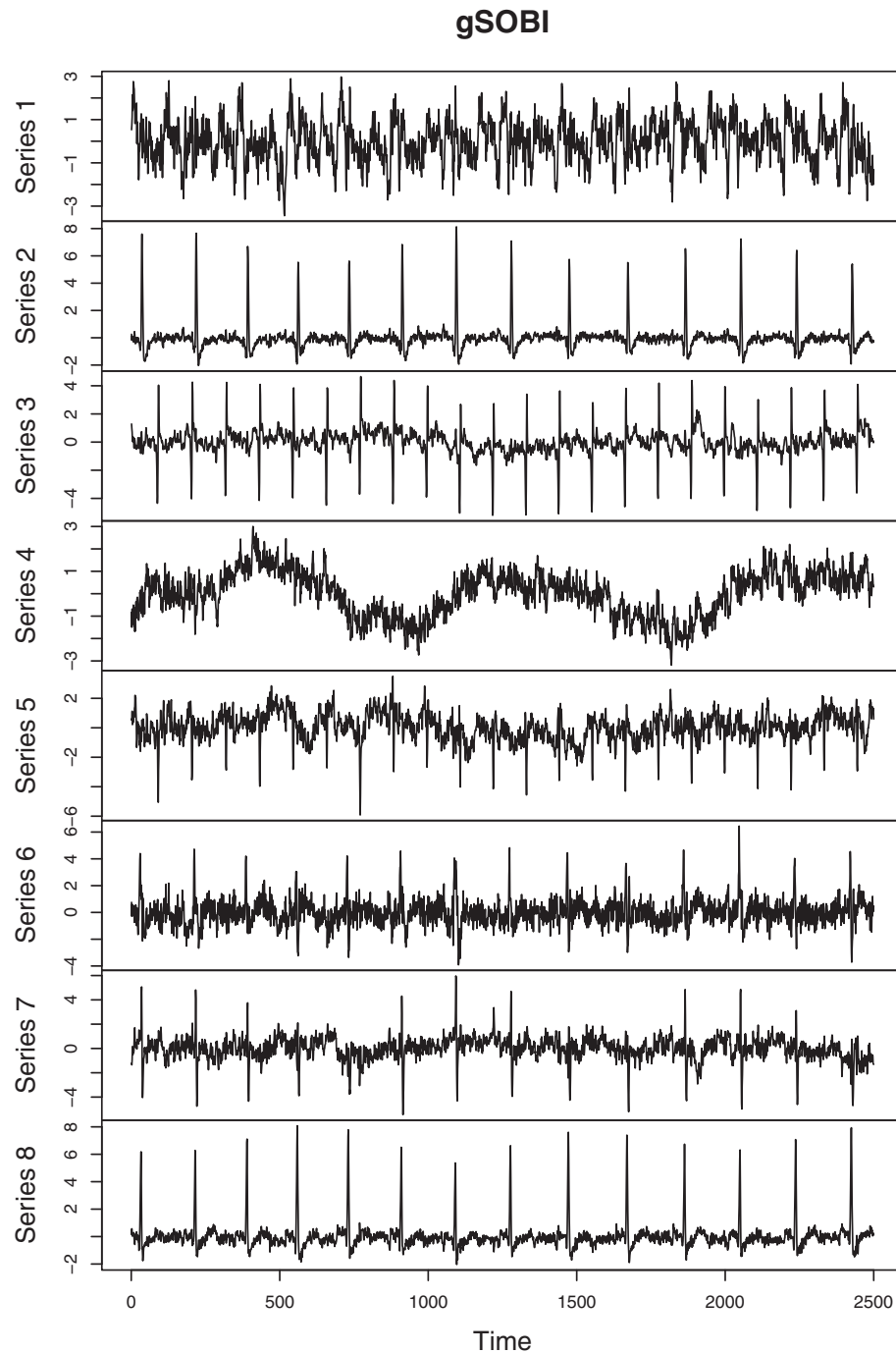
gSOBI

FIGURE 6 Electrocardiography latent source components using generalized second-order blind identification (gSOBI) with $w = 0.9$, $\mathcal{T}_1 = \{1, \ldots, 12\}$, and $\mathcal{T}_2 = \{1, 2, 3\}$

As the development of SOS methods is driven by basically two communities, signal processing and statistics implementations of the corresponding methods are usually available in the preferred software of these communities, that is, MATLAB (MATLAB, 2019) and R (R Core Team, 2020), respectively.

