Independent Component Analysis for Compositional Data

Christoph Muehlmann, Kamila Fačevicová, Alžběta Gardlo, Hana Janečková, and Klaus Nordhausen

Abstract Compositional data represent a specific family of multivariate data, where the information of interest is contained in the ratios between parts rather than in absolute values of single parts. The analysis of such specific data is challenging as the application of standard multivariate analysis tools on the raw observations can lead to spurious results. Hence, it is appropriate to apply certain transformations prior to further analysis. One popular multivariate data analysis tool is independent component analysis. Independent component analysis aims to find statistically independent components in the data and as such might be seen as an extension to principal component analysis. In this paper, we examine an approach of how to apply independent component analysis on compositional data by respecting the nature of the latter and demonstrate the usefulness of this procedure on a metabolomics dataset.

C. Muehlmann · K. Nordhausen (✉)
Institute of Statistics & Mathematical Methods in Economics,
Vienna University of Technology, Wiedner Hauptstr. 7, 1040 Vienna, Austria
e-mail: klaus.nordhausen@tuwien.ac.at

C. Muehlmann
e-mail: christoph.muehlmann@tuwien.ac.at

K. Fačevicová
Department of Mathematical Analysis and Applications of Mathematics,
Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic
e-mail: kamila.facevicova@gmail.com

A. Gardlo
Department of Clinical Biochemistry, University Hospital Olomouc and Palacký University
Olomouc, I.P. Pavlova 185/6, 779 00 Olomouc, Czech Republic
e-mail: alzbetagardlo@gmail.com

H. Janečková
Laboratory for Inherited Metabolic Disorders, Department of Clinical Biochemistry,
University Hospital Olomouc and Palacký University Olomouc, I.P. Pavlova 185/6,
779 00 Olomouc, Czech Republic
e-mail: janeckovah@gmail.com

525

1   Introduction

Independent component analysis (ICA) is a well-established data analysis method in signal processing with the goal of recovering hidden signals that are usually meant to have a physical meaning. In recent years, ICA methods have attracted increasing interest in the statistics community as an extension of normality-based multivariate methods that only use second-order moments. In principle, ICA can be seen as a refinement of principal component analysis where, after removing second-order information, higher order moments are used to search for hidden structures which are not visible in the principal components. Classical ICA methods are mainly developed for independent and identically distributed observations in a Euclidean space. Nevertheless, these methods are also applied, for example, on time series, spatial data, etc. but to the best of our knowledge not on iid compositional data.

Compositional data is special in the way that the entries (parts) of a d-variate vector are positive and carry relative rather than absolute information about the respective observation of interest. Moreover, the parts of the compositional vector are by nature not independent and in some specific situations, e.g. when all parts are bounded by a constant sum constraint, a spurious correlation between them is present. Therefore, compositional data lies on a simplex and does not follow the real Euclidean geometry. Examples of compositional data are geochemical data where the chemical composition of soil samples is of interest, the composition of nutrients of food intake or the distribution of market shares. For further details and examples of compositional data, see, for example, Aitchison (1986), Egozcue and Pawlowsky-Glahn (2019), Fačevicová et al. (2016), Filzmoser et al. (2018), Morais et al. (2018), Pawlowsky-Glahn and Buccianti (2011), Trinh et al. (2019).

It is well established that standard multivariate methods should not be applied directly to compositional data. Either methods which take the geometry of compositional data into account or methods that transform compositional data in such a way that standard multivariate analysis tools can be applied are appropriate. In this paper, we take the latter approach.

We review some basic ICA methods in Sect. 2. In Sect. 3, we describe compositional data and methods to transform such data into the real space. Based on the former two sections, we present how ICA can be performed on compositional data in Sect. 4 and conclude the paper with the analysis of a metabolomics dataset from healthy newborns in Sect. 5 and a discussion in Sect. 6.

2   Independent Component Analysis

From a statistical perspective, independent component analysis is usually formulated as a latent variable model as follows.

Definition 1 An observable p-vector $\mathbf{x}$ follows the independent component (IC) model if

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

where $\mathbf{A}$ is a $p \times p$ non-singular matrix, $\mathbf{b}$ a p-vector, and the latent p-variate random vector $\mathbf{z}$ satisfies

(A1) $\mathbf{E}(\mathbf{z}) = \mathbf{0}$ and $\mathbf{COV}(\mathbf{z}) = \mathbf{I}_p$,
(A2) the components of $\mathbf{z}$ are independent, and
(A3) at most one component of $\mathbf{z}$ is Gaussian.

Thus $\mathbf{E}(\mathbf{x}) = \mathbf{b}$ and $\mathbf{COV}(\mathbf{x}) = \mathbf{A}\mathbf{A}^\top$. The goal of ICA is to find a $p \times p$ matrix $\mathbf{W}$ such that $\mathbf{W}\mathbf{x}$ has independent components. Note however that in general it will not hold that $\mathbf{W}(\mathbf{x} - \mathbf{b}) = \mathbf{z}$ as the IC model assumptions only fix the location and scale of $\mathbf{z}$ but not the signs or the order of the components. Therefore, for every solution $\mathbf{W}$, also $\mathbf{PJW}$ is a solution, where $\mathbf{P}$ is a $p \times p$ permutation matrix (1 per row and column, 0 elsewhere) and $\mathbf{J}$ is a $p \times p$ sign-change matrix (a diagonal matrix with ± 1 on its diagonal).

There are many suggestions in the literature on how to estimate $\mathbf{W}$ based on a sample $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, and for recent reviews see, for example, Comon and Jutten (2010), Nordhausen and Oja (2018). Almost all ICA methods make, however, use of the following result:

Key result Let $\mathbf{x}$ follow the IC model and denote $\mathbf{x}^{st} = \mathbf{COV}(\mathbf{x})^{-1/2}(\mathbf{x} - \mathbf{E}(\mathbf{x}))$, then there exists an orthogonal $p \times p$ matrix $\mathbf{U}$ such that

$$\mathbf{U}^\top \mathbf{x}^{st} = \mathbf{z}.$$

This result implies that after estimating $\mathbf{COV}(\mathbf{x})$ and $\mathbf{E}(\mathbf{x})$, the problem is reduced from finding a general $p \times p$ matrix to a $p \times p$ orthogonal matrix. Also note that this means that the performance of ICA methods does not depend on the values of $\mathbf{A}$ and $\mathbf{b}$, as these are accounted for when standardizing the data. An unmixing matrix estimate is therefore obtained as $\mathbf{W} = \mathbf{U}^\top \mathbf{COV}(\mathbf{x})^{-1/2}$ and different ICA approaches differ in the way they estimate $\mathbf{U}$. In the following, we will show how some popular ICA methods estimate this rotation.

2.1 FOBI

Fourth-order blind identification (FOBI), presented in Cardoso (1989), was one of the first ICA methods but is still popular as it has a closed-form solution. For FOBI, we need to define the scatter matrix of fourth-order moments

$$\mathbf{COV}_4(\mathbf{x}) = \frac{1}{p+2} \mathbf{E}\left((\mathbf{x} - \mathbf{E}(\mathbf{x}))^\top \mathbf{COV}(\mathbf{x})^{-1}(\mathbf{x} - \mathbf{E}(\mathbf{x}))(\mathbf{x} - \mathbf{E}(\mathbf{x}))(\mathbf{x} - \mathbf{E}(\mathbf{x}))^\top \right).$$

Then we can define the following:

Definition 2 The FOBI unmixing matrix is $\mathbf{W}_{\text{FOBI}} = \mathbf{U}_{\text{FOBI}}^{\top}\mathbf{COV}(\mathbf{x})^{-1/2}$ where the columns of $\mathbf{U}_{\text{FOBI}}$ are given by the eigenvectors of $\mathbf{COV}_4(\mathbf{x}^{st})$.

