Universiteit Leiden
The Netherlands

Clusterwise Independent Component Analysis (C-ICA) using fMRI resting state networks to cluster subjects and find neurofunctional subtypes

Durieux, J.; Rombouts, S.A.R.B.; Vos, F. de; Koini, M.; Wilderjans, T.F.

Citation

Durieux, J., Rombouts, S. A. R. B., Vos, F. de, Koini, M., & Wilderjans, T. F. (2022). Clusterwise Independent Component Analysis (C-ICA): using fMRI resting state networks to cluster subjects and find neurofunctional subtypes. *Journal Of Neuroscience Methods*, *382*. doi:10.1016/j.jneumeth.2022.109718

Version:	Publisher's Version
License:	Creative Commons CC BY 4.0 license
Downloaded from:	https://hdl.handle.net/1887/3567671

Note: To cite this publication please use the final published version (if applicable).

Journal of Neuroscience Methods 382 (2022) 109718

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Clusterwise Independent Component Analysis (C-ICA): Using fMRI resting state networks to cluster subjects and find neurofunctional subtypes

Jeffrey Durieux [a,b,c,*], Serge A.R.B. Rombouts [a,b,d], Frank de Vos [a,b,d], Marisa Koini [e], Tom F. Wilderjans [a,b,f,g]

[a] Methodology and Statistics Unit, Institute of Psychology, Faculty of Social and Behavioural Sciences, Leiden University, Leiden, The Netherlands
[b] Leiden Institute for Brain and Cognition (LIBC), Leiden, The Netherlands
[c] Econometric Institute, Erasmus University Rotterdam, The Netherlands
[d] Department of Radiology, Leiden University Medical Center, The Netherlands
[e] Department of Neurology, Medical University of Graz, Austria
[f] Research Group of Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium
[g] Department of Clinical Psychology, Vrije Universiteit Amsterdam, Netherlands

ARTICLE INFO

Keywords:
ICA
Clustering
Resting-state fMRI
Patient clusters
Heterogeneity
Individual differences
Neurofunctional subtypes

ABSTRACT

Background: FMRI resting state networks (RSNs) are used to characterize brain disorders. They also show extensive heterogeneity across patients. Identifying systematic differences between RSNs in patients, i.e. discovering neurofunctional subtypes, may further increase our understanding of disease heterogeneity. Currently, no methodology is available to estimate neurofunctional subtypes and their associated RSNs simultaneously.

New method: We present an unsupervised learning method for fMRI data, called Clusterwise Independent Component Analysis (C-ICA). This enables the clustering of patients into neurofunctional subtypes based on differences in shared ICA-derived RSNs. The parameters are estimated simultaneously, which leads to an improved estimation of subtypes and their associated RSNs.

Results: In five simulation studies, the C-ICA model is successfully validated using both artificially and realistically simulated data (N = 30–40). The successful performance of the C-ICA model is also illustrated on an empirical data set consisting of Alzheimer's disease patients and elderly control subjects (N = 250). C-ICA is able to uncover a meaningful clustering that partially matches (balanced accuracy = .72) the diagnostic labels and identifies differences in RSNs between the Alzheimer and control cluster.

Comparison with other methods: Both in the simulation study and the empirical application, C-ICA yields better results compared to competing clustering methods (i.e., a two step clustering procedure based on single subject ICA's and a Group ICA plus dual regression variant thereof) that do not simultaneously estimate a clustering and associated RSNs. Indeed, the overall mean adjusted Rand Index, a measure for cluster recovery, equals 0.65 for C-ICA and ranges from 0.27 to 0.46 for competing methods.

Conclusions: The successful performance of C-ICA indicates that it is a promising method to extract neurofunctional subtypes from multi-subject resting state-fMRI data. This method can be applied on fMRI scans of patient groups to study (neurofunctional) subtypes, which may eventually further increase understanding of disease heterogeneity.

1. Introduction

An important and emerging challenge in the field of clinical neuroscience pertains to revealing differences and similarities in resting state functional connectivity networks (RSNs) between patients. Heterogeneity in patients' RSNs suggests the presence of neurofunctional subtypes and these neurofunctional subtypes may, in turn, be related to clinically determined disease subtypes (e.g., types of dementia, see Gili et al., 2011). Moreover, disclosing these functional subtypes allows for a categorization of diseases based on brain information that may complement the commonly used categorizations based on

* Corresponding author at: Methodology and Statistics Unit, Institute of Psychology, Faculty of Social and Behavioural Sciences, Leiden University, Leiden, The Netherlands.
E-mail address: j.durieux@fsw.leidenuniv.nl (J. Durieux).

https://doi.org/10.1016/j.jneumeth.2022.109718
Received 2 February 2022; Received in revised form 18 August 2022; Accepted 29 September 2022
Available online 6 October 2022
0165-0270/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

clinical symptoms (e.g., the Diagnostic and Statistical Manual of Mental Disorders; American Psychiatric Association, 2013).

RSNs are related to areas in the brain that are involved in cognitive processes such as, for example, executive functioning, auditory processing and memory (Beckmann et al., 2005). Disruptions in the integrity of RSNs are associated with neurodegenerative diseases and psychiatric disorders such as Parkinson's disease (Olde Dubbelink et al., 2013), Major Depression Disorder (Greicius et al., 2007), schizophrenia (Lynall et al., 2010) and Alzheimer's disease (AD) (Gili et al., 2011). For example, AD patients have disrupted RSNs when compared to healthy control subjects (Greicius et al., 2004; Zhang et al., 2010). Moreover, differences in RSNs are also implicated for different clinical subtypes of dementia, like early onset AD and late onset AD (Gour et al., 2014). These disruptions in RSNs integrity often precede grey matter atrophy, which is a key aspect of AD (Gili et al., 2011). This implies that such RSNs differences are very relevant for the early detection of diseases and their clinical subtypes.

A frequently used technique to disclose the most apparent RSNs underlying resting state fMRI (rs-fMRI) data is (spatial) Independent Component Analysis (ICA; see Hyvärinen et al., 2001; Beckmann and Smith, 2004; Van de Ven et al., 2004). ICA is a decomposition technique that separates a multivariate signal (e.g., fMRI data) into spatial components that are statistically independent and a linear mixing matrix. These spatially independent components, also denoted as RSNs or spatial maps, are sets of voxels that are functionally connected, whereas the mixing matrix contains information regarding the underlying time course for each independent RSN (a statistical description is presented in Appendix A).

A promising way to automatically uncover neurofunctional subtypes from multi-subject rs-fMRI data consists of clustering patients based on ICA derived shared RSNs (for a related example, see Drysdale et al., 2017). The goal of this paper is to introduce a novel unsupervised clustering method for multi-subject rs-fMRI data that combines a clustering technique with ICA into a single analysis method. Note that the combination of clustering with a decomposition method has already been successfully developed for and applied in other research areas such as psychology (De Roover et al., 2012b, 2017; Wilderjans et al., 2017), sensory profiling (Wilderjans and Ceulemans, 2013; Wilderjans and Cariou, 2016; Cariou and Wilderjans, 2018) and marketing (Bougeard et al., 2018).

The proposed method in this paper, which will be called Clusterwise Independent Component Analysis (C-ICA), clusters the subjects into homogeneous groups (i.e., neurofunctional subtypes) based on the similarities and differences in the (shared) RSNs underlying each subject's data. In particular, subjects with similar shared RSNs will be automatically clustered together, whereas subjects exhibiting RSNs that are qualitatively different will be allocated to different clusters. Note that the goal is to cluster (in an unsupervised way) *patients* and not voxels, brain regions or RSNs directly (as in ICASSO, see Himberg et al., 2004).

The remainder of this paper is organized as follows: in the next section, the mathematical formulation of the novel C-ICA model will be introduced and relationships to other (clustering) methods for multi-subject data, such as Group-ICA (Guo and Pagnoni, 2008; Calhoun et al., 2009), will be elaborated upon. Next, in the Data Analysis section, an appropriate Alternating Least Squares (ALS) algorithm to estimate the parameters of the C-ICA model will be presented, along with a procedure to select the optimal number of clusters and ICA RSNs. In the fourth section, the C-ICA model is validated in a series of studies. More specifically, the performance of the C-ICA algorithm is evaluated in four simulation studies that differ in terms of how realistic the rs-fMRI data are simulated. Additionally, comparisons to existing strategies for clustering patients into functional subtypes (e.g., a clustering procedure based on Group ICA plus dual regression) are presented. In a fifth study, the performance of the proposed procedure for model selection will be assessed. In the fifth section, an empirical data set consisting of Alzheimer's disease patients and elderly controls will be analysed to illustrate the C-ICA model. Also here, C-ICA's performance will be compared to other clustering procedures. Finally, implications of the C-ICA model and directions for future research will be discussed.

2. Clusterwise independent component analysis (C-ICA)

2.1. Model formulation of C-ICA

Starting from multi-subject (rs-)fMRI data, which are illustrated in Fig. 1, in order to cluster subjects based on similarities and differences in underlying RSNs, the C-ICA generative model assumes that data of I subjects $\mathbf{X}_i$ $(i = 1, \ldots, I)$ fall apart into R mutually exclusive clusters (i.e., a hard partitioning), which are unknown a priori and thus have to be estimated from the data, with each $\mathbf{X}_i$ containing the (rs-)fMRI data (T time points $\times$ V voxels) of subject i. Thus, as presented in Fig. 1, C-ICA decomposes each $\mathbf{X}_i$ as:[1]

$$\mathbf{X}_i = \sum_{r=1}^{R} p_{ir}\mathbf{A}_i\mathbf{S}^r + \mathbf{E}_i \tag{1}$$

where the elements p_{ir} denote the entries from the binary partition matrix $\mathbf{P}$ $(I \times R)$ which equal 1 when subject i is assigned to cluster C_r and 0 otherwise. Note that the p_{ir} elements are parameters of the C-ICA generative model that need to be estimated from the data, which implies that the clustering of the subjects is part of the C-ICA generative model. Similar as in the ICA model for a single subject (see Appendix A), $\mathbf{A}_i$ ($T \times Q$, with Q being the number of RSNs, which is kept equal across subjects and clusters) denotes the mixing matrix (i.e., time courses) for subject i and $\mathbf{S}^r$ $(Q \times V)$ represents a matrix where the rows contain the spatially independent components or RSNs for cluster C_r $(r = 1, \ldots, R)$. Note that the cluster specific RSNs in $\mathbf{S}^r$ are shared RSNs that represent the data $\mathbf{X}_i$ for all subjects i belonging to the cluster in question.[2] Hence, the C-ICA generative model assumes that only the shared RSNs for each cluster contain information that is relevant to cluster the subjects. Q thus refers to the number of *shared* RSNs that contribute to the subject clustering. This also (implicitly) means that idiosyncratic or subject-specific RSNs do not contribute to the subject clustering and thus are considered as obfuscating noise (i.e., C-ICA is not designed to extract these subject-specific RSNs). Moreover, in accordance to ICA for a single subject, the RSNs in $\mathbf{S}^r$ are assumed to be non-normally distributed and mutually statistical (spatially) independent. It is important to note that in the C-ICA generative model, the time courses $\mathbf{A}_i$ are allowed to differ for each subject but that the data from all subjects that belong to the same cluster can be described in terms of a set of RSNs $\mathbf{S}^r$. Additionally, the generative model contains an error term $\mathbf{E}_i$ for each data block i (Stegeman and Mooijaart, 2008). Note that the C-ICA generative model is defined for a given value of the number of clusters R and number of components Q. The C-ICA algorithm, that will be described in Section 3.2, will estimate the C-ICA parameters for that given value of R and Q. Determining the optimal value of R and Q, which is discussed in Section 3.3, is a model selection problem that constitutes a separate step in the analysis and thus is not part of the C-ICA algorithm. As a final note, C-ICA selecting for a single cluster ($R = 1$) is equivalent to performing Group ICA (Calhoun et al., 2001b), whereas C-ICA with all clusters being singletons ($R = I$), implying that each subject is allocated to a separate cluster, yields ICA for each subject separately (also see Section 2.2). As such, C-ICA (with $1 < R < I$) takes an intermediate position on a continuum with Group

[1] The C-ICA generative model can also be formulated as $\mathbf{X}_i = \mathbf{A}_i \sum_{r=1}^{R} p_{ir}\mathbf{S}^r + \mathbf{E}_i = \mathbf{A}_i\mathbf{S}_i^r + \mathbf{E}_i$ for $i \in C_r$ with C_r denoting the rth cluster (which is the cluster to which subject i belongs).

[2] As for Group ICA, also for C-ICA, a back-reconstruction method can be applied to project the cluster specific RSNs to the data of each subject separately.

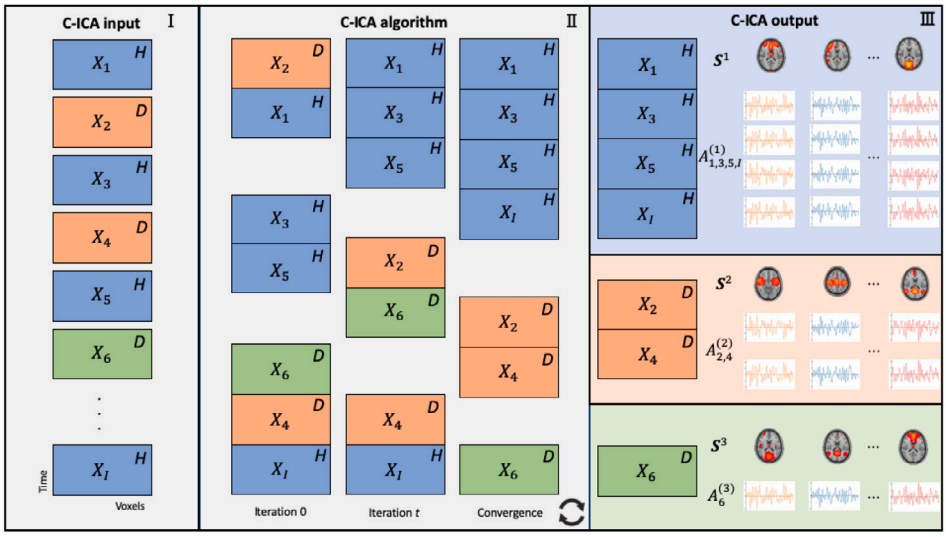

Fig. 1. Graphical illustration of Clusterwise Independent Component Analysis (C-ICA). In the left panel (I), Time × Voxel matrices of multiple subjects i ($i = 1 \dots I$) are presented. The different colours refer to the true, albeit unknown a priori, clusters which are based on differences in shared RSNs. Note that these cluster do not necessarily correspond with the known diagnostic labels (indicated by D = diseased and H = healthy) which are based on disease symptoms. In panel II, a depiction of the C-ICA algorithm (see Section 3.2) is given. During iteration 0, separately for each cluster of a (randomly) given partition, the Time × Voxel matrices are concatenated together and ICA is performed. Next, the data matrices are re-grouped by determining which data matrix optimally belongs to which cluster and again, for each cluster separately, ICA is performed on the concatenated data matrices. This procedure of re-grouping and re-estimating cluster specific ICA parameters is repeated until convergence. After convergence, as shown in panel III, subjects are allocated to their optimal cluster and each cluster r is described by a cluster specific set of RSNs $\mathbf{S}^r$ and subject specific time courses $\mathbf{A}_i$, again by performing ICA on concatenated data.

ICA ($R = 1$) and subject specific ICA's ($R = I$) as extremes, and – when comparing the cluster specific $\mathbf{S}^r$'s to each other – allows for describing subject similarities and differences based on their RSNs in a parsimonious way.

The C-ICA model suffers from four sources of non-uniqueness or ambiguities; scaling, reflection, component/RSN permutation and cluster permutation non-uniqueness (see Appendix B). Here, the first three sources of non-uniqueness also hold for ICA of a single subject (Hyvärinen et al., 2001, p.154), whereas cluster permutation non-uniqueness specifically applies to C-ICA. Cluster permutation non-uniqueness implies that cluster labels cannot be uniquely defined as the order of the clusters is arbitrary. Note that this source of non-uniqueness – also known as the label switching problem – holds for almost all existing clustering procedures.

2.2. Relationships to other (clustering) models for multi-subject data

C-ICA bears interesting relationships to other analysis methods for multi-subject data. The analysis methods most closely related to C-ICA will be discussed here. These methods fall into two main categories: ICA procedures that do not involve clustering and procedures that combine clustering with ICA or a similar decomposition method. Note that methods that incorporate clustering in ICA but only can be applied to single-subject fMRI data, like Mixture ICA (Lee et al., 2000), subspace ICA (De Ridder et al., 2000) and NSMM-ICA (Zhu and Hunter, 2019), will not be discussed. Also, methods that aim at clustering voxels, brain regions or ICA components instead of subjects, like ICASSO (Himberg et al., 2004) and self-organizing Group ICA (Esposito et al., 2005), will not be considered.

2.2.1. ICA procedures without a clustering procedure

In order to extract RSNs from multi-subject rs-fMRI data (see Fig. 1), several ICA analysis strategies can be used (for an overview see, Calhoun et al., 2009). These strategies mainly differ in whether the RSNs (i.e., $\mathbf{S}$) and/or the time courses (i.e., $\mathbf{A}$) are allowed to be specific for each subject or are restricted to be representative for an a priori defined group of subjects. Three of those strategies, relevant for a theoretical comparison to C-ICA, are discussed below.