For Matlab, a nonexhaustive list of implementations of some of the above methods can be found from the EEGLAB toolbox (Delorme & Makeig, 2004), ICALAB toolbox (Cichocki et al., 2007) or BSSGUI (Petkov & Koldovský, 2009), for example. In R, especially more current methods are available in the packages JADE Miettinen et al. (2017), BSSasymp (Miettinen et al., 2017), tsBSS (Matilainen et al., 2019) and tensorBSS (Virta et al., 2020b).

A machine learning library running on many platforms and containing for example SOBI is SHOGUN (Sonnenburg et al., 2017).

6 | CONCLUSIONS

In this article we reviewed the latest developments in the area of SOS. SOS methods originate from signal processing literature, and, as seen in this article, the classical SOBI method has been extended since its appearance to various different time series settings. Besides signal processing community, the SOS methods have recently gained interest also among statisticians. The statistical properties of two well established SOBI unmixing matrix estimators, the deflation-based estimator and the symmetric estimator, are derived in Miettinen et al. (2014, 2016). Such derivations are useful when one, for example, wishes to develop a method that aims at more efficient signal separation result. We present in this article two efficient methods (eSOBI and aSOBI), which utilize asymptotic results of SOBI estimator. Some theoretical results for the generalized SOBI estimator (gSOBI) can also be found in literature (Miettinen, Matilainen, et al., 2020; Miettinen, Vorobyov, & Ollila, 2020). However, the statistical properties of other variants of SOBI remain mostly unknown and need to be studied.

As modern datasets are often high-dimensional and highly complex, more sophisticated SOBI methods for such data need to be developed. We already discussed here the methods developed recently for matrix-valued and tensor-valued data (see, e.g., Virta & Nordhausen, 2017b, and references therein). Other directions for future research include developing methods that can handle spatial and spatiotemporal dependencies as well as methods that are suitable for functional data. Notice that in Bachoc et al. (2020) the first step toward modeling spatial data was taken. When high-dimensional data are measured the main aim of the analysis is often to separate the signals of interest from noise components. Recent developments in this area can be found in Virta and Nordhausen (2021).

This review focused on the case where the mixing is instantaneous. However, there exists also many scenarios where it would be natural to consider that at time point t the observed signal is a mixture of current and past latent values. For example, in sound applications there might be in addition to the current sounds also some echos present. In the signal processing community such models are known as convolutive blind source separation models and are reviewed in Castella et al. (2010), for example. In this context, one can also mention the dynamic factor model (DFM) considered in Forni et al. (2000); Bai and Ng (2002); Stock and Watson (2002b, 2002a); Forni et al. (2004); Forni et al. (2017), among others. DFMs can also be considered in a convolution framework and can be combined with the external noise model mentioned earlier and generalized thereof. DFMs, with the special case dynamic principal component analysis (Brillinger, 1974), work usually in the frequency domain and can be applied also when the dimension is very high, unlike the SOS models discussed in this article, which usually assume the number of components to be much smaller than the number of available time points. The connection between the models and methods proposed in the DFM literature with those proposed in signal processing community need definitely careful consideration in future research. Finally, one issue that seems to be understudied in the field of SOS, is robustness. Robustness is a major issue for multivariate analysis because outliers become increasingly common as the dimension of data increases. However, as seen in this article, only few robust proposals for SOS exist in the literature.