From denoting $\mathbf{U}_{\text{FOBI}}\mathbf{D}\mathbf{U}_{\text{FOBI}}^{\top}$ to be the eigendecomposition of $\mathbf{COV}_4(\mathbf{x}^{st})$ which is needed to compute $\mathbf{W}_{\text{FOBI}}$, it is obvious that FOBI is only unique when the eigenvalues contained in the diagonal matrix $\mathbf{D}$ are distinct. One can actually show that these eigenvalues are linked to the kurtosis values of the independent components. For FOBI to be well-defined, Assumption (A3) from the IC model needs to be replaced by the stronger assumption:

(A4) The kurtosis values of the independent components must be distinct.

FOBI is often the first ICA method applied as it is quick to compute, gives a fast first impression, and its statistical properties are well known; see, for example, Miettinen et al. (2015), Nordhausen and Virta (2019) for more details. FOBI can also be of interest outside the IC model and can be seen as an invariant coordinate selection method (Tyler et al. 2009).

2.2 JADE

Assumption (A4) is considered highly restrictive. Joint approximate diagonalization of eigenmatrices (JADE) can be seen as an extension of FOBI which relaxes this strict assumption, Cardoso and Souloumiac (1993).

For JADE, we have to define the fourth-order cumulant matrices

$$\mathbf{C}_{ij}(\mathbf{x}) = \mathbf{E}\left((\mathbf{x}^{st\top}\mathbf{E}_{ij}\mathbf{x}^{st})\mathbf{x}^{st}\mathbf{x}^{st\top}\right) - \mathbf{E}_{ij} - \mathbf{E}_{ij}^{\top} - \text{tr}(\mathbf{E}_{ij})\mathbf{I}_p,$$

where $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j^{\top}$ with $\mathbf{e}_i$ being a vector of dimension p with the ith element equals 1 and 0 otherwise. As i and j range from 1 to p, there are in total p^2 such cumulant matrices. In the IC model, $\mathbf{C}_{ij}(\mathbf{z}) = 0$ if $i \neq j$ and for the case where $i = j$ $\mathbf{C}_{ii}(\mathbf{z})$ corresponds to the kurtosis of the ith component. The matrix of fourth moments can actually be expressed as

$$\mathbf{COV}_4(\mathbf{x}) = \frac{1}{p+2}\sum_{i=1}^{p}\mathbf{C}_{ii}(\mathbf{x}) + (p+2)\mathbf{I}_p,$$

meaning that it uses not all possible cumulant information. The idea of JADE is to exploit the information contained in all cumulant matrices.

Definition 3 The JADE unmixing matrix is $\mathbf{W}_{\text{JADE}} = \mathbf{U}_{\text{JADE}}^{\top}\mathbf{COV}(\mathbf{x})^{-1/2}$ where $\mathbf{U}_{\text{JADE}}$ is the maximizer of

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

Thus, JADE tries to maximize the diagonal elements of $\mathbf{U}^\top \mathbf{C}_{ij}(\mathbf{x}^{st})\mathbf{U}$ which is equivalent to minimize the off-diagonal elements by the orthogonal invariance of the Frobenius norm $|| \cdot ||_F$. As in the IC model, only $\mathbf{C}_{ii}(\mathbf{z})$ is non-zero and corresponds to the kurtosis of z_i. This means that JADE relaxes the FOBI assumption (A4) to the following:

(A5) At most one independent component can have zero kurtosis.

For a finite sample, the joint diagonalization of more than two matrices needs to be carried out approximately; many algorithms that jointly diagonalize two or more matrices are available; see, for example, Illner et al. (2015). For the purpose of this paper, we will use an algorithm based on Givens rotations, Clarkson (1988).

The statistical properties of JADE are, for example, given in Miettinen et al. (2015); from an asymptotic point of view, FOBI is never superior compared to JADE. JADE is however computationally more expensive, especially when the number of independent components grows, as p^2 matrices need to be computed and jointly diagonalized.

As a compromise, k-JADE was suggested in Miettinen et al. (2013). The idea is to use not all matrices $\mathbf{C}_{ij}$, but only those whose indices are not too far apart, i.e. $|i - j| < k$. This requires however that the first step, the whitening step, is not done using just the covariance matrix but using $\mathbf{W}_{\text{FOBI}}$.

Definition 4 Denote $\mathbf{x}^{st'} = \mathbf{W}_{\text{FOBI}}(\mathbf{x} - \mathbf{E}(\mathbf{x}))$ and choose an integer $1 \leq k \leq p$, then the k-JADE unmixing matrix is $\mathbf{W}_{\text{kJADE}} = \mathbf{U}_{\text{kJADE}}^\top \mathbf{W}_{\text{FOBI}}$ where $\mathbf{U}_{\text{kJADE}}$ is the maximizer of

$$\sum_{|i-j|<k}^{p} ||\text{diag}(\mathbf{U}^\top \mathbf{C}_{ij}(\mathbf{x}^{st'})\mathbf{U})||_F^2.$$

The value k is basically a tuning parameter. The intuition is that the multiplicities of the distinct non-zero kurtosis values of the independent components are at most k, and that there is at most one component having kurtosis zero. Usually, k is simply chosen by the user based on expert knowledge. In Virta et al. (2020), some guidelines for the selection are offered, which are however not very practical. The statistical properties of k-JADE are given in Miettinen et al. (2013), Virta et al. (2020). It can be shown that for a value of k which fulfills the multiplicity condition, k-JADE is asymptotically as efficient as JADE but has, if k is small, a much smaller computational complexity.

2.3 FastICA

FOBI, JADE, and k-JADE are often called algebraic ICA methods. Another large group of ICA methods is based on projection pursuit ideas, where the most prominent one is FastICA. It was originally suggested in Hyvärinen (1999a). Some of the many FastICA variants are discussed below.

The general idea of FastICA is to find the column vectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$ of $\mathbf{U}$ which maximize the non-Gaussianity of the components of $\mathbf{U}^\top \mathbf{x}^{st}$. Non-Gaussianity of a univariate random variable x is measured by $|E(G(x))|$ with some twice continuously differentiable and non-quadratic function G that satisfies $E(G(y)) = 0$ for $y \sim N(0, 1)$. The most popular choices for G are

pow3: $G(x) = (x^4 - 3)/4$,
 tanh: $G(x) = \log(\cosh(x)) - c_t$, and
gauss: $G(x) = -\exp(-x^2/2) - c_g$.

The constants $c_t = E(\log(\cosh(y))) \approx 0.375$ and $c_g = E(-\exp(-y^2/2)) \approx -0.707$ are normalizing constants. The derivatives of G, denoted as g, are called non-linearities and are the name givers as pow3 : $g(x) = x^3$, tanh : $g(x) = \tanh(x)$ and gauss : $g(x) = x \exp(-x^2/2)$.

2.3.1   Deflation-Based FastICA

FastICA was first suggested in Hyvärinen and Oja (1997) using the non-linearity pow3 and finding the column vectors of $\mathbf{U}_{DF}$ one after another which is now known as deflation-based FastICA.

Definition 5 The deflation-based FastICA unmixing matrix is defined as $\mathbf{W}_{DF} = \mathbf{U}_{DF}^\top \mathbf{COV}(\mathbf{x})^{-1/2}$, where the kth column of $\mathbf{U}$, $\mathbf{u}_k$, maximizes

$$|\mathbf{E}[G(\mathbf{u}_k^\top \mathbf{x}_{st})]|$$

under the constraints $\mathbf{u}_k^T \mathbf{u}_k = 1$ and $\mathbf{u}_j^T \mathbf{u}_k = 0, \; j = 1, \ldots, k-1$.