Subject specific RSNs and time courses. The first analysis strategy for multi-subject brain data consists of performing ICA to the data of each subject separately. A possible nuisance of such a strategy is that different RSNs and time courses for each subject are obtained, which necessitates a computationally intensive post-hoc matching of RSNs across subjects in order to identify systematic differences and similarities in shared RSNs between subjects. Indeed, obtaining a one-to-one matching of RSNs across multiple subjects is a tedious endeavour as, due to the model ambiguity (i.e., component permutation non-uniqueness) of ICA (see Appendix B), RSNs and associated time courses are arbitrarily and differently ordered across subjects. Note that such a post-hoc matching step boils down to a Linear Sum Assignment Problem (LSAP) for which some algorithm was proposed by Kuhn (1955). As a second nuisance, applying the separate ICA's analysis strategy may complicate the detection of shared RSNs as these shared RSNs may become overshadowed by the presence of multiple unique subject-specific RSNs. An example of such a strategy is shown in Calhoun et al. (2001a); the authors performed ICA separately on each subject and manually labelled the estimated RSNs.

Note that when these subject specific RSNs are labelled properly, one could form a 'group' RSNs by simply averaging the RSNs across subjects. Here, the groups should be known a priori and thus no data-driven clustering (i.e., unsupervised) of subjects is performed. As mentioned before, performing ICA on each subject separately is equivalent to performing C-ICA with $R = I$.

Group RSNs and subject specific time courses. A second analysis strategy is Group ICA (Guo and Pagnoni, 2008; Calhoun et al., 2009). In Group ICA, multi-subject fMRI data (after possibly pre-processing and reducing the data dimensionality by means of PCA) of subjects belonging to the same – a priori known – group (e.g., disease subtype or neurofunctional subtype) are concatenated in the temporal domain and ICA is performed on the concatenated data matrix.

Here, each subject is characterized by a set of RSNs that are shared by all subjects belonging to the same group and a *subject specific* set of associated time courses. Often, a back-reconstruction step (Calhoun et al., 2001b; Erhardt et al., 2011) or dual regression procedure (Beckmann et al., 2009; Nickerson et al., 2017) is employed in order to project the estimated group RSNs to each subject's data space, resulting in a different set of RSNs for each subject.

Contrary to single subject ICA, a one-to-one comparison of RSNs and associated time courses across (groups of) subjects is possible and straightforward as these RSNs are uniquely ordered because of their link with the group RSNs, which as such accounts for the aforementioned ICA model ambiguity.

Note that performing Group ICA is equivalent to performing C-ICA with a single cluster ($R = 1$) and thus does not involve a clustering of the subjects. By performing Group ICA on (known) groups of subjects separately, Group ICA allows to study heterogeneity in RSNs between a priori known groups. C-ICA, however, enables researchers to identify unknown subject groups and to study RSNs heterogeneity between these unknown subject groups that are hidden in the data.

Group RSNs and Group time courses. A third method is Tensor Probabilistic ICA (Tensor PICA) which was proposed by Beckmann and Smith (2005). In Tensor PICA, multi-subject fMRI data is decomposed as a trilinear product of group spatial maps, associated group time courses and subject specific weights. Consequently, Tensor PICA results in a set of RSNs representative of the whole group and a set of associated time courses that, due to multiplication with the subject specific weights, are allowed to differ between the group members but only in a restrictive way. In particular, the time courses are constrained to be proportional to each other (i.e., parallel profiles), which, in general, is too restrictive for rs-fMRI data as large differences in brain functioning between subjects can be expected when imaging the brain at rest. C-ICA, as opposed to Tensor PICA, does not pose any restriction on the subject specific time courses. Moreover, unlike Tensor PICA and the other abovementioned ICA strategies, C-ICA involves a clustering of the subjects in order to capture heterogeneity between subjects (clusters) in RSNs.

2.2.2. Clustering procedures

Clusterwise Simultaneous Component Analysis (De Roover et al., 2012b), which – as C-ICA – especially aims at clustering multivariate multi-subject data, is a first clustering method that will be discussed. This method combines clustering with Simultaneous Component Analysis (SCA; Millsap and Meredith, 1988; ten Berge et al., 1992; Timmerman and Kiers, 2003), which can be conceived as a method for group PCA (i.e., the PCA counterpart of Group ICA). When concatenating the data of subjects along the temporal dimension, SCA estimates a loading matrix (i.e., source matrix in ICA) and subject specific component score matrices (i.e., mixing matrices in ICA). By combining clustering with SCA, Clusterwise SCA is able to partition the subjects into homogeneous clusters such that the subjects within each cluster are characterized by the same (cluster specific) component loading matrix. Note that the data can also be concatenated along the spatial dimension, like in temporal Group ICA, resulting in (cluster specific) component scores and subject specific component loadings.

A major difference with C-ICA is that in (Clusterwise-) SCA the components (i.e., RSNs in ICA) are only assumed to be orthogonal – which is closely related to uncorrelated – but not, as in (C-)ICA, non-Gaussian and statistically independent. As a consequence, the components in (Clusterwise-) SCA often have a distribution that is closer to a Gaussian distribution than the components obtained with C-ICA. Therefore, as also demonstrated in Durieux and Wilderjans (2019), (Clusterwise-) SCA is not able to retrieve meaningful RSNs from rs-fMRI data. Indeed, previous studies (e.g., Beckmann et al., 2005) showed that relevant RSNs can successfully be uncovered with ICA based methods that search for independent components. This is presumably due to the non-Gaussian properties and mutually statistical independence of the estimated ICA components/RSNs, which implies that imposing the independence restriction is necessary to retain relevant brain related RSNs.

A second clustering method related to C-ICA is Clusterwise Parafac (Wilderjans and Ceulemans, 2013), which, as Clusterwise SCA, combines clustering with component analysis. Similar as in Clusterwise SCA and C-ICA, Clusterwise Parafac clusters the subjects into homogeneous groups. A difference for this method is that the data within each cluster are modelled with Parafac (Harshman, 1972; Harshman and Lundy, 1994) instead of with SCA or (Group) ICA. Similar as in Tensor PICA, the Parafac model decomposes a three-way data array into a trilinear combination of time courses, RSNs and subject specific weights (for an interesting discussion about the similarities and differences between Tensor PICA and Parafac for multi-subject fMRI data, see Stegeman, 2007). Here, as is true for Tensor PICA, the subject specific weights constrain the time courses of the subjects to be proportional to each other.

However, similar as in Clusterwise SCA, the components that are obtained by Clusterwise Parafac are not assumed to be non-Gaussian and statistically independent, which implies that Clusterwise Parafac will not easily capture meaningful brain related RSNs. Moreover, Clusterwise Parafac, like Tensor PICA, is too strict for rs-fMRI data as one cannot expect that the differences in brain functioning between subjects in rest can be fully captured by proportional time courses.

A third related method is two-step clustering (Durieux and Wilderjans, 2019). This method consists of a two-step procedure in which subjects are partitioned based on the RSNs underlying their rs-fMRI data. In the first step, a single subject ICA – with a fixed number of RSNs – is performed on each subject's data separately in order to obtain subject specific RSNs. In the second step, for each pair of subjects, the similarity between the RSNs of the pair members is computed by means of the modified RV-coefficient, which is a correlation statistic for matrices (Smilde et al., 2009). Note that this statistic circumvents the aforementioned component/RSN permutation model ambiguity. That is, the modified RV-coefficient computes the agreement between the two subspaces spanned by the RSNs of pair members and is invariant under permutations of the RSNs. Next, a hierarchical clustering is performed on the pairwise (dis)similarities in order to partition the subjects in homogeneous groups based on similarities and differences in RSNs.

Drawbacks of this two-step procedure are that several arbitrary choices have to be made that seriously may impact the resulting clustering (e.g., which measure to compute similarity and which method to cluster the similarity matrix) and that the estimation is performed in a sequential fashion in that the RSNs are obtained separately from the subject clusters. Such a sequential procedure, which is known in the literature as tandem analysis, does not guarantee that the extracted RSNs are optimal for clustering the subjects (De Soete and Carroll, 1994; Arabie and Hubert, 1996; Vichi and Kiers, 2001; Timmerman et al., 2010; Steinley et al., 2012). Because RSN extraction (first) and clustering (second) are performed in two separate/sequential steps it is not guaranteed that the RSNs that contain cluster information are extracted and, as a consequence, the true clusters are disclosed. Indeed, in the RSN extraction (first step) the optimized loss function is not aimed at clustering. Hence, a suboptimal estimation of RSNs in step 1 could lead to a suboptimal clustering in step 2. C-ICA, however, simultaneously estimates the subject clusters, RSNs and time courses in an optimal way (see Section 3.2). Here, potentially suboptimal estimated RSNs can be corrected during the iterations of the C-ICA algorithm. In the simulation studies (see Section 4.3), this two-step method will be compared to C-ICA.

A fourth method is Generalized RAICAR (Yang et al., 2012). This method allows for a discovery of sub-groupings of subjects based on the reproducibility of ICA components; this is investigated by performing ICA on the same data set using different starting conditions or applying ICA on bootstrapped data sets. In this method, multiple fastICA analyses are performed on the (bootstrapped) data of each subject and a large similarity matrix is computed that represents both the intra-subject and inter-subject similarities between components. In a next step, components are matched across all subjects in order to produce a set of 'group-level' aligned components. Non-parametric permutation tests are then used to estimate the confidence level of assigning a subject specific component to a group level aligned component, which

indicates the relevance of a person's component for each group-level assigned component. This implies that for each group-level aligned component, multiple subject level components may pass the threshold and that the associated subjects may reflect a possible sub-group within the data. A difference between Generalized RAICAR and C-ICA is that a subject may belong to different sub-groups as his/her components may be linked to different group-level aligned components, whereas in C-ICA subjects are restricted to belong to a single cluster only. Hence, Generalized RAICAR and C-ICA aim at achieving a different goal. Moreover, no explicit clustering procedure is actually performed in this method (i.e., no objective function is minimized). In C-ICA, however, a single subject partitioning is estimated by minimizing a least squares loss function (see Section 3.1).

3. Data analysis

3.1. Aim of C-ICA and loss function

Given multi-subject rs-fMRI data $\mathbf{X}_i$ (T time courses $\times$ V voxels) for subject i ($i = 1, \ldots, I$ subjects), as presented in Fig. 1, and a pre-specified number of clusters R and independent RSNs Q, the aim of C-ICA[3] is to estimate a partitioning matrix $\mathbf{P}$ ($I \times R$), subject specific mixing matrices $\mathbf{A}_i$ ($T \times Q$) ($i = 1, \ldots, I$) and cluster specific RSNs $\mathbf{S}^r$ ($Q \times V$) ($r = 1, \ldots, R$) such that the C-ICA loss function is minimized:

$$L(\mathbf{P}, \mathbf{S}^r, \mathbf{A_i}) = \sum_{i=1}^{I} \|\mathbf{X}_i - \sum_{r=1}^{R} p_{ir}\mathbf{A}_i\mathbf{S}^r\|^2 \tag{2}$$

Moreover, in accordance with the C-ICA model (see Eq. (1)), the RSNs within each cluster $\mathbf{S}^r$ should be (maximally) mutually statistical independent and non-Gaussian (although one single RSN – at most – per cluster is allowed to be Gaussian, see Hyvärinen, 1999; Hyvärinen et al., 2001). Note that by minimizing the loss function from Eq. (2), subjects are clustered and for each cluster a different subspace is determined that optimally approximates the data of that cluster in a least squares sense. By requiring $\mathbf{S}^r$ to be (maximally) independent and non-Gaussian, the RSNs (i.e., ICA components) that determine the subspace of each cluster can be identified, herewith allowing the RSNs to be different between subject clusters.

Before analysing a data set with C-ICA, it is advised to pre-process the data by column-wise centring each data block $\mathbf{X}_i$. As a consequence, for each voxel, the data has a mean of zero across time. Moreover, it is advised to normalize each data block such that each data block has an equal amount of variability. This is simply achieved by multiplying each element of a data block by the square root of a common integer (e.g., 1000) divided by the sum of squared elements of a data block: $\mathbf{X}_i \times \sqrt{\frac{1000}{\|\mathbf{X}_i\|^2}}$. The rationale for this normalization step is that there are usually large scale differences in the observed BOLD response between data blocks, which are arbitrary and are not related to the actual RSNs of interest. The normalization of each data block results in an equal sum of squares for each data block, and therefore, ensures that one particular (subject) data block does not dominate any other data block in the analysis (Wilderjans et al., 2009). Note that this block scaling method only removes scale differences between data blocks but keeps differences within a data block intact, like between voxel differences in the variability of the BOLD response. Note that the loss function from Eq. (2) is minimized for a fixed (given) value of R and that an optimal solution with a larger value of R will always fit the data better or at least as good (i.e., have a smaller or at least equal value of the loss function from Eq. (2)) than an optimal solution with a smaller R value. A similar reason applies to the number of components Q. The optimal value of R and Q, therefore, is determined by a separate model selection procedure (see Section 3.3) that balances the fit of the model to the data with model complexity (which partially depends on the value of R). As such, optimizing R and Q is not part of the C-ICA algorithm (presented in Section 3.2), which prevents the algorithm from always selecting the trivial solution with $R = I$ clusters, which will always yield the lowest loss function value.

3.2. Alternating Least Squares (ALS) algorithm for C-ICA

In order to achieve an optimal subject clustering with the C-ICA model and to determine the associated subject specific mixing matrices and cluster specific RSNs, an Alternating Least Squares (ALS) type of algorithm is constructed. Here, the algorithm alternates between updating partitioning matrix $\mathbf{P}$ (conditionally upon $\mathbf{A}_i$ and $\mathbf{S}^r$) and re-estimating the ICA parameters $\mathbf{A}_i$ and $\mathbf{S}^r$ (conditionally upon $\mathbf{P}$) until convergence is reached. Note that the most famous algorithms for k-means clustering also imply alternatingly (conditionally) updating the clustering and (conditionally) re-estimating the parameters (i.e., centroids) per cluster. Note further that each updating step results in the loss function from Eq. (2) decreasing or staying the same (which implies convergence). More specifically, the C-ICA algorithm consists of the following four steps:

1. Initialize partition matrix $\mathbf{P}$. Here, each subject data block i ($i = 1, \ldots, I$) is allocated to one of the clusters, with each data block having the same probability of being assigned to each cluster (i.e., $\frac{1}{R}$). Note that this procedure is repeated until all clusters are non-empty (i.e., it is guaranteed that exactly R clusters are obtained). As such, this procedure yields a starting subject partition with more or less equally sized clusters. Besides using a random initial partition, it is also possible to seed the algorithm with a user-defined starting partition or to employ a data-driven starting subject partition, like, for example, the partition obtained from the procedure of Durieux and Wilderjans (2019).
2. Estimate the C-ICA parameters for each cluster (conditionally upon $\mathbf{P}$). First, all data blocks $\mathbf{X}_i$ that belong to cluster C_r are concatenated along the temporal dimension into a large matrix $\mathbf{X}^r$. Next, spatial (Group) ICA by means of fastICA (Hyvärinen, 1999) with Q RSNs is performed on each of the concatenated data blocks $\mathbf{X}^r$ ($r = 1, \ldots, R$) in order to estimate the C-ICA parameters for each cluster (i.e., the Q cluster specific RSNs $\mathbf{S}^r$ and the Q $\mathbf{A}_i$'s for the subjects i of that cluster). Note that fastICA implements ICA through the maximization of the negentropy of the independent RSNs after – possibly – data reduction and a pre-processing step known as whitening (for more information, see Appendix A) and is performed using the R package `ica` (Helwig, 2018). Note, however, that other types of ICA implementations, such as ICA that uses the Infomax algorithm (Bell and Sejnowski, 1995) can be used. Applying fastICA to $\mathbf{X}^r$ (which is a concatenation of the $\mathbf{X}_i$'s of subjects i belonging to cluster C_r) and asking for Q RSNs (which is lower than the maximum number of RSNs that could be extracted from $\mathbf{X}^r$) results in the $\mathbf{A}_i$'s and $\mathbf{S}^r$ that minimize $\sum_{i \in C_r} \|\mathbf{X}_i - \mathbf{A}_i\mathbf{S}^r\|^2$, with C_r denoting the rth subject cluster. Indeed, for small(er) Q, fastICA involves a data reduction (minimizing $\|\mathbf{X}^r - \mathbf{A}^r\mathbf{S}^r\|^2$ with $\mathbf{A}^r$ being a large matrix consisting of the concatenated $\mathbf{A_i}'s$ of subject i belonging to cluster C_r and the rank of $\mathbf{A}^r\mathbf{S}^r$ being Q) plus a whitening step along with a rotation step (in which negentropy is maximized) to find the independent RSNs for cluster C_r As a consequence, applying fastICA to each $\mathbf{X}^r$ will decrease the loss function (Eq. (2)) and will retain the RSNs for each cluster. After computing the cluster specific C-ICA parameters, the C-ICA loss function is evaluated.

[3] With the term C-ICA we mean the procedure/algorithm that estimates the parameters of the C-ICA generative model (for a given value of R and Q) as described in Section 2.1.