AUTHOR CONTRIBUTIONS
Yan Pan: Methodology. **Markus Matilainen:** Methodology. **Sara Taskinen:** Methodology. **Klaus Nordhausen:** Methodology.

CONFLICT OF INTEREST
The authors have declared no conflicts of interest for this article.

ORCID
Yan Pan https://orcid.org/0000-0002-6870-3015
Markus Matilainen https://orcid.org/0000-0002-5597-2670
Sara Taskinen https://orcid.org/0000-0001-9470-7258
Klaus Nordhausen https://orcid.org/0000-0002-3758-8501

RELATED WIREs ARTICLES

Independent component analysis: A statistical perspective

Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data

FURTHER READING

Ensor, K. B. (2013). Time series factor models. *WIREs: Computational Statistics*, *5*(2), 97–104.

Additional general references for BSS for the interested reader are Hyvärinen et al. (2001); Cichocki and Amari (2002); Chabriel et al. (2014). Joint diagonalization, that is used to solve the BSS problem, can also be used for other data structures and not just for time series data. Joint diagonalization is used for iid tensorial data in Virta et al. (2017, 2018), for functional data in Virta et al. (2020a) and for graph data in Miettinen, Vorobyov, and Ollila (2020).

Robust BSS approaches, especially for the iid setting, are discussed in Nordhausen et al. (2008); Hallin and Mehta (2015). More noise settings in the iid and time series setting are considered in Nordhausen et al. (2017); Pfister et al. (2019).

REFERENCES

Bachoc, F., Genton, M. G., Nordhausen, K., Ruiz-Gazen, A., & Virta, J. (2020). Spatial blind source separation. *Biometrika*, *107*(3), 627–646.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*, 191–221.

Belouchrani, A., Abed Meraim, K., Cardoso, J.-F., & Moulines, E. (1997). A blind source separation technique based on second order statistics. *IEEE Transactions on Signal Processing*, *45*, 434–444.

Belouchrani, A., & Amin, M. G. (1998). Blind source separation based on time–frequency signal representations. *IEEE Transactions on Signal Processing*, *46*(11), 2888–2897.

Belouchrani, A., & Cichocki, A. (2000). Robust whitening procedure in blind source separation context. *Electronics Letters*, *36*(24), 2050–2051.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Brillinger, D. (1974). Time series: Data analysis and theory. In *Number nid. 1 in International series in decision processes*. Holt, Rinehart, and Winston.

Cardoso, J.-F. (1994). On the performance of orthogonal source separation algorithms. In *Proceedings of the 7th European signal processing conference* (pp. 776–779). Retrieved from https://www.eurasip.org/index.php?option=com_content&view=article&id=80&Itemid=89

Cardoso, J.-F., & Souloumiac, A. (1993). Blind beamforming for non-Gaussian signals. *IEE Proceedings F-Radar and Signal Processing*, *140*, 362–370.

Castella, M., Chevreuil, A., & Pesquet, J.-C. (2010). Convolutive mixtures. In P. Comon & C. Jutten (Eds.), *Handbook of blind source separation. Independent component analysis and applications* (pp. 281–324). Academic Press.

Chabriel, G., Kleinsteuber, M., Moreau, E., Shen, H., Tichavsky, P., & Yeredor, A. (2014). Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications. *IEEE Signal Processing Magazine*, *31*(3), 34–43.

Choi, S., & Cichocki, A. (2000). 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* (Vol. 1, pp. 405–414). IEEE.

Cichocki, A., & Amari, S. (2002). *Adaptive blind signal and image processing: Learning algorithms and applications*. Wiley.

Cichocki, A., Ichi Amari, S., Siwek, K., Tanaka, T., Cruces, S., Georgiev, P., Leonowicz, Z., Rutkowski, T., Choi, S., Belouchrani, A., Barros, A., Thawonmas, R., Hoya, T., Hashimoto, W., Terazono, Y., & Watanabe, T. (2007). *ICALAB for signal processing*.