To obtain estimates, a modified Newton-Raphson algorithm is used which iterates the following steps until convergence:

$$\mathbf{u}_k \leftarrow \mathbf{E}[g(\mathbf{u}_k^\top \mathbf{x}_{st})\mathbf{x}_{st}] - \mathbf{E}[g'(\mathbf{u}_k^\top \mathbf{x}_{st})]\mathbf{u}_k$$

$$\mathbf{u}_k \leftarrow \left(\mathbf{I}_p - \sum_{l=1}^{k-1} \mathbf{u}_l \mathbf{u}_l^\top \right) \mathbf{u}_k$$

$$\mathbf{u}_k \leftarrow ||\mathbf{u}_k||^{-1} \mathbf{u}_k.$$

The last two steps perform the Gram-Schmidt orthonormalization.

The properties of deflation-based FastICA have been studied in detail in Ollila (2010), Nordhausen et al. (2011). One issue with deflation-based FastICA is that besides the global maximum it has many local maxima and the order in which the vectors $\mathbf{u}_k$ are found depends heavily on the initial value of the algorithm, where in turn the estimation performance depends on the order in which the vectors $\mathbf{u}_k$ are found. Using asymptotic arguments, Nordhausen et al. (2011) suggested reloaded

Table 1 Table of default candidate set of non-linearities of adaptive deflation-based FastICA, where $(x)_+ = x$ if $x > 0$ and 0 otherwise, and $(x)_- = x$ if $x < 0$ and 0 otherwise

$g_1(x) = x^3$	$g_6(x) = (x)_+^2 + (x)_-^2$	$g_{11}(x) = (x - 1.0)_+^2 + (x + 1.0)_-^2$
$g_2(x) = \tanh(x)$	$g_7(x) = (x - 0.2)_+^2 + (x + 0.2)_-^2$	$g_{12}(x) = (x - 1.2)_+^2 + (x + 1.2)_-^2$
$g_3(x) = x \exp(-x^2/2)$	$g_8(x) = (x - 0.4)_+^2 + (x + 0.4)_-^2$	$g_{13}(x) = (x - 1.4)_+^2 + (x + 1.4)_-^2$
$g_4(x) = (x + 0.6)_-^2$	$g_9(x) = (x - 0.6)_+^2 + (x + 0.6)_-^2$	$g_{14}(x) = (x - 1.6)_+^2 + (x + 1.6)_-^2$
$g_5(x) = (x - 0.6)_+^2$	$g_{10}(x) = (x - 0.8)_+^2 + (x + 0.8)_-^2$	

FastICA, which estimates first the independent components using FOBI or k-JADE and then derives an optimal order based on the estimated independent components.

The idea of reloaded FastICA to fix the extraction order based on asymptotic arguments was extended in Miettinen et al. (2014) to also select an optimal non-linearity for each component out of a candidate set of possible non-linearities. This is known as adaptive deflation-based FastICA. We will denote the adaptive deflation-based FastICA unmixing matrix as $\mathbf{W}_{\mathrm{ADF}}$. The candidate set of non-linearities suggested in Miettinen et al. (2014) contains, for example, the non-linearities presented in Table 1.

2.3.2 Symmetric FastICA

A FastICA variant estimating all directions in parallel was suggested in Hyvärinen (1999b).

Definition 6 The symmetric FastICA estimator $\mathbf{W}_{\mathrm{SF}} = \mathbf{U}_{\mathrm{SF}}^\top \mathbf{COV}(\mathbf{x})^{-1/2}$ uses as a criterion for $\mathbf{U}_{\mathrm{SF}}$

$$\sum_{j=1}^{p} |\mathbf{E}[G(\mathbf{u}_j^\top \mathbf{x}_{st})]|$$

which should be maximized under the orthogonality constraint $\mathbf{U}_{\mathrm{SF}}^\top \mathbf{U}_{\mathrm{SF}} = \mathbf{I}_p$.

The steps of the iterative algorithm to compute $\mathbf{U}_{\mathrm{SF}}$ are

$$\mathbf{u}_k \leftarrow \mathbf{E}[g(\mathbf{u}_k^T \mathbf{x}_{st})\mathbf{x}_{st}] - \mathbf{E}[g'(\mathbf{u}_k^\top \mathbf{x}_{st})]\mathbf{u}_k, \quad k = 1, \ldots, p$$
$$\mathbf{U}_{\mathrm{SF}}^\top \leftarrow (\mathbf{U}_{\mathrm{SF}}^\top \mathbf{U}_{\mathrm{SF}})^{-1/2}\mathbf{U}_{\mathrm{SF}}^\top.$$

The first update step of the algorithm is similar to that of the deflation-based FastICA estimator. The orthogonalization step can be interpreted as taking an average

over the vectors of the first step. This differs from the deflation-based approach where errors made in the kth direction carry on to the following directions and therefore the errors accumulate. This is often the reason why symmetric FastICA is usually considered superior to the deflation-based FastICA. However, there are also cases where the accumulation is preferable to the averaging. This occurs when some independent components are easier to find than the others. Statistical properties of symmetric FastICA are given in Miettinen et al. (2015), Wei (2015), Miettinen et al. (2017).

2.3.3 Squared Symmetric FastICA

One of the most recent variants of FastICA is the squared symmetric FastICA estimator (Miettinen et al. 2017). The idea of this estimator is to replace the absolute values in the objective function of the symmetric FastICA with squared values.

Definition 7 The squared symmetric FastICA estimator
$\mathbf{W}_{\text{S2F}} = \mathbf{U}_{\text{S2F}}^\top \mathbf{COV}(\mathbf{x})^{-1/2}$ obtains $\mathbf{U}_{\text{S2F}}$ as the maximizer of

$$\sum_{j=1}^{p} (\mathbf{E}[G(\mathbf{u}_j^\top \mathbf{x}_{st})])^2$$

under the orthogonality constraint $\mathbf{U}_{\text{S2F}}^\top \mathbf{U}_{\text{S2F}} = \mathbf{I}_p$.

The steps of the resulting algorithm are

$$\mathbf{u}_k \leftarrow \mathbf{E}[G(\mathbf{u}_k^\top \mathbf{x}_{st})](\mathbf{E}[g(\mathbf{u}_k^\top \mathbf{x}_{st})\mathbf{x}_{st}] - \mathbf{E}[g'(\mathbf{u}_k^\top \mathbf{x}_{st})]\mathbf{u}_k), \quad k = 1, \ldots, p,$$
$$\mathbf{U}_{\text{S2F}}^\top \leftarrow (\mathbf{U}_{\text{S2F}}^\top \mathbf{U}_{\text{S2F}})^{-1/2}\mathbf{U}_{\text{S2F}}^\top.$$

Thus, the first step of the algorithm equals the first step in the symmetric algorithm with an additional multiplication by $\mathbf{E}[G(\mathbf{u}_k^\top \mathbf{x}_{st})]$. Hence, the squared symmetric variant puts more weight on components that are "more" non-Gaussian, which most often, but not always, is advantageous. The properties of the squared symmetric FastICA estimator as well as comparisons to the deflation-based and symmetric FastICA methods are given in Miettinen et al. (2017). In Miettinen et al. (2017), it is also shown that if the non-linearity pow3 is used, symmetric squared FastICA is asymptotically equivalent to JADE.

Besides assumptions (A1)–(A3), deflation-based, symmetric, and squared symmetric FastICA need further assumptions based on G to ensure consistency. Assuming the order of the components is fixed as $|\mathbf{E}[G(z_1)]| \geq \cdots \geq |\mathbf{E}[G(z_p)]|$, then it is required that for any $\mathbf{z} = (z_1, \ldots, z_p)^\top$ with independent and standardized components and for any orthogonal matrix $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$, the following holds.