3. Update partition matrix $\mathbf{P}$ subject (data block) by subject (data block) conditionally upon $\mathbf{A}_i$ and $\mathbf{S}^r$. Here, the optimal cluster membership for data block $\mathbf{X}_i$ is determined by evaluating for each cluster C_r the fit of the data block $\mathbf{X}_i$ under consideration to that cluster by means of the partition criterion $L_{ir} = \|\mathbf{X}_i - \hat{\mathbf{X}}_i^{(r)}\|^2$; each data block $\mathbf{X}_i$ is assigned to the cluster C_r ($r = 1, \ldots, R$) for which L_{ir} is minimal. More specifically, for each data block $\mathbf{X}_i$ and each cluster C_r an estimated $\hat{\mathbf{X}}_i^{(r)}$ is computed through the formula $\hat{\mathbf{X}}_i^{(r)} = \hat{\mathbf{A}}_i^{(r)}\mathbf{S}^r$, where $\mathbf{S}^r$ is given by the previous fastICA estimate of $\mathbf{S}^r$ in step 2 of the algorithm and $\hat{\mathbf{A}}_i^{(r)}$ is computed via $\hat{\mathbf{A}}_i^{(r)} = \mathbf{X}_i\mathbf{S}^{r^T}(\mathbf{S}^r\mathbf{S}^{r^T})^{-1}$ with $(\ldots)^{-1}$ denoting matrix inversion and $\mathbf{S}^T$ the transpose of a matrix; $\hat{\mathbf{A}}_i^{(r)}$ is the associated mixing matrix for subject i when that subject is considered to belong to cluster C_r ($r = 1, \ldots, R$). This reassignment step implies that the C-ICA loss function (Eq. (2)) decreases or – when convergence is reached – stays at the same value. Note that after reassigning all subject data blocks to their optimal cluster, it could occur that some clusters are empty. In order to avoid empty clusters, a procedure is applied that puts the data block with the worst fit into an empty cluster (when this creates a new empty cluster, the next worst fitting data block is chosen). In particular, the data block with the largest L_{ir} (for optimal R) is assigned to an empty cluster. This procedure is continued until all clusters are non-empty.
4. Convergence criterion. Steps 2 and 3 are repeated until the decrease in the C-ICA loss function value (Eq. (2)) between two evaluations is smaller than the convergence criterion (e.g., .000001). Note that, instead of an absolute, it is also possible to choose a relative convergence criterion, like $\frac{L}{.000001}$.[4]

Applying this ALS algorithm, which belongs to the class of block relaxation algorithms, produces a nonincreasing sequence of loss function values (see Eq. (2)), which – due to the loss function being bounded below by zero – is convergent under mild conditions (de Leeuw, 1994). Because the C-ICA algorithm, as is true for almost all ALS (de Leeuw, 1994) and clustering algorithms (Brusco, 2006), may yield a local optimal solution, a multi-start procedure is strongly advised. In this procedure, several different runs (e.g., 30) of the C-ICA algorithm are performed, each run starting with a different random initialization of the partition matrix $\mathbf{P}$. Additionally, in order to lower the risk of the algorithm ending in a suboptimal solution, it also makes sense to start the algorithm with rational and pseudo-random starting partitions (for a discussion of different types of starts, see Ceulemans et al., 2007). A rational start can be obtained by using the subject partition that results from the two-step procedure (see Section 2.2) that was proposed by Durieux and Wilderjans (2019). Specifically, in this procedure a single-subject ICA with Q RSNs is performed on the data of each subject and the similarity in RSNs, quantified by the modified RV-coefficient (Smilde et al., 2009), is calculated for each subject pair. A rational subject partition is next obtained by applying hierarchical clustering using Ward's method to these (dis)similarities and cutting the resulting dendrogram such that R clusters are retained. A pseudo-random start can be attained by perturbing the rationally obtained starting partition through assigning at random a small number of subjects – like 10% – to another cluster. After running a combination of random, rational and pseudo-random starts, the solution with the lowest C-ICA loss function value (Eq. (2)) encountered across all runs is retained as the final solution. The aim of this procedure is to minimize the possibility that a local optimum of the C-ICA loss function is retained. Note that an open source (pilot version of an) R-package for C-ICA is available on CRAN.[5]

[4] After convergence, as an additional step, one can employ a back-reconstruction procedure (See Calhoun et al., 2001b; Beckmann et al., 2009; Erhardt et al., 2011) to project the cluster specific RSNs on the data space of each subject.

[5] https://cran.r-project.org/web/packages/CICA/index.html.

3.3. Model selection

When performing C-ICA, the number of clusters R and number of RSNs Q per cluster – which will be assumed to be equal across clusters – should be specified a priori. In general, however, no a priori information regarding the optimal number of clusters or RSNs per cluster is present. A way to determine these numbers consists of running C-ICA with increasing numbers of RSNs (e.g., from one up to seventy) and clusters (e.g., from one up to six) and then use a model selection heuristic to identify the optimal number of clusters and RSNs.

To address this problem, we propose to use a sequential model selection procedure that consists of the following two steps: (1) determine the optimal number of clusters R and, next, (2) select, conditional upon the optimal R, the optimal number of RSNs Q (for similar procedures, see De Roover et al., 2012a; Wilderjans and Ceulemans, 2013). In both steps, a procedure based on scree ratios is used to determine the optimal number of clusters/RSNs in an automated way (Ceulemans and Kiers, 2006; Wilderjans et al., 2013). Here, for the first step, the scree ratio for a certain number of clusters r (keeping the number of RSNs fixed at q) is computed as follows:

$$sr_{r|q} = \frac{L_{r-1,q} - L_{r,q}}{L_{r,q} - L_{r+1,q}}, \tag{3}$$

where $L_{r,q}$ is the C-ICA loss function value from Eq. (2) for a C-ICA model with r clusters and q RSNs. Large $sr_{r|q}$ values are preferred because such a large scree ratio value implies that a model with one cluster less ($r-1$) than the considered model fits the data substantially worse (i.e., $sr_{r|q}$ has a large numerator), whereas a model with one extra cluster ($r+1$) does not fits the data considerably better (i.e., a small denominator). After computing $sr_{r|q}$ in Eq. (3) for each possible r ($r = R_{min}+1, \ldots, R_{max}-1$) and each q ($q = Q_{min}, \ldots, Q_{max}$), the optimal number of clusters R is determined by averaging $sr_{r|q}$ over all considered number of RSNs q and selecting the number of clusters r that has the largest averaged $sr_{r|q}$-ratio. Note that a procedure based on the scree ratio does not allow to select the smallest (i.e., $R = R_{min}$) and largest number (R_{max}) of clusters considered, as the ratio is not defined for these values. As such, R_{max} should be chosen large enough and R_{min} small enough such that the optimal r is ensured to lie in between R_{min} and R_{max} (Wilderjans et al., 2013). Next, in the second step, conditional upon the optimal number of clusters R derived in step 1, the optimal number of RSNs Q is determined by selecting the number of RSNs q ($q = Q_{min}+1, \ldots, Q_{max}-1$) that maximizes the following $sr_{q|R}$-ratio:

$$sr_{q|R} = \frac{L_{q-1,R} - L_{q,R}}{L_{q,R} - L_{q+1,R}} \tag{4}$$

Also here, Q_{min} and Q_{max} cannot be selected by the model selection procedure and should thus be chosen wisely in order to not miss the optimal q.

Note that increasing either Q or R cannot result in a decrease of fit (i.e., a larger loss function value), unless, however, a very bad locally optimal solution is retained. Such suboptimal solutions should either be ignored or, preferably, restarted with more random, pseudo-rational and rational starts in order to prevent the locally optimal solution being retained. As a consequence, the numerator and denominator of Eq. (4) will always be nonnegative. Also note that it can happen that no clusters are present in the data (i.e., true $R = 1$). In order to check whether this might be the case, the mean modified-RV can be computed between RSNs across cluster pairs (for $R > 1$). Here, a mean modified-RV close to one (i.e., all cluster-specific $\mathbf{S}^r$ being almost identical) might be an indication that there is no cluster structure present in the data. Finally, as in all model selection procedures, the final decision about model retention should also be based on the interpretability of the C-ICA solution.

4. Validation of the C-ICA method

In this section, in order to validate the C-ICA method, four simulation studies are presented that vary in the degree to which the fMRI data are realistically simulated (i.e., contain properties of empirical fMRI data). The rationale and principles behind these four simulation studies align with the five principles (i.e., declaration of intent, fairness to used model, control of key aspects, start with simplicity and increase complexity gradually and make realistic simulations) discussed in Silva et al. (2014). Furthermore, in a fifth study, the proposed sequential procedure for selecting the optimal number of clusters and RSNs (see Section 3.3) is investigated. In the first simulation study, the C-ICA algorithm is evaluated with respect to its ability to correctly estimate the subject clusters and the underlying cluster specific RSNs and subject specific time courses in relatively easy clustering situations. Also, the influence of several systematically varied data characteristics on the recovery performance of C-ICA is studied and C-ICA's performance is compared to two other clustering procedures.

In the second simulation study, the C-ICA algorithm is evaluated on more realistically simulated resting-state fMRI data. In particular, RSNs are generated based on a template with existing RSNs (e.g., the default mode network; DMN), time courses are simulated with fluctuation frequencies commonly encountered in resting-state time courses and more realistic error structures are used to generate the noise in the data.

In the third study, C-ICA is evaluated in situations in which the clustering is harder to uncover from the data. Here, data sets with spatially overlapping cluster specific RSNs and noise patterns that obfuscate the true simulated clustering are generated. Moreover, the performance of the C-ICA algorithm is compared to the aforementioned two-step clustering procedure developed by Durieux and Wilderjans (2019) and also to a variation of that approach using Group ICA plus dual regression instead of separate subject-specific ICAs.

In the fourth study, a multi-subject rs-fMRI data set is generated by selecting data from an existing empirical rs-fMRI data set such that two groups of subjects are created that substantially vary in underlying RSNs. By adopting this data selection strategy, C-ICA is performed and compared to the two other clustering procedures on empirical (non-simulated) data in which a clear two-group cluster structure is present.

Finally, in a fifth study, the sequential model selection procedure from Section 3.3 is tested on some of the generated data sets from the second simulation study. The goal of the fifth study is to see whether the proposed model selection procedure is able to select the correct number of RSNs and clusters underlying a (simulated) data set. Note that all R-code for conducting the simulation studies can be found online at https://github.com/jeffreydurieux/cica_simulation

4.1. Simulation study 1

4.1.1. Problem

In the first simulation study, data were generated under a C-ICA model with a – known – true number of clusters R and a true number of independent RSNs Q and these generated data sets were subjected to a C-ICA analysis using only the true number of clusters R and RSNs Q. First, it is investigated to which extent the C-ICA algorithm succeeds in avoiding local optimal solutions. Secondly, the C-ICA algorithm is evaluated with respect to goodness of recovery, that is, whether (1) the clustering of the subject data blocks ($\mathbf{P}$) is successfully recovered, (2) the cluster specific RSNs $\mathbf{S}^r$ are correctly disclosed and (3) the time courses $\mathbf{A}_i$ are accurately retrieved.

Furthermore, it is examined whether the performance of the algorithm depends on characteristics of the data (e.g., the number of voxels) and/or on the complexity of the true underlying C-ICA model (e.g., the number of underlying clusters/RSNs) and/or on the amount of noise in the data. Based on previous research, expectations are that the C-ICA algorithm will perform better when the data contain more time points (i.e., realized samples of the independent RSNs) as this implies more available information (Brusco and Cradit, 2005; De Roover et al., 2012b). Furthermore, it can be postulated that the goodness of recovery will deteriorate with increasing complexity (i.e., more clusters and independent RSNs) of the underlying C-ICA model (Milligan et al., 1983; De Roover et al., 2012b; Wilderjans et al., 2012a,c, 2017) and when the data contain more noise (Wilderjans et al., 2008, 2011, 2012b; De Roover et al., 2012b; Wilderjans and Ceulemans, 2013).

Finally, focusing on the recovery of the clustering, the C-ICA performance is compared to the performance of the two step procedure of Durieux and Wilderjans (2019) and a Group ICA with dual regression procedure with an additional clustering step.

4.1.2. Design and procedure

In order to not have an overly complex design, the number of subjects was fixed at 40 and only clusters of equal size (i.e., 40 divided by the number of clusters R) were considered. Furthermore, five factors were systematically varied in a full five-factorial randomized design in which all factors were considered as random factors (i.e., the selected values for each factor were considered as sampled at random from a wider population of possible levels for that factor):

1. Number of voxels V, at two levels: 500 and 2000;
2. Number of (share) independent components or RSNs Q, at three levels: 2, 5 and 20;
3. Number of (equally sized) clusters R, at two levels: 2 and 4;
4. Dimension of the mixing matrices $\mathbf{A}_i$ $(T \times Q)$, which affects the number of time points T, at 2 levels: square ($T = Q$) or non-square ($T = 100$);
5. The amount of Gaussian and independent noise ϵ in the data, at three levels: 5%, 20% and 40%.

With regard to the fourth factor, in the case of a square mixing matrix, the dimensions of all subject specific $\mathbf{A}_i$'s depend on the number of independent RSNs Q (i.e., $\mathbf{A}_i$ has dimensions either $2 \times 2, 5 \times 5$ or 20×20). In the non-square conditions, however, the number of observed mixtures (i.e., time points T) was held constant at 100, a number larger than the number of independent RSNs, resulting in the mixing matrices having dimensions either $100 \times 2, 100 \times 5$ or 100×20. The subject specific mixing matrices $\mathbf{A}_i$ were generating by drawing values at random and independently from a uniform distribution $U(-2, 2)$. Note that in empirical practice, the dimensionality of the data is often reduced with PCA before performing ICA, which results in a square ICA mixing matrix.

Furthermore, as in the general ICA model, the C-ICA model assumes that the independent RSNs are non-Gaussian. Therefore, the Q RSNs for a particular cluster specific $\mathbf{S}^r$ were independently generated from a double exponential distribution, which implies a super-Gaussian distribution (i.e., a distribution where most voxels will have activation levels near zero), with a rate parameter $\lambda = \sqrt{2}$. We simulated from this particular distribution since it provides a good representation of RSNs encountered in empirical research (Calhoun and Adalı, 2006; Ashby, 2011). To this end, the R function `icasamp()` from the ica package (Helwig, 2018) was used. This function ensures that the generated RSNs are independent and non-Gaussian and have a mean of zero. Next, the independent RSNs were mixed with the subject specific mixing matrices $\mathbf{A}_i$ according to the C-ICA model from Eq. (1), resulting in the true data blocks $\mathbf{T}_i$.

Later, a noise matrix $\mathbf{E}_i$ $(T \times V)$ was added to each true data block $\mathbf{T}_i$ $(T \times V)$, with $\mathbf{T}_i = \mathbf{A}_i\ \mathbf{S}^r$ and r denoting the cluster to which subject i belongs. Here, each noise matrix $\mathbf{E}_i$ was constructed by independently drawing numbers from $\mathcal{N}(0, 1)$. Next, the noise matrices were rescaled such that their sum of squared entries (SSQ) equalled the SSQ of the corresponding noiseless data block $\mathbf{T}_i$. After equalling the SSQ of the noiseless data and the noise matrix, an appropriate rescaling of $\mathbf{E}_i$ ensures that all data blocks $\mathbf{X}_i$ contain the required percentage of noise.

More specifically, a weight parameter w was used to manipulate the percentage of noise in the data blocks $\mathbf{X}_i = \mathbf{T}_i + w\mathbf{E}_i$, where $w = \sqrt{\frac{noise}{1-noise}}$ and $noise$ equals .05, .20 and .40 for the desired percentage of noise of 5%, 20% and 40%, respectively. Note that the noise percentage levels of 5%, 20% and 40% used in the current procedure refer to a signal to noise ratio of 19, 4 and 1.5, respectively, which is defined here as $\frac{1-noiselevel}{noiselevel}$ (with the noise level being expressed as a proportion).

Lastly, for each cell in the five-factorial design, 10 replication data sets were generated. Thus, in total, 2 (number of elements) × 3 (number of independent RSNs) × 2 (number of clusters) × 2 (dimension of mixing matrix) × 3 (error) × 10 (replications) = 720 C-ICA data sets were generated. Each data set was analysed with the C-ICA algorithm with true R and Q and with 30 random starts, and the solution with the lowest value on the C-ICA loss function L (see Eq. (2)) was retained.

All 720 data sets were also analysed, using the true R and Q, with the two step procedure of Durieux and Wilderjans (2019) using Ward's hierarchical clustering method and subjected to a Group ICA plus dual regression analysis with an additional hierarchical clustering step using Ward's method. As Group ICA plus dual regression does not produce a clustering of the subjects, we computed for each subject pair the modified-RV coefficient between the RSNs of the pair members and subjected the resulting dissimilarity matrix to Partitioning Around Medoids (PAM) and Ward's hierarchical clustering with $R = 2$. This procedure is the same as the two step procedure of Durieux and Wilderjans (2019) with the only difference being that in the latter procedure the input RSNs are based on single subject ICA's instead of on (subject specific) RSNs obtained by Group ICA plus dual regression (see Fig. 5).

4.1.3. Results

Local optima of the C-ICA loss function. Two indications can be used to determine whether or not the C-ICA algorithm returned a local optimal solution. First, a solution that is locally optimal is not expected to be retained in many of the multiple (random) starts of the algorithm. As a consequence, encountering the optimal solution across most of the multiple starts increases the probability that this solution is globally optimal. Second, a solution retained by C-ICA is for sure only locally (and not globally) optimal when it has a loss function value that is larger than the loss function value from the solution that is obtained by seeding the C-ICA algorithm with the true clustering (which in the case of a simulation is known). Note that the second indication can be considered a sanity check for the C-ICA algorithm. Regarding the first indication, for each simulated data set, we computed the percentage of multi starts that yielded the same solution as the solution retained by C-ICA (which is the solution with the lowest loss value across the 30 random starts). As can be seen from Table 1, the mean percentage across all simulated data sets equals 77.30% (SD = 22.70%). This indicates that, for our data, a relatively large proportion of random starts ends up in the same optimal solution. Note that the percentage decreases, meaning the algorithm has more difficulties to avoid local minima, when the dimension of the mixing matrix is square (mean is 68.61%) compared to non-square (mean is 85.98%) and when the amount of noise present in the data increases (means are 80.63%, 77.06% and 74.21% for 5%, 20% and 40% of noise, respectively). With respect to the second indication, for all 720 simulated data sets except one,[6] the solution retained by C-ICA has a loss function value that is lower or equal to the loss function value of the solution (before and after ALS iterations) that is obtained by running C-ICA analyses with the true subject partition as a rational start. In fact, when seeding C-ICA with the true partition, ALS iterations did not improve this true partitioning. In sum, these results suggest that the C-ICA algorithm, for the current simulation study (with relatively easy clustering situations), does not suffer from a serious local optima problem.