Clarkson, D. (1988). A least squares version of algorithm as 211: The f-g diagonalization algorithm. *Applied Statistics*, *37*, 317–321.

Cohen, L. (1995). *Time-frequency analysis. electrical engineering signal processing*. Prentice Hall PTR.

Comon, P., & Jutten, C. (2010). *Handbook of blind source separation. Independent component analysis and applications*. Academic Press.

Cong, F., He, Z., Hämäläinen, J., Leppänen, P. H., Lyytinen, H., Cichocki, A., & Ristaniemi, T. (2013). 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.

De Lathauwer, L., De Moor, B., & Vandewalle, J. (1995). Fetal electrocardiogram extraction by source subspace separation. In *IEEE SP/Athos workshop on higher-order statistics* (pp. 134–138). IEEE.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Firat, U., Engin, S. N., Saraclar, M., & Ertuzun, A. B. (2010). Wind speed forecasting based on second order blind identification and autoregressive model. In *Ninth international conference on machine learning and applications* (pp. 686–691). IEEE.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*, *82*, 540–554.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2004). The generalized dynamic factor model consistency and rates. *Journal of Econometrics*, *119*, 231–255.

Forni, M., Hallin, M., Lippi, M., & Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, *199*, 74–92.

Haldane, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika*, *35*(3–4), 414–417.

Hallin, M., & Mehta, C. (2015). R-estimation for asymmetric independent component analysis. *Journal of the American Statistical Association*, *110*(509), 218–232.

Harshman, R. A., & Lundy, M. E. (1994). Parafac: Parallel factor analysis. *Computational Statistics & Data Analysis*, *18*(1), 39–72.

Hettmansperger, T. P., & Randles, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika*, *89*(4), 851–860.

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

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. John Wiley & Sons.

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

Ilmonen, P., Nevalainen, J., & Oja, H. (2010). Characteristics of multivariate distributions and the invariant coordinate system. *Statistics & Probability Letters*, *80*(23), 1844–1853.

Ilmonen, P., Nordhausen, K., Oja, H., & Ollila, E. (2010). A new performance index for ICA: Properties computation and asymptotic analysis. In V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, & E. Vincent (Eds.), *Latent variable analysis and signal separation* (pp. 229–236). Springer.

Ilmonen, P., Nordhausen, K., Oja, H., & Theis, F. (2015). An affine equivariant robust second-order BSS method. In E. Vincent, A. Yeredor, Z. Koldovský, & P. Tichavský (Eds.), *Latent variable analysis and signal separation: 12th international conference LVA/ICA 2015* (pp. 328–335). Springer International Publishing.

Joyce, C. A., Gorodnitsky, I. F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, *41*(2), 313–325.

Kalogiannis, G., Karampelas, N., & Hassapis, G. (2017). A reworked sobi algorithm based on schur decomposition for eeg data processing. In *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 268–271). IEEE.

Koldovský, Z., Tichavský, P., & Oja, E. (2006). Efficient variant of algorithm FastICA for independent component analysis attaining the cramér-rao lower bound. *IEEE Transactions on Neural Networks*, *17*(5), 1265–1277.

Kroonenberg, P. M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, *45*(1), 69–97.

Lathauwer, L. D., & Castaing, J. (2008). Blind identification of underdetermined mixtures by simultaneous matrix diagonalization. *IEEE Transactions on Signal Processing*, *56*(3), 1096–1105.

Li, H., Du, J., Du, Y., Chen, Z., & Xiao, N. (2019). P-sobi: A parallel implementation for second order blind identification algorithm. In *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* (pp. 1414–1420). IEEE.

Lietzén, N., Nordhausen, K., & Ilmonen, P. (2017). Complex valued robust multidimensional SOBI. In P. Tichavský, M. Babaie-Zadeh, O. J. Michel, & N. Thirion-Moreau (Eds.), *Latent variable analysis and signal separation: 13th international conference LVA/ICA 2017* (pp. 131–140). Springer International Publishing.