For deflation-based FastICA:

(A6) For all $k = 1, \ldots, p$, $|\mathbf{E}[G(\mathbf{u}_k^\top \mathbf{z})]| \leq |\mathbf{E}[G(z_k)]|$, when $\mathbf{u}_k^\top \mathbf{e}_j = 0$ for all $j = 1, \ldots, k-1$, where $\mathbf{e}_i$ is a p-vector with ith element one and others zero,

for symmetric FastICA

(A7) $|\mathbf{E}[G(\mathbf{u}_1^T \mathbf{z})]| + \cdots + |\mathbf{E}[G(\mathbf{u}_p^T \mathbf{z})]| \leq |\mathbf{E}[G(z_1)]| + \cdots + |\mathbf{E}[G(z_p)]|$,

and for squared symmetric FastICA

(A8) $(\mathbf{E}[G(\mathbf{u}_1^T \mathbf{z})])^2 + \cdots + (\mathbf{E}[G(\mathbf{u}_p^T \mathbf{z})])^2 \leq (\mathbf{E}[G(z_1)])^2 + \cdots + (\mathbf{E}[G(z_p)])^2$.

It was proven, for example, in Miettinen et al. (2015), that all three conditions are fulfilled with pow3. On the other hand, in the case of non-linearities like tanh and gauss some of these conditions might be violated for certain source distributions.

From a computational point of view, the advantage of both symmetric versions is that the initial value of $\mathbf{U}$ is not important when the sample size is large, as the algorithms converge usually to the global maxima.

To conclude this section we can point out that FOBI, JADE, k-JADE, symmetric FastICA, and squared symmetric FastICA are affine equivariant ICA methods which means that their performance does not depend on the mixing matrix. So, from this point of view, only deflation-based FastICA differs, which can be overcome when the reloaded version or adaptive version is used. Affine equivariance will be of relevance later when applying the ICA methods to compositional data.

3 Compositional Data and Its Real Space Representation

A specific family of d-dimensional vectors is present when each entry (part) of a vector is positive and carries information about its contribution to the whole. In the following, such multivariate observations are called (vector) compositional data, whose specifics were already described, utilized, and analyzed in a wide range of applications (Pawlowsky-Glahn and Buccianti 2011). The main property of compositional data is its relative nature, when the relevant information is contained in the ratios between parts rather than in the absolute values of the parts. Consider, e.g. a vector describing a geochemical structure of soil, where each part represents the quantity of the given element in the sample. The quantity can be given either in absolute scale, like in mg of the component contained in the sample, or some of its relative alternatives, typically ppm. While the mg representation depends on the overall size of the sample, the ppm one does not, despite the ratios between parts remaining unchanged. Both representations are therefore from the compositional point of view equivalent.

Due to the relative nature of compositional data, the sample space of representations of a d-part compositional vector $\mathbf{x}$ forms a d-part simplex

$$\mathcal{S}^d = \left\{ \mathbf{x} = (x_1, \ldots, x_d)^\top, \sum_{i=1}^{d} x_i = \kappa, \kappa > 0 \right\},$$

where the Aitchison geometry holds. The whole sample space is formed by equivalence classes of proportional vectors (Pawlowsky-Glahn et al. 2015, Chaps. 2, 3). Since most of the standard statistical methods are designed for real-valued data following the usual Euclidean geometrical structure, it is favorable to express compositional data in real coordinates prior to their analysis. One of the possible representations is the centered log-ratio (clr) transformation from $\mathcal{S}^d$ to $\mathbb{R}^d$ given by

$$\text{clr}(\mathbf{x})_i = \ln \frac{x_i}{g_m(\mathbf{x})} = \frac{1}{d} \sum_{j=1}^{d} \ln \frac{x_i}{x_j}, \quad \text{for} \quad i = 1, \ldots, d,$$

where $g_m(\mathbf{x})$ denotes the geometrical mean of all parts. The parts of the resulting clr vector can be interpreted in terms of the dominance of the compositional part in the numerator within the whole composition or equivalently as its mean dominance over each part of the whole composition. The use of logarithm symmetrizes this relationship. Let us stress here that the clr values depend on the set of compositional parts used for its computation and therefore the above interpretation holds true only when the whole composition is considered. Within the whole manuscript, the clr transformation based on all compositional parts will be of interest. On the other hand, from its construction, the clr coefficients/variables are not linearly independent, as they sum up to zero and, therefore, the whole clr vector falls in a $(d-1)$-dimensional subspace of $\mathbb{R}^d$. This feature prevents direct use of the clr representation within methods that require full rank data, like robust PCA (Filzmoser et al. 2009) or the above stated ICA methods.

One possible workaround is the isometric log-ratio (ilr) transformation, which represents the compositional vector $\mathbf{x}$ in a system of $d-1$ orthonormal real coordinates. This system can be obtained directly from the clr vector as

$$\text{ilr}(\mathbf{x}) = \mathbf{V}^\top \text{clr}(x),$$

where the columns of the $d \times d - 1$ log-contrast matrix $\mathbf{V}$ are given as $\mathbf{v}_i = \text{clr}(\xi_i)$ and the vectors ξ_i, $i = 1, \ldots, d-1$ constitute an orthonormal basis in $\mathcal{S}^d$. See Pawlowsky-Glahn and Buccianti (2011), Ch. 11 for details.

The system of basis vectors $\{\xi_1, \ldots, \xi_{d-1}\}$ is not uniquely given and can be chosen according to the purpose of further analysis. Since each system of ilr coordinates can be obtained as an orthogonal rotation of the others, its specific choice does not affect the results of their analysis, like predictions of the regression model with a compositional regressor or scores of the robust PCA model (Filzmoser et al. 2009; Hron et al. 2012). When it is required, a specific coordinate system can be selected by some data-driven method, like hierarchical clustering of the compositional parts, or using expert knowledge. In both cases, the main aim is to obtain such an interpretation of the coordinates at hand, which is favorable according to the given problem (Egozcue

and Pawlowsky-Glahn 2005). Since a specific interpretation of the ilr coordinates is not the main purpose here, the same system as in Nordhausen et al. (2015) is used. The basis vectors ξ_i have the value $\exp\left(\sqrt{1/i(i+1)}\right)$ at the first i positions, $\exp\left(-\sqrt{i/(i+1)}\right)$ at the position $i+1$, and 1 at the remaining ones. Consequently, the columns of the log-contrast matrix are

$$\mathbf{v}_i = \sqrt{\frac{i}{i+1}}\left(\frac{1}{i}, \ldots, \frac{1}{i}, -1, 0, \ldots, 0\right)^\top, \quad i = 1, \ldots, d-1.$$

The ilr coordinates have the form of balances between the ith part of the composition and all parts with lower indices

$$\text{ilr}(\mathbf{x})_i = \sqrt{\frac{i}{i+1}} \ln\left(\frac{(x_1 \cdots x_i)^{1/i}}{x_{i+1}}\right), \quad \text{for} \quad i = 1, \ldots, d-1.$$

Finally, the clr and ilr representations are mutually transferable through the contrast matrix $\mathbf{V}$

$$\text{clr}(\mathbf{x}) = \mathbf{V}\text{ilr}(\mathbf{x})$$

and also the back-transformation to the simplex is possible by using

$$\mathbf{x} = \exp(\text{clr}(\mathbf{x})) = \exp(\mathbf{V}\text{ilr}(\mathbf{x})).$$