[6] For this data set, the algorithm did not converge to the global minimum and the associated ARI equals .933.

*Recovery of the clustering of the data blocks (**P**)*. In order to evaluate the goodness of recovery for the clustering of the subject data blocks, the Adjusted Rand Index (ARI; Hubert and Arabie, 1985) is computed between the true partition of the data blocks and the estimated partition (i.e., the partition that yields the lowest loss function value across the 30 random starts used). The ARI equals 1 if two partitions are identical and 0 when the overlap between both partitions is at chance level. The overall mean ARI, across all 720 data sets, equals .9999 (SD = .0026). It can be concluded that the C-ICA algorithm recovers the clustering of the data blocks to a very large extent.

To study how the recovery of the clustering of the data blocks changes as a function of the manipulated factors, Table 1 gives an overview of the mean ARI (and standard deviation of ARI) for each level of the five manipulated factors. From this table, it can be seen that when the amount of noise is low or moderate (i.e., 5% or 20%), a perfect recovery is encountered for each data set. Moreover, a perfect recovery of the simulated partitioning is achieved when the mixing matrix is non-square. Finally, recovery decreases when the mixing matrix becomes square (i.e., $Q = T$) instead of non-square (i.e., $Q < T = 100$), when the number of RSNs and voxels becomes smaller and when the number of clusters and the amount of noise in the data becomes larger.

Recovery of the cluster specific independent RSNs $\mathbf{S}^r$. To evaluate the extent to which the true independent components or RSNs were recovered, for each RSN separately, the Tucker congruence coefficient (Tucker, 1951) is computed between the simulated independent RSN and the corresponding estimated independent RSN. To arrive at a single Tucker congruence value for each $\mathbf{S}^r$, for each of the Q RSNs (i.e., columns of $\mathbf{S}^r$) the Tucker value is computed (after accounting for the C-ICA ambiguities, see further) and the mean across these Q obtained Tucker values is calculated. To obtain a single Tucker value for each generated data set, the (mean) Tucker values of the cluster specific $\mathbf{S}^r$ were averaged across the R clusters. Tucker's congruence coefficient equals the normalized inner product between two vectors and ranges from −1 to 1, with 1 indicating perfect recovery and 0 chance level. A value in the range of .85-.94 denotes a fair similarity between the two vectors, whereas a value of .95 or larger indicates that the two vectors are very similar (Lorenzo-Seva and Ten Berge, 2006).

Determining the degree to which the simulated independent RSNs are recovered by the estimated independent RSNs is not straightforward as the C-ICA model suffers from four ambiguities (see Appendix B): scaling ambiguity, reflectional ambiguity and component/RSN and cluster permutational freedom ambiguity. To take these ambiguities into account when computing Tucker's congruence, the following procedure was followed: (1) the absolute value of the Tucker coefficient is taken to account for reflectional freedom; (2) to account for the permutational freedom of both the RSNs and the clusters, all possible combinations of cluster and RSN permutations are considered and for each combination of these permutations the associated mean Tucker congruence (averaged across RSNs and $\mathbf{S}^r$'s) is computed. Next, the combination of a cluster and a RSN permutation with the largest Tucker congruence value is retained and the associated averaged Tucker value is reported. Note that as the Tucker coefficient is invariant under a scaling of the RSNs with a positive scalar, this coefficient automatically accounts for the scaling ambiguity of the C-ICA model. By taking the absolute value (see reflectional ambiguity), also scaling with a negative number is accounted for.

As the overall mean Tucker congruence equals .9826 (SD = .0197), it can be concluded that the C-ICA algorithm recovers the independent RSNs very good. The mean Tucker congruence value (and standard deviation) for each level of each factor can be found in Table 1. From this table it can be seen that the mean Tucker congruence varies as a function of the amount of error, with recovery deteriorating when error increases. Similar as for the partitioning recovery results, a small deterioration in recovery of the independent RSNs is encountered when the mixing matrix is square compared to non-square. Finally, recovery

Table 1
Mean ARI (cluster recovery), Tucker's congruence (cluster specific spatial maps $\mathbf{S}^r$ and subject specific time courses $\mathbf{A}_i$ recovery) and percentage multi-starts resulting in the same lowest loss function value (and standard deviation) for each level of the manipulated factors of the first simulation study.

Factor	Level	ARI	$\mathbf{S}^r$ recovery	$\mathbf{A}_i$ recovery	Percentage
Number of voxels	500	0.9998 (0.0036)	0.9754 (0.0235)	0.9809 (0.0210)	77.0278 (22.7216)
	2000	1 (0.000)	0.9897 (0.0112)	0.9964 (0.0036)	77.5648 (22.7163)
Number of RSNs	2	0.9997 (0.0044)	0.9903 (0.0159)	0.9980 (0.0045)	62.3472 (30.6341)
	5	1 (0.0000)	0.9889 (0.0110)	0.9952 (0.0034)	88.9167 (12.0101)
	20	1 (0.0000)	0.9685 (0.0223)	0.9727 (0.0211)	80.6250 (9.8491)
Number of Clusters	2	1 (0.0000)	0.9865 (0.0154)	0.9891 (0.0158)	84.1389 (20.6562)
	4	0.9998 (0.0036)	0.9795 (0.0229)	0.9881 (0.0180)	70.4537 (22.6244)
Dimension of the Mixing matrix	Square	0.9998 (0.0036)	0.9758 (0.0219)	0.9881 (0.0180)	68.6111 (27.0020)
	Non-square	1.000 (0.000)	0.9893 (0.0145)	0.9892 (0.0158)	85.9815 (12.3309)
Amount of noise	0.05	1.000 (0.000)	0.9895 (0.0132)	0.9909 (0.0132)	80.6250 (20.8474)
	0.20	1.000 (0.000)	0.9847 (0.0151)	0.9895 (0.0151)	77.0556 (22.8356)
	0.40	0.9997 (0.0044)	0.9735 (0.0252)	0.9855 (0.0210)	74.2083 (23.9580)
Overall		0.9999 (0.0026)	0.9826 (0.0197)	0.9886 (0.0169)	77.2963 (22.7047)

also decreases when the number of voxels decreases and the number of RSNs and clusters increases.

Recovery of the subject specific time courses $\mathbf{A}_i$. In order to evaluate to what extent the time courses (i.e., mixing matrices) $\mathbf{A}_i$ are recovered, Tucker's congruence coefficient is computed between the simulated and estimated mixing matrices. This measure is computed in a similar way as was done for determining the recovery of the independent RSNs, herewith also accounting for the C-ICA ambiguities (see Appendix B) except for the cluster permutation non-uniqueness as the $\mathbf{A}_i$'s are subject specific and not cluster specific.

The mean Tucker congruence for the mixing matrices across all data sets is .9886 (SD = .0169). Therefore, it can be concluded that the C-ICA algorithm recovers the underlying time courses to a very large extent, albeit to a slightly lesser extent than the independent RSNs. As can be seen in Table 1, the recovery especially decreases when the number of RSNs Q increases. Further, analogue to the results of the other recovery indices, the time courses are recovered less accurately when the mixing matrix is square (compared to non-square) and when the data contain more noise, more clusters and less voxels.

Analysis of variance. To evaluate potential main and interaction effects between the manipulated factors, three separate analyses of variance (ANOVA) were performed. In these analyses, the aforementioned five factors pertaining to data characteristics (see Section 4.1.2) were treated as between-subjects factors and the ARI and Tucker's congruence coefficients for the RSNs and the time courses were used as the dependent variables (i.e., each outcome measure was used in a separate ANOVA).

When discussing the results of these analyses, only significant effects with a medium effect size as measured by the generalized eta squared η_G^2 (Olejnik and Algina, 2003) are considered (i.e., $\eta_G^2 > .15$). Main effects and interaction effects from the ANOVA's with $p < .05$ and $\eta_G^2 > .15$ are displayed in Table 2 (for the recovery of the RSNs) and Table 3 (for the recovery of the time courses). Note that no results are presented for the ANOVA with ARI as dependent variable as no significant main and interaction effects were encountered for that outcome measure. This is caused by the very large mean ARI's – indicating the almost perfect cluster recovery for all generated data sets – which results in a negligible variation in ARI's (see also Table 1).

From Table 2, which presents all significant main and interaction effects with an effect size larger than .15, it appears that the recovery of the cluster specific RSNs mainly depends on the number of components ($\eta_G^2 = .74$) and to a lesser extent on the number of voxels ($\eta_G^2 = .60$), the amount of noise present in the data ($\eta_G^2 = .57$) and the dimensions of the mixing matrix ($\eta_G^2 = .57$); the number of clusters present in the data only plays a minor role ($\eta_G^2 = .22$). However, as shown in Table 2, these main effects are qualified by four two-way interactions. First, as can be seen in the upper left panel of Fig. 2, the recovery of the RSNs deteriorates when the number of components Q increase, but this

Table 2
ANOVA table resulting from the ANOVA with Tucker congruence of the cluster specific spatial maps $\mathbf{S}^r$ as the dependent variable (first simulation study), considering only significant (at $\alpha = .05$) main and interaction effects with an effect size $\eta_G^2 > .15$.

Effect	SS effect	SS error	DF effect	DF error	F	η_G^2
# voxels	.036	.025	1	648	955.10	.60
# RSNs	.071	.025	2	648	932.86	.74
# Clusters	.007	.025	1	648	178.40	.22
Dim. mixing matrix	.033	0.25	1	648	858.98	.57
Amount of noise	.033	.025	2	648	428.18	.57
# voxels × # RSNs	.035	.025	2	648	453.67	.58
# Clusters × Dim. mixing matrix	.005	.025	1	648	137.44	.17
# Clusters × Amount of noise	.005	.025	2	648	65.08	.17
Dim. mixing Matrix × Amount of noise	.022	.025	2	648	290.53	.47

Note. Dim. mixing matrix = dimension of mixing matrix (square or non-square).

deterioration is slightly more pronounced when the data contain less voxels ($\eta_G^2 = .58$). Further, the upper right panel of Fig. 2 shows that the recovery of the RSNs becomes worse when noise intensifies, with this effect being more pronounced when the mixing matrix is square compared to non-square ($\eta_G^2 = .47$). Next, as presented in the bottom left panel of Fig. 2, the decline of the recovery due to the increase of noise ameliorates when less clusters are present in the data ($\eta_G^2 = .17$). Finally, the bottom right panel of Fig. 2 demonstrates that the recovery of the RSNs deteriorates slightly when the number of clusters increases and that this effect is more obvious when the mixing matrix is square instead of non-square ($\eta_G^2 = .17$).

In Table 3, the significant main and interaction effects with an effect size larger than .15 are presented for the ANOVA on the recovery of the time courses. It appears from this table that, in terms of main effects, the recovery of the subject specific time courses mainly depends on the number of RSNs ($\eta_G^2 = .89$), number of voxels ($\eta_G^2 = .79$) and the amount of noise present in the data ($\eta_G^2 = .25$). These main effects, however, are qualified by two two-way interactions (see Table 3 and Fig. 3). Similar as for the recovery of the spatial maps, the recovery of the subject specific time courses depends on the interaction between the number of voxels and the number of RSNs ($\eta_G^2 = .80$). More specifically, as showed in the left panel of Fig. 3, the mean Tucker congruence decreases when the number of RSNs increases, but this decrease is more pronounced when the number of voxels is low. Finally, the second two-way interaction effect pertains to the effect of the number of RSNs and the amount of noise in the data ($\eta_G^2 = .21$). Here, as shown in the right panel of Fig. 3, the recovery of the subject specific time courses decreases when there is more noise in the data, but this decrease is stronger when more RSNs are present in the data.

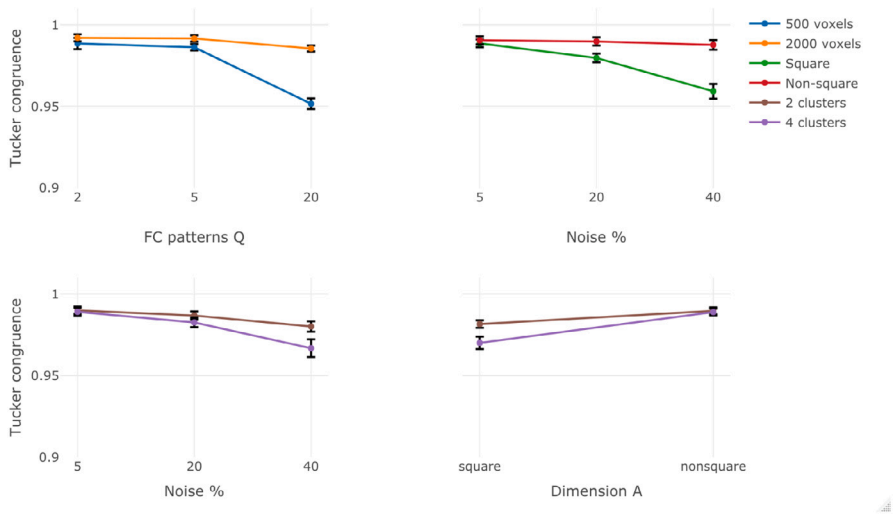

Fig. 2. Four two-way interaction effects from the analysis of variance of the first simulation study with Tucker's congruence coefficient between the simulated and estimated RSNs as outcome variable: (upper left panel) Congruence as a function of the number of simulated RSNs and the number of voxels; (upper right panel) Congruence as a function of the percentage of noise added to the data and the dimension of the mixing matrix (either square or non-square); (bottom left panel) Congruence as a function of the percentage of noise and the number of underlying clusters in the data; (bottom right panel) Congruence as a function of the dimension of the mixing matrix and the number of underlying clusters. Error bars represent two times the standard error (SE).

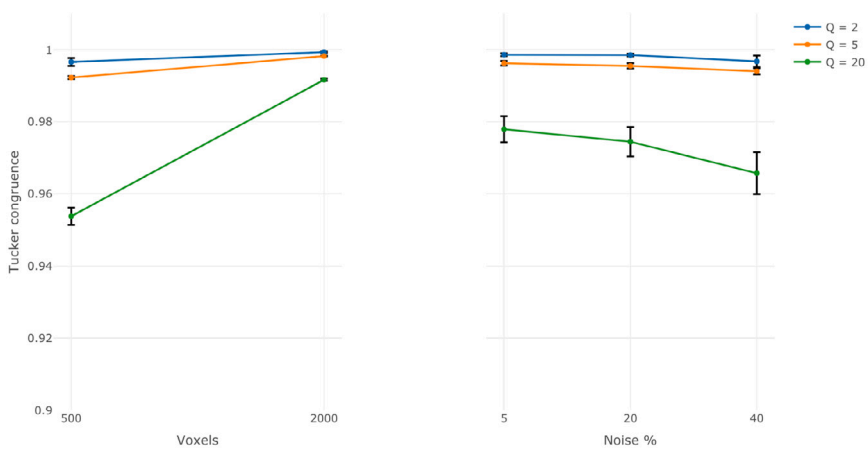

Fig. 3. Two two-way interaction effects from the analysis of variance of the first simulation study with Tucker's congruence coefficient between the simulated and estimated time courses as outcome variable: (left panel) Congruence as a function of the number of voxels and the number of simulated RSNs; (right panel) Congruence as a function of the percentage of noise added to the data and the number of simulated RSNs. Error bars represent two times the standard error (SE).

Table 3
ANOVA table resulting from the ANOVA with Tucker congruence of the subject specific time courses $\mathbf{A}_i$ as the dependent variable (first simulation study), considering only significant (at $\alpha = .05$) main and interaction effects with an effect size $\eta_G^2 > .15$.

Effect	SS effect	SS error	DF effect	DF error	F	η_G^2
# voxels	.043	.011	1	648	2502.59	.79
# RSN	.092	.011	2	648	2661.89	.89
Amount of noise	.003	.011	2	648	109.59	.25
# voxels × # RSN	.045	.011	2	648	1306.80	.80
# RSN × Amount of noise	.003	.011	4	648	41.93	.21

Note. Dim. mixing matrix = dimension of mixing matrix (square or non-square).

Comparison of C-ICA with two other clustering methods. Results indicate that both the two-step method of Durieux and Wilderjans (2019) and a cluster analysis based on results of Group ICA plus dual regression recover the simulated clustering to a very large extent. The overall mean ARI (computed over all 720 analyses) for the two-step method using Ward's hierarchical clustering equals 1 (SD = 0), indicating that all simulated clusterings were correctly estimated irrespective of the manipulated factors. These results are comparable to the results of C-ICA (overall mean ARI = 0.9999; SD = 0.0026). The overall mean ARI for Ward's hierarchical clustering using information from a Group ICA plus dual regression equals 0.7986 (SD = 0.2744), implying that this procedure recovers the correct clustering to a substantially smaller extent than C-ICA. When the results were computed for each manipulated factor, a strong effect was found for the dimension of the mixing matrix factor. The mean ARI for data sets with a square underlying mixing matrix equals 0.5973 (SD = 0.2637), whereas the mean ARI conditional on a non-square mixing matrix equals 1 (SD = 0).

It can be concluded that for randomly simulated data the C-ICA algorithm does not suffer from a serious local optima problem and recovers the underlying clusters, RSNs and time courses to a very large extent. Moreover, although for some data sets the difference is negligible, C-ICA outperforms a cluster procedure based on the results of Group ICA plus dual regression in terms of uncovering the underlying clusters. C-ICA yields comparable results to the two-step method of Durieux and Wilderjans (2019) for the current simulation

study. A possible reason for these very good recovery results could be that clusters were generated in such a way that they did have a very small amount of spatial overlap in RSNs and thus are relatively easy to uncover from the data and to separate from each other. The relatively good performance of Group ICA plus dual regression and the two-step clustering method can also be attributed to this. In the third simulation study, therefore, we will generate data with noise that obfuscates the true clustering and with cluster with more spatial overlap in RSNs. We expect that the C-ICA performance will drop a (little) bit but that the performance of the other two clustering methods will deteriorate drastically.