Matilainen, M., Croux, C., Miettinen, J., Nordhausen, K., Oja, H., Taskinen, S., & Virta, J. (2019). *tsBSS: Blind source separation and supervised dimension reduction for time series*. R package version 0.5.5.

Matilainen, M., Miettinen, J., Nordhausen, K., Oja, H., & Taskinen, S. (2017). On independent component analysis with stochastic volatility models. *Austrian Journal of Statistics*, *46*(3–4), 57–66.

Matilainen, M., Nordhausen, K., & Oja, H. (2015). New independent component analysis tools for time series. *Statistics & Probability Letters*, *105*, 80–87.

Matilainen, M., Nordhausen, K., & Virta, J. (2018). On the number of signals in multivariate time series. In *International conference on latent variable analysis and signal separation* (pp. 248–258). Springer.

MATLAB. (2019). *Version 9.7.0 (R2019b)*. The MathWorks Inc.

Matteson, D. S., & Tsay, R. S. (2011). Dynamic orthogonal components for multivariate time series. *Journal of the American Statistical Association*, *106*(496), 1450–1463.

Miettinen, J. (2015). Alternative diagonality criteria for SOBI. In K. Nordhausen & S. Taskinen (Eds.), *Modern nonparametric, robust and multivariate methods: Festschrift in honour of Hannu Oja* (pp. 455–469). Springer International Publishing.

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

Miettinen, J., Matilainen, M., Nordhausen, K., & Taskinen, S. (2020). Extracting conditionally heteroskedastic components using independent component analysis. *Journal of Time Series Analysis*, *41*, 293–311.

Miettinen, J., Nordhausen, K., Oja, H., & Taskinen, S. (2012). Statistical properties of a blind source separation estimator for stationary time series. *Statistics & Probability Letters*, *82*, 1865–1873.

Miettinen, J., Nordhausen, K., Oja, H., & Taskinen, S. (2014). Deflation-based separation of uncorrelated stationary time series. *Journal of Multivariate Analysis*, *123*, 214–227.

Miettinen, J., Nordhausen, K., & Taskinen, S. (2017). Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp. *Journal of Statistical Software*, *76*(2), 1–31.

Miettinen, J., Taskinen, S., Nordhausen, K., & Oja, H. (2015). Fourth moments and independent component analysis. *Statistical Science*, *30*, 372–390.

Miettinen, J., Vorobyov, S. A., & Ollila, E. (2020). Blind source separation of graph signals. In *ICASSP 2020—2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 5645–5649). IEEE.

Nordhausen, K. (2014). On robustifying some second order blind source separation methods for nonstationary time series. *Statistical Papers*, *55*(1), 141–156.

Nordhausen, K., & Oja, H. (2018). Independent component analysis: A statistical perspective. *WIREs Computational Statistics*, *10*(5), e1440.

Nordhausen, K., Oja, H., & Ollila, E. (2008). Robust independent component analysis based on two scatter matrices. *Austrian Journal of Statistics*, *37*, 91–100.

Nordhausen, K., Oja, H., Tyler, D., & Virta, J. (2017). Asymptotic and bootstrap tests for the dimension of the non-gaussian subspace. *IEEE Signal Processing Letters*, *24*(6), 887–891.

Nordhausen, K., & Tyler, D. E. (2015). A cautionary note on robust covariance plug-in methods. *Biometrika*, *102*(3), 573–588.

Nordhausen, K., & Virta, J. (2018). Ladle estimator for time series signal dimension. In *IEEE statistical signal processing workshop (SSP)* (pp. 428–432). IEEE.

Onatski, A. (2009). A formal statistical test for the number of factors in the approximate factor models. *Econometrica*, *77*(5), 1447–1480.