4 ICA for Compositional Data

As described above, ICA is not reasonable for data following the Aitchison geometry in its raw form. Therefore, it is natural to transform the data first into the Euclidean space. As ICA methods start with whitening and therefore require full rank data, the ilr space is the most natural representation. Due to the affine equivariance property of the discussed ICA methods, the particular used basis for the ilr transformation at most affects the order and signs of the estimated independent components. Hence, for compositional ICA we have the following model assumption:

$$\text{ilr}(\mathbf{x}) = \mathbf{A}_{\text{ilr}}\mathbf{z} + \mathbf{b},$$

where $\mathbf{A}_{\text{ilr}}$ is a $(d-1) \times (d-1)$ full rank mixing matrix specific for a chosen ilr basis, $\mathbf{b}$ a $d-1$-dimensional location vector, and $\mathbf{z} = (z_1, \ldots, z_{d-1})^\top$ a random vector with independent components, which are standardized so that $\mathbf{E}(\mathbf{z}) = \mathbf{0}$ and $\mathbf{COV}(\mathbf{z}) = \mathbf{I}_{d-1}$. When the unmixing matrix $\mathbf{W}_{\text{ilr}}$ is estimated using one of the ICA methods described in Sect. 2, the system of independent components is given by

$$\mathbf{z} = \mathbf{W}_{\text{ilr}}(\text{ilr}(\mathbf{x}) - \mathbf{b}) = \mathbf{W}_{\text{ilr}}(\mathbf{V}^\top \text{clr}(\mathbf{x}) - \mathbf{b}).$$

As ilr coordinates are not directly related to the dominance of the original parts within the considered composition, the relationship between ilr and clr spaces can be exploited yielding a $(d - 1) \times d$ "clr" loading matrix $\mathbf{W}_{\mathrm{clr}} = \mathbf{W}_{\mathrm{ilr}} \mathbf{V}^{\top}$, allowing interpretation of the independent components in the clr space. In the context of principal component analysis performed in the clr space, principal components lead to a new system of ilr coordinates (Pawlowsky-Glahn et al. 2011). This is not the case for ICA, as the unmixing matrix $\mathbf{W}_{\mathrm{ilr}}$ (and consequently also $\mathbf{W}_{\mathrm{clr}}$) is generally not restricted to be orthogonal. Even if the independent component model does not hold, ICA transformations remain affine equivariant which means that $\mathbf{z}$ can be seen as an intrinsic data representation with a coordinate system, whose components are as independent as possible.

After performing ICA, one is usually interested in either using $\mathbf{z}$ itself for further analysis, such as classification and outlier identification, with possible interpretation in ilr or clr space using the former defined loading matrices $\mathbf{W}_{\mathrm{ilr}}$ or $\mathbf{W}_{\mathrm{clr}}$, or, using ICA for noise or artifact removal. For that purpose, the components of $\mathbf{z}$ are divided into a signal part $\mathbf{z}_s$ and a noise/artifact part $\mathbf{z}_n$. This defines also the partition of the unmixing matrix $\mathbf{W}_{\mathrm{ilr}}$ into $\mathbf{W}_{\mathrm{ilr}}^s$ and $\mathbf{W}_{\mathrm{ilr}}^n$ and the mixing matrix $\mathbf{A}_{\mathrm{ilr}} = (\mathbf{W}_{\mathrm{ilr}})^{-1}$ into $\mathbf{A}_{\mathrm{ilr}}^s$ and $\mathbf{A}_{\mathrm{ilr}}^n$. $\mathbf{A}_{\mathrm{ilr}}^s$ is formed only by those columns of $\mathbf{A}_{\mathrm{ilr}}$ that correspond to the signal components $\mathbf{z}_s$. The pure signal can then be restored in the ilr, clr, and original space by using

$$\mathrm{ilr}(\mathbf{x})_s = \mathbf{A}_{\mathrm{ilr}}^s \mathbf{z}_s + \mathbf{b}, \ \ \mathrm{clr}(\mathbf{x})_s = \mathbf{V}\left(\mathbf{A}_{\mathrm{ilr}}^s \mathbf{z}_s + \mathbf{b}\right), \ \text{and} \ \mathbf{x}_s = \exp\left[\mathbf{V}\left(\mathbf{A}_{\mathrm{ilr}}^s \mathbf{z}_s + \mathbf{b}\right)\right],$$

respectively.

5   A Case Study in Metabolomics

In order to demonstrate the above-described methods, the data from a neonatal screening program in the Czech Republic was analyzed. Anonymous data were obtained from a retrospective study approved by the Ethics Committee of the University Hospital Olomouc which was part of a larger international study described in Fleischman et al. (2013). Newborn screening is a preventive program that allows for early detection of a selected spectrum of inborn metabolic diseases. At an age of 48–72 hours after birth, several drops of blood from the heel of the child were sampled on a special paper and sent for analysis to the screening laboratory. The data at hand were constituted by the metabolite profile of over 10 000 healthy newborns. For each neonate, the values of 48 metabolites were measured. Moreover, information about sex and birth weight was available. More specifically, the birth weight ranged from 300 to 5 570 grams and for newborns with very low birth weight (less than 1500 grams) a different metabolite structure can be expected, due to their prematurity and the artificial nutrition they receive. One of the main goals of metabolomics is to investigate interactions between metabolites, their dynamic changes, and responses to stimuli. Biofluids, e.g. blood or urine, and also tissues are used for the analysis. On the one
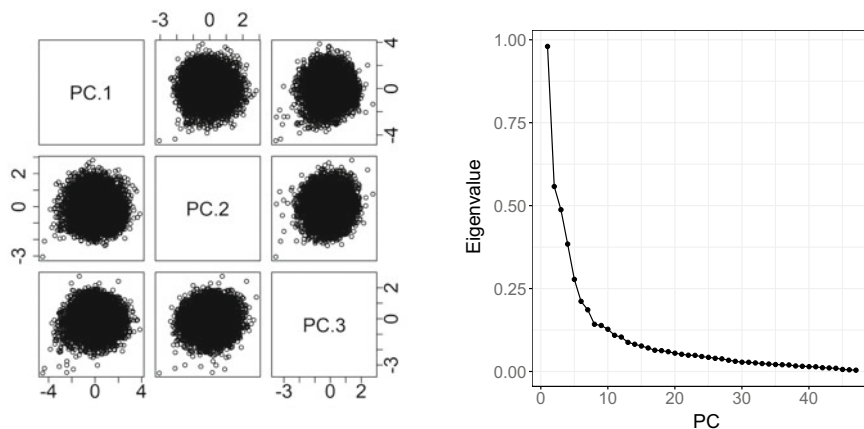
Fig. 1 Scatterplots of the first three principal components resulting from the compositional PCA (left) and scree plot of the respectively explained variability (right)

hand, the most frequently used approach for the data analysis is done through comparison of absolute values of biomarkers and reference ranges (data from the healthy population). On the other hand, the new trend of data evaluation is based on the use of ratios of metabolite data. Relative changes are more relevant/informative than absolute values in diagnostics based on profiling. Therefore, metabolomic data can be considered as observations carrying relative information, i.e. as compositional data (Kalivodová et al. 2018), and as such the above-discussed methods can be applied.

The following analysis was carried out in R 3.6.1 (R Core Team 2019) with the help of the packages JADE (Miettinen et al. 2017), fICA (Miettinen et al. 2018), compositions (van den Boogaart et al. 2019), and robCompositions (Templ et al. 2011). As the first step, standard principal component analysis (PCA) was performed on the clr transformed data. There were no significant patterns visible within the first three principal components; see Fig. 1, left. The whole dataset forms one quite compact cluster with no outliers. Moreover, the variance explained by the first components is low (around 20 % for the first PC) (Fig. 1, right), and therefore PCA does not seem to deal well with the issue of outlier detection, grouping, as well as dimension reduction in that case.

As PCA seems not to reveal any clear structure, we applied FOBI, k-JADE, with $k = 5$, and adaptive deflation-based FastICA to the ilr representation of the data (the dimension $p = 47$ was already too large for JADE). For easier comparison, the components from all three ICA methods were ordered according to their kurtosis values. As all three ICA methods showed similar results, we focus on our presentation and discussion of the components on those from adaptive deflation-based FastICA.