4.2. Simulation study 2

4.2.1. Problem

In the second simulation study, C-ICA is evaluated on more realistically simulated resting-state fMRI data and noise structures. As in the previous simulation study, data were generated according to the C-ICA model with a (known) true number of clusters R and a true number of RSNs Q. Contrary to the first simulation study, the simulated time courses were not randomly generated numbers from a uniform distribution, but simulated rs-fMRI time courses that are often observed in empirical data. Moreover, instead of randomly generated, the RSNs were simulated based on four existing functional connectivity networks that are described in Beckmann et al. (2005). Finally, more realistic error structures were used to generate the noise in the data. In particular, besides Gaussian noise (as in the first simulation, see Section 4.1), also autoregressive temporal and spatial noise structures were adopted. In this simulation, it is evaluated whether C-ICA – with the true R and Q – is able to recover the clustering of the patient data blocks ($\mathbf{P}$), the cluster specific RSNs ($\mathbf{S}^r$) and the subject specific rs-fMRI time courses ($\mathbf{A}_i$) underlying the simulated data.

4.2.2. Procedure

For this simulation study, we fixed the following simulation parameters: the number of subjects I at 30, the number of clusters R at 3, the number of true shared RSNs Q at 4, the number of voxels V at 4761 and the number of time points (mixtures) T at 150. Further, we only adopted equally sized clusters of 10 subjects each and only used non-square mixing matrices $\mathbf{A}_i$ (of size 150×4). To generate realistic RSNs, we used $Q = 4$ template RSNs from Beckmann et al. (2005)[7] that correspond with known disease related brain networks (for a visualization of these RSNs, see the most left column of Fig. 4): the medial visual network, the right frontoparietal network, the left frontoparietal network and the default mode network. In order to create cluster specific RSNs $\mathbf{S}^r$ that differ between the $R = 3$ clusters, we manually removed some active voxels from each of the four template RSNs, herewith obtaining disrupted RSNs. As such, we created a cluster representing healthy subjects, a cluster with moderately diseased patients and a cluster with severely diseased patients, with the latter two clusters having less or more severely disrupted RSNs. More specifically, as can be seen in the three most left columns of Fig. 4, for the first "healthy" cluster (first column) we took the template RSNs. For the second "moderately diseased" (second column) and third "heavy diseased" cluster (third column) we deactivated (i.e., replaced with a value of zero) about 4.74% and 8.30% of the voxels (averaged over the four networks) from the four template RSNs, respectively. Note that the procedure of deactivating voxels might have resulted in RSNs being not fully independent anymore (Silva et al., 2014), which violates an assumption of both the ICA and C-ICA model (i.e., the components/RSNs being mutually statistically independent). As a consequence, this results – to some extent – in data being generated that do not fully comply with a C-ICA model. The current simulation study could therefore also be seen as a – to some extent difficult – test case for C-ICA.

The average modified-RV statistic (a matrix correlation statistic, see Smilde et al., 2009) between all pair-wise cluster specific RSNs equals .65, indicating that there is some spatial overlap between the simulated RSNs. More specifically, the modified-RV between the RSNs of the healthy and the moderately diseased cluster equals .80, between the moderately and the severely diseased cluster .63 and between the healthy and the severely diseased cluster .52.

Further, in line with the C-ICA model, we simulated $I = 30$ subject specific resting-state time courses $\mathbf{A}_i$ (each with a length of $T = 150$ scans) for each of the $Q = 4$ RSNs, yielding non-square mixing matrices. The time courses were generated with the *neuRosim* package (Welvaert et al., 2011) and a repetition time of 2 s was used. We linearly combined these subject specific time courses with the cluster specific RSNs. As we used an equal cluster size of 10 subjects per cluster, this resulted in 30 noise-free data sets $\mathbf{T}_i$.

Next, we generated a noise structure that is – more – realistic for rs-fMRI data and added this noise to the noise-free data sets $\mathbf{T}_i$. In particular, we added either a Gaussian or a first-order autoregressive noise structure (AR1) to the data. For the Gaussian noise structure, we added (and scaled) Gaussian noise in the same way as we did for the first simulation study (see Section 4.1) such that in total 10%, 30% or 70% of the data pertains to noise. The noise percentage levels of 10%, 30% and 70% used in the current simulation study refer to a signal to noise ratio of 9, 2.33 and .43, respectively. In order to simulate a realistic autoregressive noise structure, we generated both an AR(1) structure in the temporal domain (i.e., column-wise by simulating V voxels with an autoregressive time series of length T, which represents correlated time points for each voxel) and an AR(1) structure in the spatial domain (i.e., row-wise by generating T autoregressive time series of length V that indicate correlated voxels at each time point). In order to manipulate the amount of autocorrelation of the noise in the data, we varied the AR correlation parameter – which was kept always the same for the spatial and temporal noise – as $\rho = .5$ or .7. Note that setting ρ to 0 would result in simulating noise very similar to Gaussian noise. Next, in order to manipulate the amount of noise in the data, we scaled the spatial and temporal noise structure such that both sources of AR(1) noise have an equal sum of squared entries and we added both noise structures together. Subsequently, we scaled this aggregated noise matrix (i.e., the sum of spatial and temporal noise) and added this matrix to the noise-free data matrix $\mathbf{T}_i$, herewith assuming that both adjacent time points and adjacent voxels are autocorrelated to the same extent and both types of noise being equally strong. Here, we scaled the aggregated noise using the same levels as in the Gaussian noise case. That is, we scaled the aggregated AR(1) noise such that either 10%, 30% of 70% of the data pertains to AR(1) noise.

In sum, we generated noise-free data $\mathbf{T}_i$ and added several types of noise to this noise-free data. That is, we added a small, intermediate or large amount of Gaussian or AR(1) noise (with ρ being .50 or .70.) to the data. We replicated the noise generating part 10 times, keeping the generation of the $\mathbf{A}_i$'s and $\mathbf{S}^r$'s fixed. This resulted in a total of 2 (AR(1) ρ noise) $\times$ 3 (amount of noise) $\times$ 10 (replications) = 60 data sets with AR(1) noise, and 3 (amount of noise) $\times$ 10 (replications) = 30 data sets with Gaussian noise. These data sets were analysed with C-ICA using the true number of $R = 3$ clusters and $Q = 4$ RSNs and adopting 30 random starts. Next, similar as for the first simulation study, the true and estimated C-ICA parameters were compared and the recovery measures (i.e., ARI and Tucker congruence) were computed, herewith accounting for the ambiguities of the C-ICA model (see Appendix B).

4.2.3. Results

From Table 4, which displays the average recovery values for each level of the two manipulated factors, one can see that for each noise structure C-ICA recovers the true underlying subject partition perfectly (i.e., mean ARI = 1). Further, C-ICA uncovers the true RSNs $\mathbf{S}^r$ to a very

[7] These networks can be downloaded from http://www.fmrib.ox.ac.uk/datasets/royalsoc8/.

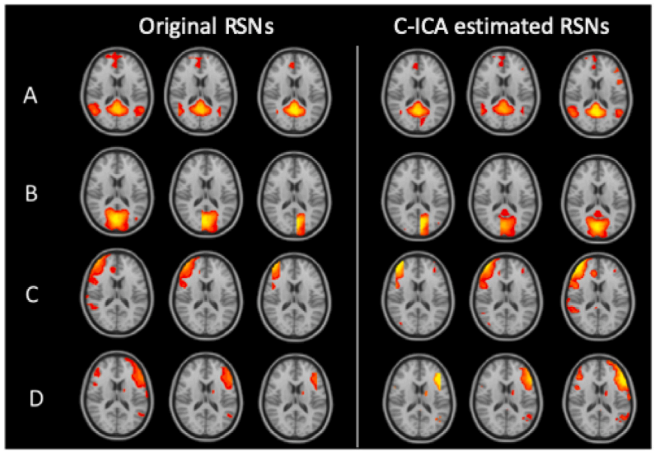

Fig. 4. Axial images (taken at coordinate $z = 45$) of four (A–D, in the rows) cluster specific true RSNs (first three columns) and C-ICA estimated cluster specific RSNs (last three columns) for $R = 3$ clusters. Each column refers to one of the $R = 3$ clusters, with the first true cluster (column 1) denoting healthy subjects and the second and third true cluster (column 2–3) consisting of moderately and severely diseased patients with moderately and severely disrupted RSNs. The true RSNs from the first (healthy) true cluster are templates taken from Beckmann et al. (2005) and refer to four brain networks: (A) default mode network, (B) medial visual network, (C) right frontoparietal network and (D) left frontoparietal network. The true RSNs for the second and third cluster are obtained by removing some active voxels from the four templates. The estimated cluster specific RSNs (column 4–6) were obtained by analysing one of the simulated data sets to which 30% Gaussian noise was added (see Section 4.2) with C-ICA with true R and Q. Note that the label switching problem inherent to clustering is clearly demonstrated here as the first (column 4), second (column 5) and third (column 6) retained cluster corresponds to the third (column 3), second (column 2) and first (column 1) true cluster, respectively.

large extent as indicated by the mean Tucker congruence being .93 – .96. As an example, for a single selected data set with 30% Gaussian noise, in Fig. 4, the true (column 1–3) and estimated (column 4–6) cluster specific RSNs are displayed, which demonstrates that C-ICA is able to retrieve the true RSNs to a large extent. The mean Tucker congruence value decreases when the strength of the noise autocorrelation (ρ for AR(1) noise) and the percentage of noise increases, with this effect being a little bit stronger for Gaussian noise than for AR(1) noise. Finally, the results show that C-ICA is also able to recover the subject specific time courses $\mathbf{A}_i$ to a very large extent, with mean Tucker congruence values around .99. Similarly to the recovery of the RSNs, the decline in time courses recovery with increasing noise is smaller for the auto-regressive noise structure than for the Gaussian noise structure. In sum, the results of the second simulation study with more realistic rs-fMRI noise structures and true RSNs and time courses show that C-ICA accurately estimates the underlying subject partition and discloses the underlying cluster specific spatial maps and subject specific time courses to a very large extent.

When comparing the results of both simulation studies, it appears that in both studies the recovery of the true subject partition is (almost) perfect. The reason for this is that the true clusters are simulated such that they do not show a large amount of spatial overlap in RSNs, which facilitates their recovery. In particular, the average RV coefficient of RSNs across cluster pairs for all generated data sets equals .00 and .65 for the first and second simulation study, respectively. Further, in the second simulation study the RSNs are recovered to a slightly lesser extent than in the first simulation study (i.e., a mean Tucker congruence around .95 and .96, respectively). A possible explanation for this minor discrepancy between the results of both simulation studies is the different way in which the RSNs per cluster are generated. Whereas in the first simulation study RSNs were generated randomly (from a double exponential distribution), in the second study the RSNs are created by using existing RSN templates and deactivating voxels

Table 4
Mean ARI (cluster recovery) and Tucker's congruence (RSN and TC recovery) value (and standard deviation) for all levels of the manipulated noise structures of the second simulation study.

Factor	ρ_{AR}	Noise level	ARI	RSN recovery	TC recovery
AR	.50	10%	1 (0)	.9607 (.0001)	.9915 (.0004)
AR	.70	10%	1 (0)	.9607 (.0001)	.9914 (.0004)
AR	.50	30%	1 (0)	.9602 (.0001)	.9912 (.0004)
AR	.70	30%	1 (0)	.9602 (.0001)	.9909 (.0004)
AR	.50	70%	1 (0)	.9607 (.0001)	.9915 (.0004)
AR	.70	70%	1 (0)	.9593 (.0001)	.9905 (.0004)
Gaussian	–	10%	1 (0)	.9432 (.0001)	.9917 (.0004)
Gaussian	–	30%	1 (0)	.9425 (.0001)	.9914 (.0004)
Gaussian	–	70%	1 (0)	.9385 (.0003)	.9901 (.0004)
Overall			1 (0)	.9540 (.0090)	.9911 (.0006)

Note. – = no ρ_{AR} parameter was manipulated; AR = autoregressive noise, ARI = Adjusted Rand Index, RSN = Resting state network, TC = subject specific time courses.

from these templates. The voxel deactivation procedure used could – to a varying degree for each cluster – have resulted in a model assumption violation of the C-ICA model (i.e., statistically independent RSNs). This, in turn, could have resulted in a slightly suboptimal recovery of the simulated RSNs, however, without affecting the uncovering of the true clustering (i.e., ARI = 1 for all data sets). The effect on the recovery of the time courses is minimal (i.e., a Tucker congruence around .99) as the time courses are uncovered almost perfectly.

4.3. Simulation study 3

4.3.1. Problem

In the previous two simulation studies, the recovery of the clustering (and the other C-ICA parameters) was excellent and this had a two-fold reason. First, clusters were generated which did have (almost) no spatial overlap in RSNs and thus were relatively easy to detect. Second, random noise was added in which no information was present that obfuscated the true cluster structure, which also facilitated the discovery of these true clusters. Therefore, the goal of this simulation study is to test C-ICA in more difficult circumstances in which the true clusters are harder to uncover from the data. In particular, clusters that show a considerable amount of spatial overlap in RSNs will be generated and structured noise that is targeted at obfuscating the true cluster structure will be added to a varying degree. The performance of C-ICA in terms of cluster recovery will be evaluated in these more difficult circumstances. Moreover, for these harder clustering conditions, C-ICA will also be compared to two other clustering methods: the two step procedure of Durieux and Wilderjans (2019) and the Group ICA plus dual regression clustering procedure (see Fig. 5).

4.3.2. Procedure

To not make the simulation study too complex, we kept the following simulation parameters constant: the number of subjects I at 20, the number of equally sized clusters R at 2, the number of true shared RSNs per cluster Q at 4, the number of voxels V at 1000 and the dimension of the mixing matrices at $T = 50$ (i.e., non-square mixing matrices). Next, three factors were systematically varied in a full three-factorial design:

1. The amount of spatial overlap between cluster specific shared RSNs, at two levels: medium and large spatial cluster overlap;
2. The strength of the obfuscating noise that contains a cluster structure that obfuscates the true clusters, at two levels: equal obfuscating noise and larger obfuscating noise;
3. The amount of independent Gaussian noise, at two levels: 70% and 90%.

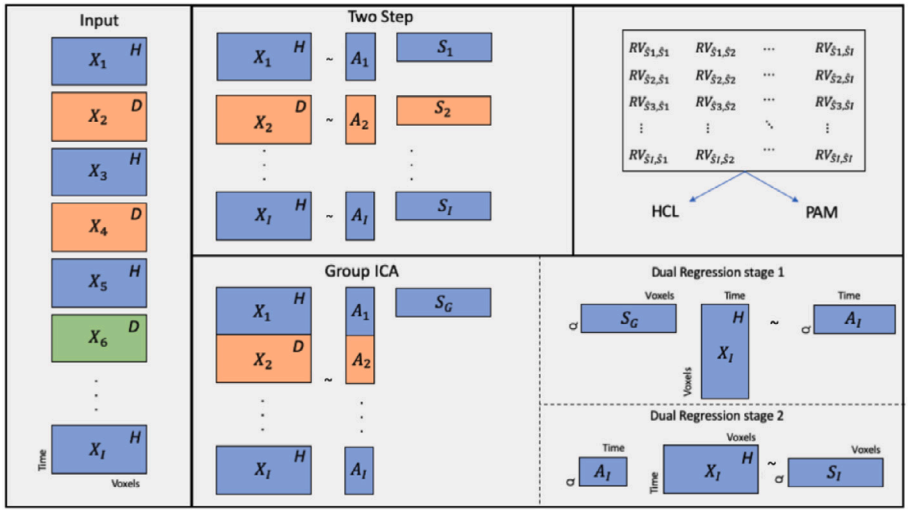

Fig. 5. Graphical illustration of the two step procedure (Durieux and Wilderjans, 2019) and a clustering procedure using Group ICA and dual regression estimated RSNs as input. In the left panel, Time × Voxel matrices $\mathbf{X}_i$ of multiple subjects i ($i = 1 \ldots I$) are presented. The different colours refer to the true, albeit unknown a priori, sub groups (i.e., clusters), whereas the letter in the top right corner of each matrix refers to the diagnostic label (which not necessarily perfectly matches with the true sub groups): H (healthy) or D (diseased). In the upper middle panel, a depiction of the two step procedure is given. Here, ICA is performed on each of the data matrices $\mathbf{X}_i$ separately in order to estimate subject specific RSN matrices $\mathbf{S}_i$. In the lower panel, a depiction of Group ICA including dual regression is given. Here the Time × Voxel matrices $\mathbf{X}_i$ of multiple subjects are temporally concatenated in order to estimate a group RSN matrix $\mathbf{S}_G$. These group RSNs are used in the first stage of dual regression in order to estimate time course matrices $\mathbf{A}_i$ for each subject. In the second dual regression stage, the subject specific time course matrices $\mathbf{A}_i's$ are used in order to estimate subject specific RSN matrices $\mathbf{S}_i$. In the upper right panel, a (dis)similarity matrix is constructed by computing the modified-RV coefficient between all possible pairs of the estimated $\mathbf{S}_i$ matrices (obtained by the two step or the Group ICA plus dual regression procedure). To obtain a subject clustering, a cluster analysis method such as hierarchical clustering using Ward's method or Partitioning Around Medoids is performed on this (dis)similarity matrix.