Petkov, J., & Koldovský, Z. (2009). Bssgui—A package for interactive control of blind source separation algorithms in matlab. In *Cross-modal analysis of speech, gestures, gaze and facial expressions* (pp. 386–398). Springer.

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

Pham, D. (2001). Joint approximate diagonalization of positive definite hermitian matrices. *SIAM Journal on Matrix Analysis and Applications*, *22*, 1136–1152.

Piccolotto, N., Bögl, M., Gschwandtner, T., Muehlmann, C., Nordhausen, K., Filzmoser, P., and Miksch, S. (2020). *TBSSvis: Visual analytics for temporal blind source separation*. Retrieved from https://github.com/npiccolotto/tbss-vis.

Popescu, T. D., & Manolescu, M. (2007). Blind source separation of traffic-induced vibrations in building monitoring. In *IEEE international conference on control and automation* (pp. 2101–2106). IEEE.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. R version 4.0.2.

Rajih, M., Comon, P., & Harshman, R. A. (2008). Enhanced line search: A novel method to accelerate parafac. *SIAM Journal on Matrix Analysis and Applications*, *30*(3), 1128–1147.

Shi, Z., Jiang, Z., & Zhou, F. (2009). Blind source separation with nonlinear autocorrelation and non-Gaussianity. *Journal of Computational and Applied Mathematics*, *223*(1), 908–915.

Sonnenburg, S., Strathmann, H., Lisitsyn, S., Gal, V., García, F. J. I., Lin, W., De, S., Zhang, C., frx, tklein23 Andreev, E., JonasBehr, sploving, Mazumdar, P., Widmer, C., Zora, P. D., Toni, G. D., Mahindre, S., Kislay, A., Hughes, K., … Esser, B. (2017). *shogun-toolbox/shogun: Shogun 6.1.0.* https://doi.org/10.5281/zenodo.1067840

Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Tang, A. (2010). Applications of second order blind identification to high-density EEG-based brain imaging: A review. In L. Zhang, B.-L. Lu, & J. Kwok (Eds.), *Advances in neural networks—ISNN 2010: 7th international symposium on neural networks, part II* (pp. 368–377). Springer.

Tang, A. C., Liu, J.-Y., & Sutherland, M. T. (2005). Recovery of correlated neuronal sources from eeg: The good and bad ways of using sobi. *NeuroImage*, *28*(2), 507–519.

Tang, A. C., Sutherland, M. T., & McKinney, C. J. (2005). Validation of SOBI components from high-density EEG. *NeuroImage*, *25*(2), 539–553.

Taskinen, S., Miettinen, J., & Nordhausen, K. (2016). A more efficient second order blind identification method for separation of uncorrelated stationary time series. *Statistics & Probability Letters*, *116*, 21–26.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes—A study of daily sugar prices 1961–79. In O. D. Anderson (Ed.), *Time series analysis: Theory and practice 1* (pp. 203–216). Springer.

Theis, F. J., & Inouye, Y. (2006). On the use of joint diagonalization in blind signal processing. In *IEEE international symposium on circuits and systems* (pp. 3589–3593). IEEE.

Theis, F. J., Meyer-Bäse, A., & Lang, E. W. (2004). Second-order blind source separation based on multi-dimensional autocovariances. In C. G. Puntonet & A. Prieto (Eds.), *Independent component analysis and blind signal separation: Fifth international conference* (pp. 726–733). Springer.

Theis, F. J., Müller, N. S., Plant, C., & Böhm, C. (2010). Robust second-order source separation identifies experimental responses in biomedical imaging. In V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, & E. Vincent (Eds.), *International conference on latent variable analysis and signal separation* (pp. 466–473). Springer.

Tichavský, P., Doron, E., Yeredor, A., & Nielsen, J. (2006). A computationally affordable implementation of an asymptotically optimal BSS algorithm for AR sources. In *14th European signal processing conference* (pp. 1–5). European Association for Signal Processing.