Due to the kurtosis ordering, the first components show heavy-tailed distributions, and they are expected to find outliers or small groupings, while the last components show light-tailed distributions and hence might find more balanced groupings. Scores of the first and last three independent components are plotted in Fig. 2, and the chosen
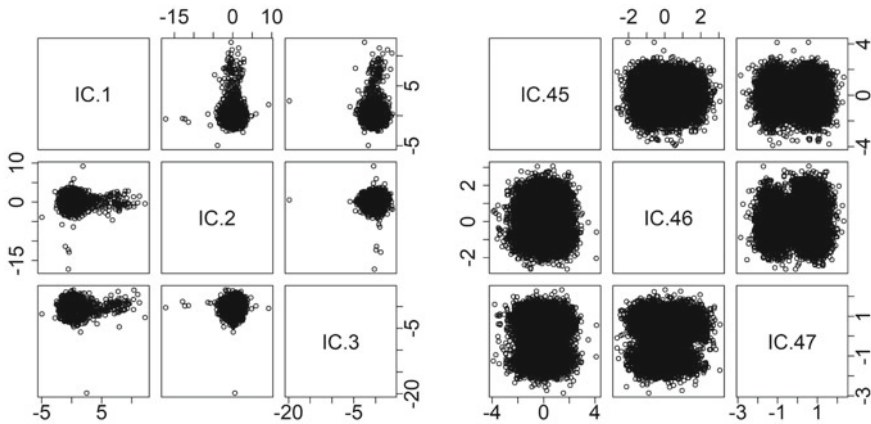
Fig. 2 Scatterplots of the first (left) and last (right) three independent components resulting from the compositional FastICA, using adaptive deflation-based FastICA

non-linearities are given in Table 2 for all independent components. According to the left plot of Fig. 2, one outlier is clearly detected due to its high negative value in the third component (IC.3). According to its loadings, which are collected in Table 3, IC.3 mostly reflects the relative dominance (with respect to concentrations of all 48 measured metabolites) of phenylalanine (Phe), hexadecanoylcarnitine (C16), octade-cenoylcarnitine (C18:1), valine (Val), and hexadecenoyl- and octadecanoylcarnitines in the form of C16:1 and C18, respectively, when the higher dominance of the first three metabolites results in a decrease of IC.3 and vice versa for the last three stated metabolites. The high loadings of the clr coefficients of these six metabolites imply that IC.3 reflects mostly (but not solely) the balance between subcompositions formed by Phe, C16, C18:1, and Val, C16:1, C18. The value of this balance was for the outlier significantly lower than that within the rest of the sample. After a deeper investigation of the outlying sample, it turned out that it belongs to a newborn suffering from Phenylketonuria, a metabolic disease which is typically followed by distinctly high absolute blood concentrations of phenylalanine. The measured value was 1 014.7 $\mu mol/l$, which significantly exceeds the upper norm value set on 120 $\mu mol/l$ (van Wegberg et al. 2017) and which is represented with the respective high clr value 6.76. The levels of the remaining metabolites were comparable with the other samples, but particularly the atypical high dominance of Phe over all measured metabolites, which for the rest of samples ranged from 5.72 to 3.58 for their clr values, resulted in the high negative value of the third component, and therefore clear identification of this non-standard observation.

The next interesting feature is presented by IC.1. According to Fig. 2, the values of this component are not very homogeneous across the whole dataset and therefore some specific groups of neonates might be identified. A deeper graphical analysis of the first component (presented in Fig. 3) shows that for newborns with a birth weight smaller than 1500 grams, higher values of IC.1 are typical. The independent

Table 2 Chosen non-linearities g_i for each independent component computed with the adaptive deflation-based FastICA algorithm. Non-linearities are ordered according to kurtosis values of the corresponding ICs. In the original ordering, IC.44 was the last component, thus no non-linearity is given. See Table 1 for the definitions of the functions g_i

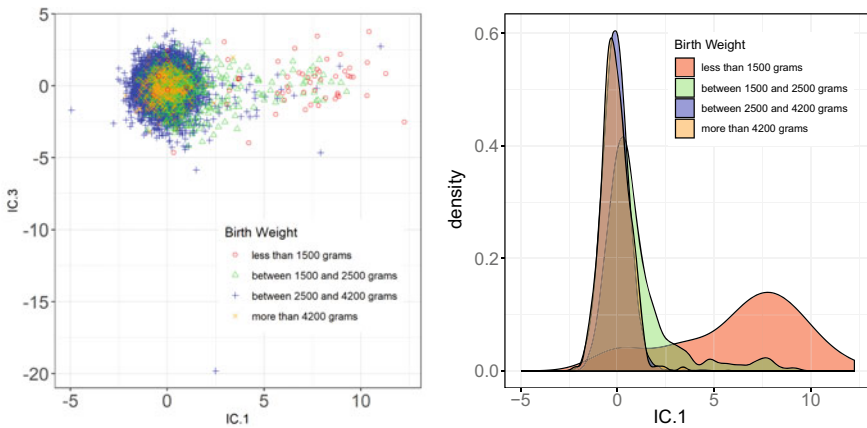| IC	g_i	IC	g_i	IC	g_i	IC	g_i	IC	g_i
IC.1	g_2	IC.11	g_2	IC.21	g_5	IC.31	g_1	IC.41	g_5
IC.2	g_2	IC.12	g_9	IC.22	g_6	IC.32	g_5	IC.42	g_4
IC.3	g_6	IC.13	g_9	IC.23	g_9	IC.33	g_{14}	IC.43	g_4
IC.4	g_2	IC.14	g_8	IC.24	g_1	IC.34	g_5	IC.44	–
IC.5	g_6	IC.15	g_6	IC.25	g_1	IC.35	g_5	IC.45	g_5
IC.6	g_8	IC.16	g_{10}	IC.26	g_4	IC.36	g_5	IC.46	g_3
IC.7	g_8	IC.17	g_5	IC.27	g_{14}	IC.37	g_{12}	IC.47	g_6
IC.8	g_8	IC.18	g_6	IC.28	g_5	IC.38	g_{14}		
IC.9	g_9	IC.19	g_{10}	IC.29	g_{10}	IC.39	g_4		
IC.10	g_8	IC.20	g_6	IC.30	g_8	IC.40	g_{11}		

Fig. 3 Scatterplots of IC.1 and IC.3 (left) and the kernel density plot of IC.3 (right) with the groups defined according to the birth weight

component IC.1 is mostly formed by clr values of acylcarnitines dodecanoylcarnitine (C12), C16, and C18:1, whose high relative dominance over all measured metabolites results in low values of the component and, e.g. clr values of acylcarnitines isovalerylcarnitine/methylbutyrylcarnitine C5, and linoleoylcarnitine (C18:2) increase the IC.1 values. Even though there are also other metabolites contributing with a high weight to the values of IC.1 (all clr loadings are collected in Table 3), the clr values of the selected ones systematically differ for the group of the newborns with low birth weight, and therefore these acylcarnitines seem to be responsible for their separation from the remaining neonates. The differences in the selected metabolites are clearly visible in Fig. 4. Let us stress here that the immature neonates tend