With respect to the first factor, in order to obtain spatially overlapping RSNs between clusters, we adopted a similar data generation scheme as used in Durieux and Wilderjans (2019). We first generated a set of RSNs $\mathbf{S}^{base}$ by sampling from a uniform distribution $U(-1, 1)$. Note that, although RSNs usually follow a super-Gaussian distribution, the current simulation study uses RSNs that are generated from a sub-Gaussian distribution. This, however, is not an issue for C-ICA since a logcosh contrast function is used within the fastICA estimation procedure, which is a well performing general purpose contrast function for the estimation of both super- and sub-Gaussian RSNs (Hyvärinen, 1997).

Next, a temporary RSNs $\mathbf{S}^r_{temp}$ matrix was sampled for each cluster ($r = 1, \ldots, R$) from a uniform distribution $U(-1, 1)$; next, these temporary RSNs matrices were added to $\mathbf{S}^{base}$: $\mathbf{S}^r = \mathbf{S}^{base} + w\mathbf{S}^r_{temp}$. By varying w, spatial cluster overlap was manipulated. For this study, we choose a w value of .23 and .15 as a pilot study indicated that using these weights results in an average modified-RV coefficient between cluster specific RSNs of .90 and .95, respectively (hence the factor levels; medium and large spatial cluster overlap).

Regarding the second factor, obfuscating noise was added by generating a C-ICA model with $Q = 20$ and $R = 20$. Note that in this case $R = I$ and therefore this obfuscating noise (only) consists of subject specific RSNs which do not contain any clustering information. We added this type of noise in order to obfuscate the true clustering, which is defined by the (spatially overlapping) cluster specific RSNs that are mentioned above, and make it harder for C-ICA to retrieve the true clusters. As such, the data generating model for a subject ($\mathbf{X}_i$) is defined as:

$$\mathbf{X}_i = \sum_{r=1}^{R} p_{ir}\mathbf{A}_i^{signal}\mathbf{S}^r + \mathbf{A}_i^{noise}\mathbf{S}_i + \mathbf{E}_i \tag{5}$$

where $\mathbf{A}_i^{signal}$ denotes the time courses of subject i that are used to mix the cluster specific RSNs $\mathbf{S}^r$ (of the cluster C_r to which subject i belongs) and $\mathbf{A}_i^{noise}$ denotes the time courses of subject i used for mixing the subject specific RSNs $\mathbf{S}_i$ of subject i. Similar as in the second simulation study, the time courses $\mathbf{A}_i^{signal}$ and $\mathbf{A}_i^{noise}$ were generated with the *neuroSim* package (Welvaert et al., 2011). The $\mathbf{S}^r$ and $\mathbf{S}_i$ matrices were generated by sampling random numbers from U(−1,1). Note that Eq. (5) can be written as $\mathbf{X}_i = \sum_{r=1}^{R} p_{ir}\,\mathbf{X}_i^{signal,r} + \mathbf{X}_i^{noise} + \mathbf{E}_i$. Here $\mathbf{X}_i^{signal,r}$ represents the mixed data of subject i based on the cluster specific RSNs $\mathbf{S}^r$ and thus contains information about the true cluster structure, whereas $\mathbf{X}_i^{noise}$ represents the mixed data based on the subject specific RSNs of subject i and contains obfuscating clustering information. Note further that $20 + 4 = 24$ RSNs are present in each subject's data and that only $Q = 4$ of these RSNs are shared across subjects, which is somewhat unrealistic as normally a larger proportion of the RSNs present in the data is expected to be shared across subjects. However, we believe that increasing the number of shared RSNs across subjects will not substantially change the performance results.

In order to vary the strength of the obfuscating noise, we scaled $\mathbf{X}_i^{signal,r}$ and $\mathbf{X}_i^{noise}$ by a factor $\sqrt{\frac{2000\times(1-z)}{\|\mathbf{X}_i^{signal,r}\|^2}}$ and $\sqrt{\frac{2000\times z}{\|\mathbf{X}_i^{noise}\|^2}}$, respectively, and manipulated z with the following values: $z = .5$ or .6 (corresponding to the factor levels equal and larger strength of the obfuscating noise). As a consequence, when $z = .5$, both $\mathbf{X}_i^{signal,r}$ and $\mathbf{X}_i^{noise}$ have a sum of squares of 1000, whereas a z value of .6 results in the sum of squares of $\mathbf{X}_i^{noise}$ being larger than the sum of squares of $\mathbf{X}_i^{signal,r}$ (i.e., 1200 vs. 800), which means that the obfuscating noise will mask the true cluster structure to a large extent.

Similarly as in the previous simulation studies, after scaling the two matrices $\mathbf{X}_i^{signal,r}$ and $\mathbf{X}_i^{noise}$, a noise matrix $\mathbf{E}_i$ is added in such a way that it has the desired percentage of independent Gaussian noise of either 70% or 90% (i.e., low signal to noise ratios of .43 and .11 respectively)

Lastly, for each of the 8 cells in the three-factorial design (i.e., 2 amount of spatial cluster overlap × 2 strength of structured noise × 2 noise levels), 10 replication data sets were generated, resulting in 80 simulated data sets. Each data set is analysed with C-ICA selecting for $R = 2$ clusters and $Q = 4$ RSNs, which equals the amount of shared true RSNs across subjects in the data, and with 30 random and (pseudo-) rationally defined starts. The rationally defined starts (2 in total) are determined by the two-step clustering procedures and the Group ICA plus dual regression clustering procedure. Pseudo-rationally defined starts (28 in total) are determined by perturbating 10% of the cluster labels from the rational starts. Additionally, in order to investigate the

effect of an overestimation of the number of RSNs, we also analysed each data set selecting for $R = 2$ clusters, $Q = 10$ and $Q = 24$ RSNs (note that the true shared $Q = 4$). Finally, for comparative purposes, we also performed the clustering procedure described in Durieux and Wilderjans (2019) and a Group ICA plus dual regression clustering procedure (see Section 4.1.2.) on these data sets. Here, we selected for $Q = 4$, $Q = 10$ and $Q = 24$ RSNs and applied PAM and a hierarchical clustering procedure using Ward's criterion on the dissimilarity matrix, both selecting for $R = 2$ clusters (also see Fig. 5).

4.3.3. Results

In Table 5, the mean ARI (and standard deviation) between the true partition and the estimated partition obtained from C-ICA, the two step and the Group ICA plus dual regression clustering procedure is presented for each manipulated factor level of the design conditional on either $Q = 4$ (upper part), $Q = 10$ (middle part) and $Q = 24$ (bottom part). The overall means (bottom row; averaged across all simulated data sets and values of Q used in the analysis) indicate that C-ICA (mean ARI of .637) outperforms both the Group ICA plus dual regression (mean ARI of .225 and .179 for HCL and PAM, respectively) and the two step (mean ARI of .517 and .431 for HCL and PAM, respectively) clustering procedure. When the true number of shared RSNs ($Q = 4$) is selected, C-ICA clearly outperforms (mean ARI of .695) the two step procedure (mean ARI of .329 and .327 for HCL and PAM, respectively) and to a lesser extent the Group ICA plus dual regression clustering procedure (mean ARI of .623 and .495 for HCL and PAM, respectively). In general, the recovery of the clustering deteriorates when the amount of independent Gaussian noise increases, when the spatial cluster overlap increases and when the strength of the structured noise becomes larger.

When selecting more RSNs ($Q = 10$ or $Q = 24$) than the true number of shared RSNs ($Q = 4$), C-ICA has a similar performance as the two step procedure using HCL when $Q = 10$ (mean ARI of .594 versus .600, respectively) but slightly outperforms the two step procedure using HCL when $Q = 24$ (mean ARI of .662 vs. .621). For both Q values, C-ICA has a better performance than the two step procedure using PAM (mean ARI of .496 and .470 for $Q = 10$ and $Q = 24$, respectively). The performance of the Group ICA plus dual regression for both HCL and PAM completely fails (mean ARI's close to 0 for each condition and $Q = 10$ or $Q = 24$). This indicates that an overestimation of the number of RSNs for this procedure in these difficult conditions, results in a random clustering that does not map the true clustering at all. Group ICA plus dual regression works relatively well with $Q = 4$ and for the condition with a larger strength of the structured noise since it concatenates all subjects and the sum of squares of shared RSNs is accumulated and this is retained in the group PCA results. On the other hand, the failure for Group ICA plus dual regression when spatial overlap in RSNs is large is likely due to cluster differences being too small and therefore the clustering fails. The good performance of the two step procedure with HCL suggests to always run C-ICA with a (too) large Q. It should be noted however that (1) selecting a (too) large Q increases the computation time of all clustering procedures, thus also for C-ICA in which the computationally intensive fastICA has to be run R times per iteration (see Section 3.2)[8] and (2) the true number of RSNs is never known (i.e., even a large Q may be smaller than the true (shared) Q, which is a situation where the two step procedure fails and C-ICA may fail due to an under-specification of the shared Q). Note further that often the two step procedure with HCL is performed anyways as it is often used to obtain a good (rational) starting partition for C-ICA (see Section 3.2). Similar as in the $Q = 4$ condition, an alike pattern for the manipulated factors occurs for C-ICA and the two step procedure. That is, cluster recovery deteriorates when the amount of independent Gaussian noise increases, when the spatial cluster overlap increases and when the strength of the structured noise becomes larger.

In sum, results of this simulation study show that C-ICA performs – compared to the other clustering procedures – well under very difficult circumstances (i.e., large amounts of independent Gaussian noise, structured subject specific noise and spatial cluster overlap). Moreover, given that the true number of RSNs Q is always unknown, C-ICA is the safest option as it performs the best when the data are analysed with the true (shared) Q and as good as the two step procedure using HCL when a larger Q is selected.

4.4. Simulation study 4

4.4.1. Problem

In this subsection, the C-ICA method is evaluated on a sub selection of the subjects of an empirical multi-subject rs-fMRI data set (for a full description of this data set, see Section 5). This empirical data set contains samples of two subpopulations, namely, patients that are diagnosed with Alzheimer's disease (AD, 77 patients) and elderly control subjects (EC, 173 subjects). In order to evaluate the performance of C-ICA in retrieving an underlying cluster structure, we took a subsample of the subjects of this empirical data set in such a way that two groups of subjects were retained that differ the most as possible in RSNs. More precisely, we selected the 20 AD patients and 20 EC subjects that show the largest differences (between groups) in underlying RSNs. The rationale behind this procedure is that we now can apply C-ICA to an actual rs-fMRI data set – instead of simulated data as before – in which – due to the sub selection – a clear underlying cluster structure is present. Note that the information regarding the true (clinically defined) cluster structure, although known in this case, will not be used in the clustering analysis.

4.4.2. Procedure

In order to select two groups of subjects that differ the most (between groups) in terms of RSNs from the empirical data set, we performed a Group ICA on the group of AD patients and EC subjects separately. Here, for each group, data were concatenated along the temporal dimension and ICA was performed and 20 RSNs were extracted, resulting in a set of RSNs that are representative for the group of AD patients and EC subjects separately. Next, subject specific RSNs were extracted by applying a single-subject ICA with 20 RSNs to each subject's data block separately (i.e., 250 ICA analyses in total). Then, for each subject, we computed the mean Tucker congruence between the subject specific RSNs for the subject in question and the representative Group ICA RSNs for both groups, which yielded two mean Tucker congruence values per subject; these Tucker values indicate how good the RSNs of a subject resemble the RSNs of each group (i.e., AD or EC) of subjects. The 20 AD patients and 20 EC subjects that had the largest absolute difference between these two mean Tucker congruence values were retained for the current analysis. As such, subjects were selected that are representative for one group of subjects (i.e., AD or EC) but are the most as possible dissimilar from the other group of subjects in terms of RSNs.

The data sub selection with 20 AD patients and 20 EC subjects was analysed with C-ICA, extracting $Q = 20$ RSNs and $R = 2$ clusters – which is the true number of clusters here – and using 30 random starts. Additionally, for comparative purposes, the data was also analysed with the two step clustering procedure and the clustering procedure using Group ICA and dual regression estimated RSNs, both selecting for $R = 2$ clusters and different values of Q. More specifically, for the Group ICA with dual regression clustering procedure $Q = 5, 10, 15$ and 20 were selected and for the two step clustering procedure $Q = 20, 25, 30$ and 35 were selected. These selected values were chosen since, based on the results of the third simulation study (see Section 4.3),

[8] It is advised, especially when one also wants to do model selection (i.e., determining the optimal R and Q), to perform these analyses on a parallel computer infrastructure, where, for example, the different starts can be performed in parallel.

Table 5
Mean ARI (and standard deviation) for C-ICA, two step clustering and G-ICA DR clustering for all combinations of levels of the manipulated factors of the third simulation study.

	Q	C-ICA	GICA DR HCL	GICA DR PAM	Two step HCL	Two step PAM
Medium overlap	4	0.912 (0.178)	0.872 (0.279)	0.708 (0.322)	0.545 (0.373)	0.509 (0.369)
High overlap	4	0.478 (0.442)	0.373 (0.465)	0.283 (0.365)	0.114 (0.245)	0.144 (0.259)
SS equal	4	0.724 (0.373)	0.604 (0.473)	0.577 (0.410)	0.415 (0.423)	0.433 (0.401)
SS larger	4	0.667 (0.427)	0.641 (0.444)	0.414 (0.384)	0.244 (0.317)	0.220 (0.295)
70% noise	4	0.890 (0.304)	0.840 (0.361)	0.698 (0.394)	0.491 (0.420)	0.498 (0.404)
90% noise	4	0.500 (0.392)	0.406 (0.441)	0.293 (0.300)	0.168 (0.254)	0.155 (0.218)
Overall mean $Q = 4$	4	0.695 (0.400)	0.623 (0.456)	0.495 (0.403)	0.329 (0.381)	0.327 (0.366)
Medium overlap	10	0.798 (0.309)	0.059 (0.139)	0.072 (0.126)	0.783 (0.295)	0.655 (0.347)
High overlap	10	0.390 (0.455)	0.002 (0.054)	0.009 (0.074)	0.416 (0.437)	0.338 (0.344)
SS equal	10	0.698 (0.412)	0.016 (0.091)	0.031 (0.082)	0.698 (0.399)	0.504 (0.385)
SS larger	10	0.489 (0.441)	0.045 (0.123)	0.050 (0.128)	0.501 (0.410)	0.489 (0.377)
70% noise	10	0.875 (0.291)	0.014 (0.068)	0.035 (0.102)	0.879 (0.273)	0.777 (0.301)
90% noise	10	0.312 (0.374)	0.047 (0.136)	0.046 (0.114)	0.320 (0.335)	0.216 (0.196)
Overall mean $Q = 10$	10	0.594 (0.437)	0.030 (0.108)	0.041 (0.108)	0.600 (0.414)	0.496 (0.379)
Medium overlap	24	0.832 (0.303)	0.023 (0.078)	0.000 (0.062)	0.823 (0.265)	0.630 (0.401)
High overlap	24	0.412 (0.447)	0.020 (0.074)	0.003 (0.078)	0.420 (0.420)	0.310 (0.313)
SS equal	24	0.690 (0.386)	0.013 (0.082)	−0.008 (0.055)	0.665 (0.398)	0.487 (0.391)
SS larger	24	0.554 (0.473)	0.030 (0.069)	0.011 (0.081)	0.578 (0.418)	0.453 (0.397)
70% noise	24	0.895 (0.244)	0.024 (0.069)	−0.006 (0.059)	0.895 (0.208)	0.761 (0.292)
90% noise	24	0.350 (0.413)	0.019 (0.083)	0.009 (0.079)	0.348 (0.366)	0.179 (0.227)
Overall mean $Q = 24$	24	0.662 (0.435)	0.021 (0.076)	0.001 (0.070)	0.621 (0.404)	0.470 (0.392)
Overall mean		0.637 (0.424)	0.225 (0.392)	0.179 (0.331)	0.517 (0.420)	0.431 (0.385)

Note. Overlap = spatial cluster overlap, SS = strength of the structured noise (.6 and .5 for SS larger and SS equal, respectively), Noise = independent Gaussian noise, Q = number of cluster specific RSNs, ARI = Adjusted Rand Index, C-ICA = Clusterwise Independent Component Analysis, HCL = hierarchical clustering (Ward's criterion), PAM = Partitioning Around Medoids, Two step = two step method of Durieux and Wilderjans (2019), GICA DR = Group ICA plus dual regression.

it can be conjectured that the two step procedure might benefit from selecting a relatively large number of RSNs, whereas Group ICA with dual regression might benefit from extracting a relatively low number of RSNs.

4.4.3. Results

Table 6 shows a cross-tabulation of the true partition versus the estimated C-ICA partition (with $R = 2$ and $Q = 20$ on the empirical data sub selection). As can be seen in this confusion table, 19 AD patients are allocated to cluster A and 21 subjects – consisting of 20 EC subjects and 1 AD patient – are allocated to cluster B. Comparing the true to the estimation partition gives a balanced accuracy (García et al., 2009) of .976 and an ARI of .90, which implies that C-ICA uncovered the true subject partition almost perfectly (i.e., 1 AD patient is erroneously marked as an EC subject). Note that the analysis is completely unsupervised as no information about the diagnostic labels was used in the C-ICA estimation. An interpretation of the obtained subject clusters and cluster specific RSNs is left for Section 5.