Tichavský, P., Koldovský, Z., Doron, E., Yeredor, A., & Herrero, G. G. (2006). Blind signal separation by combining two ICA algorithms: HOS-based EFICA and time structure-based WASOBI. In *14th European signal processing conference* (pp. 1–5). European Association for Signal Processing.

Tichavský, P., & Yeredor, A. (2009). Fast approximate joint diagonalization incorporating weight matrices. *IEEE Transactions on Signal Processing*, *57*(3), 878–891.

Tong, L., Inouye, Y., & Liu, R.-W. (1992). A finite-step global convergence algorithm for the parameter estimation of multichannel ma processes. *IEEE Transactions on Signal Processing*, *40*(10), 2547–2558.

Tong, L., Soon, V., Huang, Y., & Liu, R. (1990). AMUSE: A new blind identification algorithm. In *Proceedings of IEEE international symposium on circuits and systems* (pp. 1784–1787). IEEE.

Virta, J., Li, B., Nordhausen, K., & Oja, H. (2017). Independent component analysis for tensor-valued data. *Journal of Multivariate Analysis*, *162*, 172–192.

Virta, J., Li, B., Nordhausen, K., & Oja, H. (2018). Jade for tensor-valued observations. *Journal of Computational and Graphical Statistics*, *27*(3), 628–637.

Virta, J., Li, B., Nordhausen, K., & Oja, H. (2020a). Independent component analysis for multivariate functional data. *Journal of Multivariate Analysis*, *176*, 104568.

Virta, J., Li, B., Nordhausen, K., & Oja, H. (2020b). *tensorBSS: Blind source separation methods for tensor-valued observations*. R package version 0.3.6.

Virta, J., & Nordhausen, K. (2017a). Blind source separation for nonstationary tensor-valued time series. In *2017 IEEE 27th international workshop on machine learning for signal processing (MLSP)* (pp. 1–6). IEEE.

Virta, J., & Nordhausen, K. (2017b). Blind source separation of tensor-valued time series. *Signal Processing*, *141*, 204–216.

Virta, J., & Nordhausen, K. (2019). Estimating the number of signals using principal component analysis. *Stat*, *8*, e231.

Virta, J., & Nordhausen, K. (2021). Determining the signal dimension in second order source separation. *Statistica Sinica*, *31*, 135–156.

Visuri, S., Koivunen, V., & Oja, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference*, *91*, 557–575.

Wax, M., & Kailath, T. (1985). Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *33*, 387–392.

Weisman, T., & Yeredor, A. (2006). Separation of periodically time-varying mixtures using second-order statistics. In *International conference on independent component analysis and signal separation* (pp. 278–285). Springer.

Yeredor, A. (2000). Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting. *IEEE Signal Processing Letters*, *7*(7), 197–200.

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

Yeredor, A. (2003). TV-SOBI: An expansion of SOBI for linearly time-varying mixtures. In *Proceedings of the 4th international symposium on independent component analysis and blind source separation*. ISBN 4-9901531-1-1.

Zhang, X., Vialatte, F.-B., Chen, C., Rathi, A., & Dreyfus, G. (2015). Embedded implementation of second-order blind identification (SOBI) for real-time applications in neuroscience. *Cognitive Computation*, *7*(1), 56–63.

Zhao, L., Krishnaiah, P., & Bai, Z. (1986). On detection of the number of signals in presence of white noise. *Journal of Multivariate Analysis*, *20*(1), 1–25.

Ziehe, A., Laskov, P., Müller, K. R., & Nolte, G. (2003). A linear least-squares algorithm for joint diagonalization. In *4th international symposium on independent component analysis and blind signal separation*, ISBN 4-9901531-1-1. (pp. 469–474).

Ziehe, A., & Müller, K. R. (1998). TDSEP—An efficient algorithm for blind separation using time structure. In *ICANN 98. Perspectives in neural computing* (pp. 675–680). Springer.