to have different diet supplementation, therefore the metabolic profile can substantially differ within this group, but despite the proposed ICA method being able to find some similar patterns, detect the important metabolites, and separate the low birth weight newborns from the remaining ones. More specifically, artificial nutrition consists of amino acids, lipids, sugars, vitamins, etc. Essential unsaturated fatty acids including linoleic acid may be responsible for increased C18:2. The increased blood concentration of the long-chain acylcarnitines (C12, 16, C18:1) as well as of the short-chain C5 carnitine, which then results in high respective clr values, corresponds with previous studies. In Gucciardi et al. (2015), the significantly lower amounts of acylcarnitines except the branched-chain acylcarnitines (e.g. C5), which were significantly higher in preterm infants, were described. The latter mentioned are direct products of branched-chain amino acid (BCAA) catabolism, therefore its elevated levels may be related to BCAA overfeeding (Gucciardi et al. 2015; Wilson et al. 2014). The difference of several amino acids measured for the premature newborns compared to the others agrees with findings in Wilson et al. (2014), where increased levels of several amino acids (arginine, leucine, Orn, Phe, and Val) in the blood spots of premature infants were described. This observation may be related to the catabolic state of organisms in these children, amino acid supplementation, and immaturity of preterm infants (hepatic maturation, renal insufficiency, etc.) (Wilson et al. 2014; te Braake et al. 2005). The raw concentrations of valine (Val) and leucine/isoleucine (Xle) are known to be highly positively correlated, therefore the opposite signs of the respective loadings of IC.1 seem to be counter-intuitive at the first glance. However, the values of the loadings suggest that the resulting value of IC.1 is affected by the difference of clr values of the respective metabolites, or equivalently by the log-ratio of their measured concentrations, when the higher relative dominance of Val over Xle results in a higher value of IC.1. These findings agree with the data, since slightly higher values of the Val-Xle log-ratio are typical for newborns with a low birth weight (see Fig. 4). Finally, an even more complex interpretation can be based on the ilr loading matrix $\mathbf{W}_{ilr}$. According to the values of this matrix, IC.1 is mainly influenced by the balance between C18 and subcompositions C18:1, C18:OH, C18:2, and C18:2OH. This balance corresponds to the highest positive loading, and its values are systematically higher for the group of newborns with low birth weight than for the rest of the samples.

An even better visible pattern is formed by the last independent component IC.47, which clearly divides the whole dataset into two groups as seen in Fig. 5. According to the loadings (collected in Table 3), the most contributing are clr values of metabolites Xle, ornithine (Orn), and lysine (Lys) with a negative effect and methionine (Met), proline (Pro), and valine (Val) with a positive one. This suggests that the value of IC.47 is highly affected by the balance between subcompositions Met, Pro, Val and Xle, Orn, Lys. The dataset is roughly separated into two groups of observations with values of IC.47 higher and lower than -0.34; this value was chosen as the corresponding value of IC.47 at the local minimum in the middle of the density presented in Fig. 5 (this density was computed with Gaussian kernels and a bandwidth selection with Silverman's rule of thumb). The relative dominance of the six above-mentioned metabolites itself over all measured concentrations does not significantly differ in its
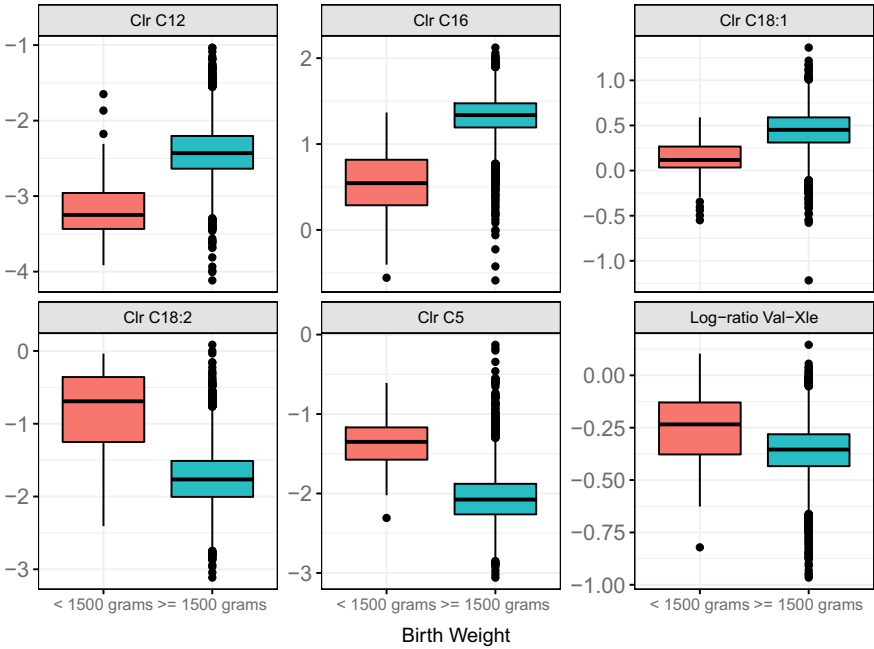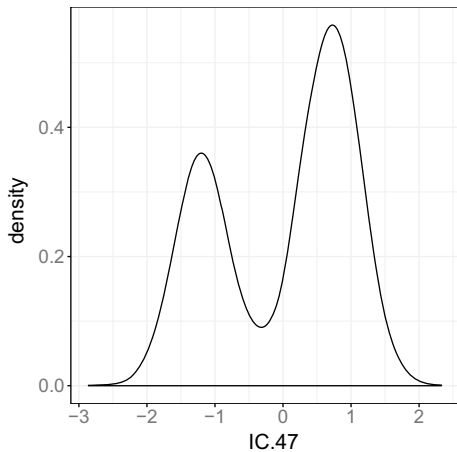
Fig. 4 Boxplots of clr as well as log-ratio values of the selected metabolites, which significantly differ for newborns with very low ($< 1500\,\text{g}$) and normal ($>= 1500\,\text{g}$) weight at birth

Fig. 5 Density plot of IC.47, the bimodal shape shows a clear grouping

values between the two groups. Therefore, the grouping effect of IC.47 is hidden in some of their more complex combinations, e.g. the suggested balance between Met, Pro, Val and Xle, Orn, Lys, which is distinctly higher by cases with IC.47 higher than -0.34.

Table 3 The list of loadings for IC.1, IC.3, and IC.47 computed with the adaptive deflation-based FastICA algorithm regarding clr transformed data

	IC.1	IC.3	IC.47		IC.1	IC.3	IC.47
Ala	0.02	0.24	0.41	C5DC/C6OH	−0.08	0.14	0.04
Arg	0.13	−0.33	0.04	C5:1	−0.05	0.11	0.13
ArgSucc	0.00	−0.09	−0.01	C6	−0.22	0.22	0.07
Cit	−0.05	0.38	0.01	C8	0.84	−0.05	0.21
Glu	−0.16	0.70	0.20	C8:1	−0.05	−0.16	0.09
Gly	0.08	1.33	−0.16	C10	0.25	−0.21	−0.09
His	−0.56	0.44	0.55	C10:1	−1.05	−0.01	−0.21
Lys	0.20	0.38	−0.46	C10:2	0.28	0.20	0.06
Met	0.53	1.03	1.01	C12	−1.27	−0.01	−0.13
Orn	0.48	0.40	−0.78	C12:1	−0.09	0.10	−0.09
Phe	1.02	−7.30	−0.40	C14	−0.39	0.45	−0.38
Pro	−1.15	−0.66	0.65	C14:1	0.88	−0.46	0.04
Thr	−0.17	−1.71	−0.01	C14:2	0.15	−0.01	0.17
Trp	−1.33	−0.77	0.14	C14OH	0.03	0.27	0.08
Tyr	−0.28	1.25	−0.02	C16	−2.00	−1.49	−0.02
Val	3.09	3.86	0.59	C16:1	0.97	1.60	0.09
Xle	−1.73	−0.15	−1.03	C16OH	0.24	0.09	0.01
C0	0.51	0.60	0.17	C16:1OH	0.05	0.01	0.24
C2	0.08	−0.30	−0.25	C18	2.38	1.72	0.12
C3	−0.57	−0.34	−0.01	C18:1	−3.46	−2.64	0.00
C3DC/C4OH	0.29	−0.23	0.34	C18:2	1.80	0.41	0.16
C4	0.05	0.14	0.01	C18:1OH	−0.00	0.27	−0.15
C4DC/C5OH	0.07	0.49	−1.42	C18:2OH	0.19	0.23	−0.05
C5	0.35	−0.32	−0.08	C18OH	−0.33	0.20	0.09