When comparing C-ICA to the other clustering procedures, it appears that C-ICA clearly outperforms these other procedures in terms of uncovering the true clustering. In particular, the ARI for the two step clustering procedure (Durieux and Wilderjans, 2019) using Ward's hierarchical clustering and PAM, selecting for $Q = 20$, equals .81 and .35, respectively. When the number of RSNs was increased in steps of 5 for the two step procedure (i.e., $Q = 25$, 30 and 35) the performance of the two step procedure slightly increased for Ward's hierarchical clustering (ARI = .90 for every selected Q, the same results as for C-ICA selecting for $Q = 20$ RSNs) but completely failed for PAM (ARI = 0 for every selected Q). As such, the two step procedure using Ward's method is able to find a similar subject clustering as C-ICA but needs a more complex (i.e., less parsimonious) model with more RSNs to obtain this. For the Group ICA plus dual regression clustering procedure, the ARI for both Ward's hierarchical clustering and PAM, selecting for $Q = 20$, equals -.015. The same pattern occurred when a lower number of RSNs was selected. In particular, the ARI for Group ICA with dual regression using Ward's hierarchical clustering equals −0.013, −0.014 and −0.015 for $Q = 5$, 10 and 15, respectively. For Group ICA with dual regression adopting PAM, the ARI equals 0, 0 and 0.805, respectively.

Table 6
Confusion matrix of the diagnostic labels (AD versus EC; in the rows) against the estimated clustering (labelled as A and B; in the columns) from a C-ICA with $Q = 20$ and $R = 2$ on the data sub selection (fourth simulation study).

	A	B	Sum
AD	19	1	20
EC	0	20	20
Sum	19	21	40

The latter solution has two subjects that were erroneously clustered when compared to the diagnostic labels. The results show that Group ICA with dual regression does not retrieve – except when $Q = 15$ is selected in conjunction with PAM clustering – the relevant clustering information in the data, resulting in clusters that do not resemble the diagnostic (patients) groups at all.

It can be concluded that when a clear cluster structure with substantial differences in underlying RSNs is present in the data, C-ICA is able to uncover this subject clustering, whereas other clustering procedures that make use of ICA (e.g., Group ICA with dual regression) do this to a (way) lesser extent or need a more complex solution with a larger Q to do so.

4.5. Evaluating the sequential model selection procedure

4.5.1. Problem

To perform C-ICA, one has to specify the number of clusters R and RSNs Q a priori to the C-ICA analysis. However, often no prior information regarding the optimal number of clusters and/or RSNs is available and thus the optimal R and Q should be estimated based on the data set at hand by using some model selection procedure. Therefore, in this fifth study, it is investigated whether the proposed sequential model selection procedure (see Section 3.3) is able to select the optimal number of clusters R and RSNs Q underlying a data set. In this procedure, first the optimal number of clusters R is determined and next, conditional on the optimal R, the optimal number of RSNs Q is estimated.

Table 7
Result of the sequential model selection procedure applied to the selected simulated data set from the second simulation study. Computed scree ratios $sr_{r|q}$ from step 1 of the procedure are displayed in the columns $R = 2$ to $R = 4$ for all values of Q (rows) and averaged across Q (bottom row). The computed scree ratios $sr_{q|R}$ – conditional on the optimal R – from step 2 of the procedure are presented in the last column.

	$R = 2$	$R = 3$	$R = 4$	Optimal Q
$Q = 2$	1.31	8.11	2.45	–
$Q = 3$	1.30	4.95	4.70	1.55
$Q = 4$	1.32	64.10	1.05	**111.27**
$Q = 5$	2.53	21.16	1.06	1.01
$Q = 6$	3.88	8.60	1.06	1.01
$Q = 7$	5.81	3.48	1.21	–
Average	2.69	**18.40**	1.92	

Note. In the first step of the sequential model selection procedure, the scree ratios are not defined for $R_{min} = 1$ and $R_{max} = 5$, whereas, similarly for the second step, the scree ratios are not defined for $Q_{min} = 2$ and $Q_{max} = 7$. The largest average scree ratio $sr_{r|q}$ is highlighted in bold (bottom row), with the corresponding $R = 3$ denoting the optimal number of clusters. The largest scree ratio $sr_{q|R}$ conditional upon the optimal R is highlighted in bold (last column), indicating that $Q = 4$ is the optimal number of RSNs.

Table 8
Mean variance accounted for (VAF) and standard deviation (between brackets), computed across all simulated data sets from two design cells (Upper panel: 30% Gaussian noise; lower panel: 70% Gaussian) of the second simulation study, for each number of RSNs Q and number of clusters R.

	R = 1	R = 2	**R = 3**	R = 4	R = 5
Q = 2	44.84 (0.61)	47.98 (0.59)	50.28 (0.62)	50.47 (0.62)	50.54 (0.58)
Q = 3	52.77 (0.49)	56.42 (0.38)	59.00 (0.41)	59.31 (0.43)	59.40 (0.40)
Q = 4	57.61 (0.35)	61.73 (0.19)	**64.54 (0.22)**	64.58 (0.22)	64.62 (0.22)
Q = 5	60.22 (0.18)	63.40 (0.20)	64.59 (0.22)	64.65 (0.22)	64.70 (0.22)
Q = 6	61.64 (0.22)	64.01 (0.22)	64.64 (0.22)	64.71 (0.21)	64.78 (0.22)
Q = 7	62.68 (0.21)	64.38 (0.22)	64.69 (0.22)	64.77 (0.22)	64.85 (0.22)
	R = 1	R = 2	**R = 3**	R = 4	R = 5
Q = 2	19.27 (0.27)	20.64 (0.26)	21.65 (0.27)	21.75 (0.27)	21.82 (0.26)
Q = 3	22.69 (0.22)	24.29 (0.16)	25.44 (0.18)	25.61 (0.19)	25.70 (0.19)
Q = 4	24.79 (0.15)	26.59 (0.08)	**27.86 (0.10)**	27.96 (0.10)	28.05 (0.10)
Q = 5	25.93 (0.08)	27.39 (0.09)	27.98 (0.10)	28.11 (0.10)	28.23 (0.10)
Q = 6	26.57 (0.10)	27.67 (0.10)	28.09 (0.10)	28.21 (0.09)	28.35 (0.10)
Q = 7	27.04 (0.09)	27.88 (0.10)	28.06 (0.09)	28.21 (0.11)	28.33 (0.10)

Note. True underlying model: $R = 3$ and $Q = 4$.

4.5.2. Procedure

For this study, we selected the data sets from the second simulation study that contained a moderate amount – 30% – and a large amount – 70% – of Gaussian noise. We then applied C-ICA using 30 random starts to these data sets with a range of RSNs going from 2 to 7 and the number of clusters ranging from 1 to 5. We selected these ranges for R and Q as we know in this case that these ranges contain the correct R and Q, which is $R = 3$ clusters and $Q = 4$ RSNs. In practice, however, no information about the optimal number of clusters and RSNs is available and it is advised to investigate a larger range for the number of RSNs and clusters.

After obtaining the loss function value for each combination of a number of RSNs Q and clusters R, we applied the sequential model selection procedure as described in Section 3.3. Here, in the first step, the optimal number of clusters R^{opt} is determined by computing the $sr_{r|q}$ values from Eq. (3) and selecting the R that yields the largest mean $sr_{r|q}$ value, with the mean computed across all considered values of Q. Next, in the second step, the optimal number of RSNs Q^{opt} is selected conditional on the optimal number of clusters R^{opt} by selecting the Q that maximizes the $sr_{q|R}$ scree ratio from Eq. (4).

4.5.3. Results

For demonstration purposes, we first show the working and result of the sequential model selection procedure when applied to one of the data sets with 30% Gaussian noise. Next, the results for all analysed data sets are presented. Table 7 shows the computed scree ratios $sr_{r|q}$ (columns $R = 2$, $R = 3$ and $R = 4$) for each value of Q (rows) and averaged across Q (bottom row). Note that the scree ratios for $R_{min} = 1$ and $R_{max} = 5$ are not displayed as these ratios are undefined, implying that the sequential model selection procedure does not allow to select the smallest and largest considered value for R (Wilderjans et al., 2013). The largest averaged scree ratio (displayed in the bottom row of Table 7) indicates the optimal number of clusters R, which in this case is $R = 3$ clusters. To select the optimal number of RSNs Q, conditional on the optimal number of clusters $R = 3$ (i.e., the second step of the model selection procedure), the $sr_{q|R}$ values from Eq. (4) are displayed in the last column of Table 7. Note that the scree ratios for $Q_{min} = 2$ and $Q_{max} = 7$ are not defined and are thus not displayed. It appears that, as $sr_{q|R}$ is maximal for $Q = 4$, the optimal number of RSNs Q equals four. In sum, the sequential model selection procedure applied to the selected simulated data set retains the solution with $R = 3$ clusters and $Q = 4$ RSNs. As such, the sequential model selection procedure selected the correct solution as this solution has the true number of clusters and RSNs for this data set.

In Table 8, the mean variance accounted for (VAF) and standard deviation, computed across the 20 analysed data sets, is displayed for each combination of Q and R.[9] Note that the VAF is displayed instead of the original loss function values since the VAF values are more intuitive due to their 0%–100% range (with the VAF and the original loss function values being inversely proportionally related to each other). Note that the VAF values in Table 8 are increasing relatively fast from a less complex model (i.e., $Q < 4$ and $R < 3$) to the true model ($Q = 4$ and $R = 3$), compared to the increase in VAF for more complex models (i.e., $Q > 4$ and $R > 3$). In other words, the proposed model selection procedure automatically searches for this 'tipping point'. For all 20 simulated data sets, the model selection procedure indicated that the optimal number of RSNs equals 4 (i.e., $Q = 4$) and the optimal number of clusters equals 3 (i.e., $R = 3$). These results show that the proposed model selection procedure may be a valuable tool for selecting the optimal number of RSNs and clusters for a C-ICA analysis.

5. Illustrative application of the C-ICA model

5.1. Motivation and data

In this section, the proposed C-ICA model will be illustrated on an empirical multi-subject rs-fMRI data set that consists of a total of 77 clinically diagnosed Alzheimer patients (AD) and 173 elderly controls (EC). The goal of the current application is to cluster the subjects based on similarities and differences in the RSNs and time courses underlying their data. It has to be stressed that this analysis is performed completely unsupervised, implying that no information about the diagnostic labels is used. After clustering the subjects, the diagnostic labels will only be used to interpret and validate the derived clustering (i.e., classification is not the goal of this analysis). To further validate the obtained clustering, an ad hoc procedure is adopted in which the cluster specific RSNs in each cluster are matched to eight known template RSNs described in Beckmann et al. (2005).

The AD patient data was collected in the prospective registry on dementia (PRODEM, see Seiler et al., 2012). Participants from the control group were scanned as part of the Austrian Stroke Prevention Family Study. For both groups, an equal scanning protocol was used and the scanning took place at the same site (i.e., the Medical University of Graz). For a more thorough description of the data, see de Vos et al. (2018).

Participants were scanned on a Siemens Magnetom TrioTrim 3T MRI scanner. For the rs-fMRI session, 150 volumes were acquired with

[9] The VAF can be computed as follow; $VAF = \frac{\|\mathbf{X}\|^2 - L}{\|\mathbf{X}\|^2} \times 100$, where L is the loss function value of a C-ICA analysis and $\|\underline{\mathbf{X}}\|^2$ the total sum of squares of a multi-subject data set.

TR = 3000 ms, TE = 30 ms, flip angle = 90°, 40 axial slices and an isotropic voxel size of 3 mm. During the rs-fMRI session, participants were instructed to lie still with their eyes closed and to stay awake. No tasks were performed before acquiring the rs-fMRI data and, therefore, no task related after-effects could potentially compromise the RSNs (Grigg and Grady, 2010).

The data were pre-processed using the *FMRIB* Software Library (FSL, version 5.0; Jenkinson et al., 2012; Smith et al., 2004). Here, for the rs-fMRI data, the pre-processing steps included brain extraction and motion correction (Jenkinson et al., 2002), temporal high-pass filtering with a cut-off point of 100 s and 3 mm *FWHM* spatial smoothing. Moreover, the ICA-based *Xnoiseifier* (*FIX*, version 1.06) from *FMRIB* with the standard provided training data was used in order to automatically identify and remove statistically independent noise components from the data (Salimi-Khorshidi et al., 2014). The subject's fMRI scans were registered to MNI152 space in a two step procedure. In a first step, the subject's fMRI scans were linearly registered to the subject's structural MRI scan, only allowing for translation and rotation (in x, y and z direction). In a second step the subject's structural MRI was nonlinearly registered to MNI152 space. These two steps were then combined to register the fMRI volumes to MNI152 space. These steps resulted in pre-processed data (in MNI space) with a voxel size of $2 \times 2 \times 2$ mm.

Lastly, to make the analysis computationally more feasible, we down sampled the data of each subject using the *subsamp2* command from *fslmath*. This means that the dimension of each voxel changed from $2 \times 2 \times 2$ mm (i.e., the original size) to $8 \times 8 \times 8$ mm. Additionally, a mask is applied such that only voxels that are part of the brain are included in the analysis, resulting in a total of 2553 voxels being used for this analysis. Finally, as noted in Section 3.1, we applied (columnwise) mean-centring, resulting in each voxel having an average activation of zero over time, to the data of each subject. Moreover, by block scaling the data we ensured, in order to remove trivial scale differences between subjects, that each subject's data block had the same variability (i.e., equal sum of squared values).

5.2. Procedure

We applied C-ICA on the data set with Q = 20, 25, 30 and 35 RSNs and a range of number of clusters R going from 1 to 5. We initialized the C-ICA algorithm with 1 rational and 30 random starts (see step 1 in Section 3.2). For the rational start, we used the subject partition obtained by the clustering procedure described in Durieux and Wilderjans (2019). Here, a single-subject ICA with Q RSNs is performed on the data of each subject and the similarity in RSNs, quantified by the modified RV-coefficient (Smilde et al., 2009), is calculated for each subject pair. A rational subject partition is obtained by applying hierarchical clustering using Ward's method to these (dis)similarities and cutting the resulting dendrogram such that R clusters are retained. For each C-ICA analysis, the run with the lowest C-ICA loss function value (see Eq. (2)) was kept as the final solution. These optimal loss function values for all considered number of RSNs Q and number of clusters R were used in the sequential model selection procedure (see Section 3.3 and the simulation example from Section 4.4) in order to determine the optimal number of clusters and RSNs for this data set. To validate the clustering results (i.e., the neurofunctional subtypes), the obtained subject clusters (i.e., for each model with $R = 2$ to $R = 5$) are compared to the diagnostic labels (i.e., AD versus EC). Additionally, to validate the estimated cluster specific RSNs, we applied an ad hoc procedure where we matched the estimated RSNs for each cluster to a template with eight known RSNs that are thoroughly described in Beckmann et al. (2005). As such, a one-to-one matching of RSNs across clusters is obtained, which facilitates the interpretation of the cluster specific RSNs.

5.3. Results

5.3.1. Clustering P

The result of the sequential model selection procedure, which is presented in Table 9, indicates that the optimal C-ICA solution is a model with $R = 2$ clusters and $Q = 25$ RSNs. This optimal C-ICA solution was obtained from the C-ICA run in which the algorithm was seeded with the subject clustering from the two step procedure using Ward's method (i.e., rational start). Note that the optimal solution has a loss function value of 175 144.2 (after ALS iterations), whereas the solution obtained from the two step procedure adopting Ward's method (i.e., the solution before ALS iterations) has a clearly larger loss value of 175 354.9. The C-ICA algorithm needed 5 ALS iterations to converge to this optimal solution. As such, the optimal C-ICA solution is better (i.e., lower loss value) than the subject partition obtained from the two step clustering procedure. Table 10 shows four crosstabulations in which the diagnostic labels are presented against the derived clustering for $R = 2, \ldots, 5$ clusters with $Q = 25$ RSNs. As can be seen in the upper left panel ($R = 2$) of this table, C-ICA allocates predominately elderly control (EC = 139) subjects to cluster A and predominately Alzheimer patients (AD = 51) to cluster B. Again, note that the allocation of the subjects by C-ICA is done completely unsupervised, that is, no information of the diagnostic labels is used when estimating the clusters. Comparing the obtained clusters to the diagnostic labels, the balanced accuracy (BA) equals .72 and the ARI equals .26. Note that we used the balanced accuracy – instead of the regular accuracy – since this measure is more suitable for situations with imbalanced classes (García et al., 2009). Both measures show a clear above chance performance and indicate that C-ICA retains clusters that can be matched to the diagnostic labels to a substantial extent. Also note that the unsupervised C-ICA method (BA of .72) retrieves the diagnostic labels almost as good as a supervised penalized regression classifier (BA of .79) that uses derived resting state fMRI features (e.g., functional connectivity between brain regions or amplitude of low frequency fluctuations) computed from the same data set as input (de Vos et al., 2018). The obtained C-ICA clustering, however, does not perfectly match the diagnostic labels, which should not be surprising given the noisy nature of rs-fMRI data and the fact that the obtained clustering is based on differences in RSNs, whereas mainly clinical information is used to obtain the labels (i.e., separate AD from EC).

As a comparison, we also employed the two step (Durieux and Wilderjans, 2019) and the Group ICA plus dual regression clustering procedure mentioned in Section 4.3 on the empirical data. Note that the two step clustering procedure was already performed in order to find the rational starts to seed the C-ICA algorithm. For both analyses, in order to make a fair comparison, we selected $R = 2$ clusters and $Q = 25$ RSNs. The results show that Group ICA plus dual regression yields a cluster structure that does not match the diagnostic labels at all (i.e., ARI equals .02 and .05 and the BA .60 and .60 for Ward's hierarchical clustering and PAM, respectively). Also the two step clustering procedure gives a subject clustering that matches the diagnostic labels only to a very small, although larger than Group ICA plus dual regression, extent (i.e., ARI of .15 and .01 and BA of .68 and .63 for HCL and PAM, respectively). Comparing the clustering retained by the three clustering procedures (see Table 11), it appears that the C-ICA clusters are very different then the clusters obtained by the Group ICA plus dual regression (ARI of .02 and .04 for HCL and PAM, respectively) and the two step (ARI of .29 and .08 for HCL and PAM, respectively) procedure. It can be concluded that, when selecting for $Q = 25$, C-ICA extracts a meaningful clustering, although not perfectly matching the diagnostic labels, from an empirical multi-subject fMRI data set, whereas the other two clustering procedures yield more random clusters that are hard to interpret.