6 Discussion

In this paper, we reviewed some classical independent component analysis methods and showed how these can be applied to compositional data. The key finding here is that when the ICA methods are affine equivariant it is most natural to use an ilr transformation, as the choice of the basis constituting the ilr coordinate system does not matter. For interpretability, the link between ilr coordinates and clr coefficients/variables can be easily exploited, which allows interpreting the results either in terms of the dominance of single compositional parts with respect to the whole composition, or, e.g. based on values of balances between subcompositions formed according to values of clr loadings. Finally, since the clr loadings are derived from the ilr ones, it is also possible to provide the interpretation directly in terms of the ilr coordinates. The proposed technique is demonstrated on a metabolomics dataset

where PCA, which is probably the most used multivariate transformation, reveals no specific feature on the first few components while ICA reveals several interesting features visible when exploiting the higher order moments information. Independent component analysis belongs to the larger class of blind source separation methods where for the separation of the latent components often also temporal or spatial information is used. In the context of compositional data such blind source separation methods are, for example, discussed in Nordhausen et al. (2015), Nordhausen et al. (2020). But these methods would not be applicable to the metabolomics dataset from Sect. 5 as there is no temporal or spatial information present. The current results, which were discussed mostly in terms of the relative dominance of a single compositional part respective to the highest loading of an IC, open new challenges for further research. An alternative interpretation can be reached, e.g. by adaptation the approach based on principal balances (Pawlowsky-Glahn et al. 2011). However, the loadings of ICs are in general not orthonormal and therefore the principal balances approach is not as straightforward as in the case of PCA. Finally, an extension of the dataset with a group of blood samples collected from neonates with a diagnosed disease can further prove the usefulness of the method.

References

Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman & Hall.

Cardoso, J.-F. (1989). Source separation using higher order moments. *International Conference on Acoustics, Speech, and Signal Processing*, *1989*, 2109–2112.

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

Clarkson, D. B. (1988). A least squares version of algorithm AS 211: The FG diagonalization algorithm. *Journal of the Royal Statistical Society C*, *37*, 317–321.

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

Egozcue, J. J., & Pawlowsky-Glahn, V. (2005). Groups of parts and their balances. *Mathematical Geology*, *37*, 795–828.

Egozcue, J. J., & Pawlowsky-Glahn, V. (2019). Compositional data: The sample space and its structure. *Test*, *28*, 599–638.

Fačevicová, K., Bábek, O., Hron, K., & Kumpan, T. (2016). Element chemostratigraphy of the devonian/carboniferous boundary - a compositional approach. *Applied Geochemistry*, *75*, 211–221.

Filzmoser, P., Hron, K., & Reimann, C. (2009). Principal component analysis for compositional data with outliers. *Environmetrics*, *20*, 621–632.

Filzmoser, P., Hron, K., & Templ, M. (2018). *Applied compositional data analysis*. Cham: Springer.

Fleischman, A., Thompson, J. D., & Glass, M. (2013). Systematic data collection to inform policy decisions: Integration of the region 4 stork (r4s) collaborative newborn screening database to improve ms/ms newborn screening in Washington State. In J. Zschocke, K. M. Gibson, G. Brown, E. Morava, & V. Peters (Eds.), *JIMD reports - case and research reports* (Vol. 13, pp. 15–21). Berlin: Springer.

Gucciardi, A., Zaramella, P., Costa, I., Pirillo, P., Nardo, D., Naturale, M., et al. (2015). Analysis and interpretation of acylcarnitine profiles in dried blood spot and plasma of preterm and full-term newborns. *Pediatric Research*, *77*(1–1), 36–47.

Hron, K., Filzmoser, P., & Thompson, K. (2012). Linear regression with compositional explanatory variables. *Journal of Applied Statistics*, *39*(5), 1115–1128.

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

Hyvärinen, A. (1999). Gaussian moments for noisy independent component analysis. *IEEE Signal Processing Letters*, *6*, 145–147.

Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, *9*, 1483–1492.

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

Kalivodová, A., Hron, K., Filzmoser, P., Najdekr, L., Janečková, H., & Adam, T. (2018). PLS-DA for compositional data with application to metabolomics. *Journal of Chemometrics*, *29*, 21–28.

Miettinen, J., Nordhausen, K., Oja, H., & Taskinen, S. (2013). Fast equivariant JADE. In *IEEE international conference on acoustics, speech and signal processing (ICASSP) 2013* (pp. 6153–6157).

Miettinen, J., Nordhausen, K., Oja, H., & Taskinen, S. (2014). Deflation-based FastICA with adaptive choices of nonlinearities. *IEEE Transactions on Signal Processing*, *62*, 5716–5724.

Miettinen, J., Nordhausen, K., Oja, H., Taskinen, S., & Virta, J. (2017). The squared symmetric FastICA estimator. *Signal Processing*, *131*, 402–411.

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., Nordhausen, K., & Taskinen, S. (2018). fICA: FastICA algorithms and their improved variants. *The R Journal*, *10*, 148–158.

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

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, *47*, 1–25.

Nordhausen, K., Fischer, G., & Filzmoser, P. (2020). Blind source separation for compositional time series. *To appear in Mathematical Geosciences* (pp. 1–21).

Nordhausen, K., Ilmonen, P., Mandal, A., Oja, H., Ollila, E.: Deflation-based FastICA reloaded. In *Proceedings of 19th European signal processing conference* (pp. 1854–1858).

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

Nordhausen, K., Oja, H., Filzmoser, P., & Reimann, C. (2015). Blind source separation for spatial compositional data. *Mathematical Geosciences*, *47*, 753–770.

Nordhausen, K., & Virta, J. (2019). An overview of properties and extensions of FOBI. *Knowledge-Based Systems*, *173*, 113–116.

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

Pawlowsky-Glahn, V., & Buccianti, A. (Eds.). (2011). *Compositional Data Analysis, Theory and Applications*. Chichester: Wiley.

Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R. (2011). Principal balances. In J.J. Egozcue, R. Tolosana-Delgado, & MI. Ortego (Eds.), *Proceedings of the 4th International Workshop on Compositional Data Analysis*.

Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (Eds.). (2015). *Modelling and analysis of compositional data*. Chichester: Wiley.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Te Braake, F. W., van den Akker, C. H., Wattimena, D. J., Huijmans, J. G., van Goudoever, J. B. (2005). Amino acid administration to premature infants directly after birth. *The Journal of Pediatrics*, *147*, 457–461.

Templ, M., Hron, K., & Filzmoser, P. (2011). robCompositions: an R-package for robust statistical analysis of compositional data. In *Compositional data analysis: Theory and applications* (pp. 341–355). New York: Wiley.

Trinh, H. T., Morais, J., Thomas-Agnan, C., & Simioni, M. (2019). Relations between socio-economic factors and nutritional diet in Vietnam from 2004 to 2014: New insights using compositional data analysis. *Statistical Methods in Medical Research*, *28*, 2305–2325.

Tyler, D., Critchley, F., Dümbgen, L., & Oja, H. (2009). Invariant coordinate selection. *Journal of Royal Statistical Society B*, *71*, 549–592.

van den Boogaart, K. G., Tolosana-Delgado, R., & Bren, M. (2019). *Compositions: Compositional data analysis*. R package version 1.40-3.

van Wegberg, A., MacDonald, A., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A. M., et al. (2017). The complete European guidelines on phenylketonuria: Diagnosis and treatment. *Orphanet Journal of Rare Diseases*, *12*, 162.

Virta, J., Lietzen, N., Ilmonen, P., Nordhausen, K. (2020). Fast tensorial JADE. *To appear in Scandinavian Journal of Statistics*.

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

Wilson, K., Hawken, S., Ducharme, R., Potter, B. K., Little, J., Thébaud, B., et al. (2014). Metabolomics of prematurity: Analysis of patterns of amino acids, enzymes, and endocrine markers by categories of gestational age. *Pediatric Research*, *75*, 367–373.