Based on the results from the third simulation study (see Section 4.3), it can be expected that for Group ICA with dual regression

Table 9
Result of the sequential model selection procedure applied to the empirical data set. Computed scree ratios $sr_{r|q}$ from step 1 of the procedure are displayed in the columns $R = 2$ to $R = 4$ for all values of Q (rows) and averaged across Q (bottom row). The computed scree ratios $sr_{q|R}$ – conditional on the optimal R – from step 2 of the procedure are presented in the last column.

	$R = 2$	$R = 3$	$R = 4$	Optimal Q
$Q = 20$	2.38	0.84	0.67	–
$Q = 25$	2.32	0.83	0.54	**1.21**
$Q = 30$	2.18	0.74	0.48	1.11
$Q = 35$	2.02	0.74	0.51	–
Average	**2.26**	0.79	0.53	

Note. In the first step of the sequential model selection procedure, the scree ratios are not defined for $R_{min} = 1$ and $R_{max} = 5$, whereas, similarly for the second step, the scree ratios are not defined for $Q_{min} = 20$ and $Q_{max} = 35$. The largest average scree ratio $sr_{r|q}$ is highlighted in bold (bottom row), with the corresponding $R = 2$ denoting the optimal number of clusters. The largest scree ratio $sr_{q|R}$ conditional upon the optimal R is highlighted in bold (last column), indicating that $Q = 25$ is the optimal number of RSNs.

clustering a relatively lower number of RSNs may be beneficial for estimating the subject clustering. Conversely, selecting for a larger number of RSNs may be beneficial for the two step clustering procedure. Therefore, we also selected for $Q = 5$, 10, and 15 for the Group ICA with dual regression clustering procedure and $Q = 30$ and 35 for the two step clustering procedure. Results indicate that for the Group ICA with dual regression clustering procedure the clustering performance improves but stays below C-ICA's performance when $Q = 5$ (i.e., the ARI equals .13 and .12 for Ward's hierarchical clustering and PAM, respectively). No improvement in clustering performance is encountered for Group ICA with dual regression when $Q = 10$ (i.e., ARI of .03 and .07 for Ward's hierarchical clustering and PAM, respectively). Finally, when $Q = 15$, Group ICA with dual regression only improves the clustering – but still below the level of C-ICA – when PAM is used (i.e., ARI of .14 for PAM and ARI of 0 for Ward's hierarchical clustering). The associated BA values for the Group ICA with dual regression clustering solutions range from .03 to .65, which is clearly below the BA value obtained with C-ICA ($BA = .72$). When adopting the two step clustering procedure and selecting for a larger Q value, no improvement in the clustering is observed. In particular, nor using Ward's method (ARI of .06 and .09 for $Q = 30$ and $Q = 35$, respectively), nor adopting PAM (ARI of -.02 and -.04 for $Q = 30$ and $Q = 35$, respectively) yields a better subject partition (BA values range from .62 to .64). In conclusion, for these data, both the Group ICA with dual regression clustering procedure and the two step clustering procedure were not able to find a clustering that matches the diagnostic labels when selecting for either a relatively low number of RSNs (for Group ICA with dual regression clustering) or a relatively large(r) number of RSNs (for the two step clustering procedure).

We also investigated how the derived clustering for a C-ICA model with $R = 2$ and $Q = 25$ differs from C-ICA models with a lower number of RSNs ($Q = 20$) and larger number of RSNs ($Q = 30$ and $Q = 35$) for $R = 2$. To this end, we computed the ARI between the optimal model ($R = 2$ and $Q = 25$) and the abovementioned C-ICA models. Results indicate that the derived subject partition shows a large stability when Q is decreased or increased. In particular, the ARI for comparing the optimal model with $Q = 25$ to models with $Q = 20, 30$ and 35 equals .87, .79 and .76, respectively (BA values of .97, .93 and .92, respectively).

An interesting aspect of C-ICA is that one can investigate how the subject clustering changes when a larger number of clusters is retained. As can be seen in the upper right panel of Table 10, when $R = 3$, the estimated clusters A and C resemble the derived clusters from the two-cluster solution. However, a new cluster (cluster B) emerges with mainly elderly controls (i.e., 24 EC and 4 AD subjects). As this cluster almost exclusively contains EC subjects, this cluster can be considered as a "predominantly healthy" cluster, whereas cluster A can be conceived as a "subjects at risk" cluster. For $R = 4$, a small cluster D (i.e., 6 AD and 9 EC subjects) emerges, which contains a subset of subjects from cluster C from the three-cluster solution. For $R = 5$, an extra cluster with mainly AD subjects appears (i.e., cluster C with 15 AD and 5 EC subjects).

Table 10
Crosstabulations of the diagnostic labels (AD versus EC) against the subject clustering obtained by applying C-ICA to the empirical data set with $Q = 25$ RSNs and $R = 2$ (first table), $R = 3$ (second table), $R = 4$ (third table) and $R = 5$ (fourth table) clusters.

	A	B	Sum
AD	26	51	77
EC	139	34	173
Sum	165	85	250

	A	B	C	Sum
AD	21	4	52	77
EC	118	24	31	173
Sum	139	28	83	250

	A	B	C	D	Sum
AD	20	4	47	6	77
EC	116	24	24	9	173
Sum	136	28	71	15	250

	A	B	C	D	E	Sum
AD	14	4	15	38	6	77
EC	106	24	5	29	9	173
Sum	120	28	20	67	15	250

Note. AD = Alzheimer's disease patients, EC = elderly control subjects.

Table 11
Crosstabulation of the subject clustering obtained by C-ICA (always presented in the rows) against the subject clustering obtained by Two Step HCL (first table), Two Step PAM (second table), GICA DR HCL (third table) and GICA DR PAM (fourth table).

	A	B	Sum
A	161	4	165
B	50	35	85
Sum	211	39	250

	A	B	Sum
A	78	87	165
B	75	10	85
Sum	153	97	250

	A	B	Sum
A	163	2	165
B	81	4	85
Sum	244	6	250

	A	B	Sum
A	62	103	165
B	48	37	85
Sum	110	140	250

Note. HCL = Hierarchical clustering using Ward's method, PAM = Partitioning Around Medoids, Two Step = two step method of Durieux and Wilderjans (2019), GICA DR = Group Independent Component Analysis with Dual Regression. The labels A and B refer to the estimated clusters (for C-ICA in the rows, for the other methods in the columns).

5.3.2. *Cluster specific spatial maps* $\mathbf{S}^r$

To interpret and compare the underlying cluster specific RSNs $\mathbf{S}^r$, we matched these RSNs per cluster to a set of template RSNs that are related to known brain regions. In Fig. 6, we plotted each template RSNs (most left column) against the matched RSN for each cluster of the C-ICA solution with $R = 2$ clusters and $Q = 25$ RSNs (middle panel, column 2–3) and the C-ICA solution with $R = 3$ clusters and $Q = 25$ RSNs (right panel, column 4–6). The RSNs are matched to the template RSNs by, first, computing the Tucker congruence between the

templates[10] and the cluster specific RSNs. Next, for each template, the estimated RSNs with the largest (absolute) Tucker congruence value are selected from each cluster, resulting in a one-to-one mapping of the RSNs across all clusters. Note that this matching procedure provides a solution to the RSN/component permutation ambiguity of (C-)ICA (see Appendix B).

As can be seen in the middle panel of Fig. 6, which presents the two-cluster C-ICA solution, subtle differences between (some) RSNs of both clusters exist. For example, a difference is visible between the estimated executive network (template T6 in Fig. 6) of the first cluster (middle panel, column A, predominately EC) and the second cluster (middle panel, column B, mainly AD patients). In particular, the executive network seems to have an intensified functional connectivity for the second cluster (predominantly AD patients) compared to the first cluster (mostly EC subjects). Moreover, differences in functional connectivity are visible in the right frontoparietal network (template T7), for which subjects from the second cluster (cluster B; mainly AD) show more widespread connectivity than subjects belonging to the first cluster (cluster A; mainly EC). The opposite pattern occurs for the lateral visual network (template T2) and the sensorimotor network (template T4). That is, subjects from cluster A show slightly more widespread activity than subjects allocated to cluster B.

The estimated RSNs of the three-cluster solution (see right panel of Fig. 6, columns 4–6) also show subtle differences between clusters in terms of matched RSNs. For example, a subtle difference in activation pattern can be observed between the estimated salience networks of the three clusters (template T3). Further, for the executive network (template network T6), the matched RSN from the first cluster (cluster A) is quite different from the corresponding template, implying that none of the RSNs from that cluster really correspond to this template (for this specific z coordinate). Further, for network T2, the corresponding RSN from cluster B shows connectivity in the visual areas of the human cortex, which is not the same in the matched RSN of cluster A and C. To fully capture the differences between the obtained clusters, we strongly advise to inspect all $Q = 25$ estimated RSNs per cluster. Indeed, the derived clustering is based not only on the eight ad hoc matched RSNs but on all 25 cluster specific RSNs and associated subject specific time courses.

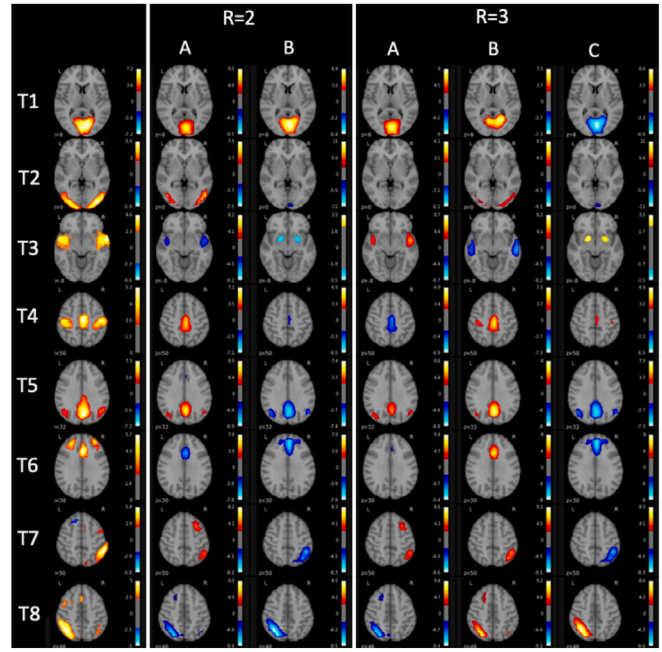

Fig. 6. Axial images of the cluster specific RSNs – matched to template RSNs in terms of Tucker congruence value – obtained by applying C-ICA with $R = 2$ and $R = 3$ clusters and $Q = 25$ RSNs to the empirical data set. In the most left column, eight known RSNs from a template from Beckmann et al. (2005) are displayed (in the rows, indicated by T1–T8). These eight known template RSNs refer to eight important brain networks: (T1) medial visual network, (T2) lateral visual network, (T3) salience network, (T4) sensorimotor network, (T5) default mode network, (T6) executive control network, (T7) right frontoparietal network and (T8) left frontoparietal network. In the middle panel (column 2–3), the corresponding estimated cluster specific RSNs of the two cluster C-ICA solution are presented. Note that the second column belongs to cluster 1 and the third column to cluster 2, with these RSNs being the most similar ones in terms of Tucker congruence to the template patterns. In the right panel (column 4–6), the estimated cluster specific RSNs – matched to the templates – for the three cluster C-ICA solution are displayed, with the fourth, fifth and sixth column belonging to the first, second and third cluster of this C-ICA solution, respectively. Note that the z coordinates of the eight rows are: 8, 0, −8, 50, 32, 30, 50, 46 and are equal across columns. For each estimated RSN, only values above 2.3 or below −2.3 are shown.

6. Discussion

In this paper, the C-ICA model was presented, which is a novel unsupervised clustering model that combines a clustering technique with ICA in order to cluster patients based on differences and similarities in underlying RSNs. In this model, patients are clustered into homogeneous groups (or neurofunctional subtypes) such that patients from the same group have the same RSNs that are representative for that group and patients belonging to different groups can be characterized using RSNs that are qualitatively different. Additionally, to estimate the parameters of the C-ICA model, an Alternating Least Squares (ALS) algorithm was constructed. Further, in order to determine the optimal number of clusters and RSNs underlying a data set at hand, a sequential model selection procedure was proposed. This sequential procedure consists of two steps in which, first, the optimal number of clusters R is determined, and, next, conditional on this optimal number of clusters, the optimal number of RSNs/components Q is selected. Here, an automated scree test like procedure based on computing ratios of the loss function values for several solutions with varying R and Q is used in both steps.

The novel C-ICA model and its associated ALS algorithm were successfully validated in four extensive simulation studies that differed in how realistically the data were generated in terms of underlying RSNs, time courses and noise structures. The results of the first simulation study demonstrated that C-ICA can accurately estimate the underlying cluster structure, cluster specific RSNs and subject specific time courses. These C-ICA parameters were successfully estimated under a wide variety of data characteristics, such as, noisiness of the data and a varying number of clusters and RSNs. Moreover, although the difference being small for some conditions, C-ICA outperformed a clustering procedure based on Group ICA plus dual regression and the two step procedure of Durieux and Wilderjans (2019). In the second simulation study, C-ICA was tested under conditions that are more realistic for rs-fMRI data. In particular, four known RSNs were used as true underlying RSNs and time courses and noise structures were generated in a more realistic way. Similar to the first simulation study, the C-ICA method recovered the C-ICA parameters to a very large extent. In the third simulation study, the uncovering of the clustering was made harder as the RSNs, generated from a sub-Gaussian distribution for this simulation study, for each cluster were allowed to overlap spatially and subject specific noise that obfuscated the true cluster structure – which C-ICA is not designed to recover – was added to the data. The results indicated that also in these more difficult situations C-ICA yields an excellent recovery performance (except in the extremely hard conditions) and outperforms the other two clustering procedures. The latter is also true when larger values for the number of RSNs (Q) than the true number of shared RSNs is used, except for the two step clustering procedure adopting Ward's

[10] If no template(s) is available, a matching of RSNs across clusters can also be obtained by selecting one cluster as the reference cluster and comparing the RSNs from the other cluster(s) to the RSNs from this reference cluster.

method. In the fourth simulation study, in which rs-fMRI data derived from an existing data set with a clear underlying cluster structure was used, it was shown that C-ICA is able to retrieve this clustering almost perfectly. Thus, when a cluster structure with clear differences in RSNs is present in a data set, C-ICA can correctly identify this cluster structure. Also here, C-ICA outperforms the other two clustering procedures, also when varying the number of RSNs except for the two step procedure with Ward's method. In a fifth study, we successfully tested the sequential model selection procedure, implying that this procedure can be employed to aid the user in selecting the optimal number of RSNs and clusters for a data set at hand.

Finally, C-ICA was illustrated on an empirical rs-fMRI data set consisting of Alzheimer's disease patients (AD) and elderly controls (EC). The results showed that C-ICA retrieved a meaningful subject clustering with differences in RSNs between the clusters. Moreover, the obtained clustering partially matched the clinically determined diagnostic labels (i.e., AD-EC), however, without making use of these labels when estimating the clustering. Applying the two other clustering procedure to the empirical data set resulted in a less meaningful clustering that (almost) did not match the diagnostic labels. This was also true when other values for the number of RSNs were tested. One benefit of applying C-ICA is the possibility to discover within group-heterogeneity which may point at the existence of an unknown neurofunctional subtype(s) in the data. Indeed, the three cluster C-ICA solution revealed a small third cluster that mainly contained healthy elderly controls along with a cluster containing 'subjects at risk'. In a follow up study, more information regarding the subjects should be used (e.g., clinical information and behavioural measures) such that the obtained clustering can be better validated and, as such, becomes more insightful. One potential limitation of the current application is that the voxel resolution was down-sampled such that the analysis was more computationally feasible. Future research should incorporate higher voxel resolution data sets, such that C-ICA may disclose more fine-grained differences in RSNs between subject clusters.

From the validation studies and the empirical application, it can be concluded that C-ICA to a very large extent retrieves the subject clusters that are present in multi-subject fMRI data. As such, C-ICA uncovers heterogeneity in RSNs between (a priori unknown groups of) subjects, whereas the Group ICA with dual regression and the two step clustering procedure are substantially worse at detecting subject partitions under the several conditions evaluated here in this study. This implies that these two clustering procedures cannot easily identify the subject clusters underlying the data for the conditions tested in our study. In sum, C-ICA offers a structured way to discover a natural subject partition present in the data, which cannot be obtained by these other two clustering procedures as they are not designed to do this.

6.1. Limitations and directions for future research

Although its good performance on simulated and empirical data, the C-ICA model has some limitations, which may be used as starting points for future research regarding model improvements and extensions. A first limitation of the C-ICA model is that it does not specifically account for longitudinal data. Here, longitudinal refers to the collection of data for the same group of individuals at different moments in time (e.g., follow up scans every year). Collecting and analysing longitudinal fMRI data is particularly interesting for the field of developmental psychology and the study of neurodegenerative diseases. For example, the analysis of individual growth trajectories and individual differences ultimately leads to a better understanding of brain development (Crone and Elzinga, 2015). In a similar way, analysing longitudinal imaging data of neurogenerative diseases, such as AD, may lead to a more profound understanding of cognitive decline and may point to biomarkers that are sensitive for these neurodegenerative diseases (Staffaroni et al., 2018). Note, however, that it is possible to perform C-ICA on each follow up scan separately and to investigate whether the clusters estimated by C-ICA on baseline remain stable or change over time. Perhaps more interestingly to study is whether healthy subjects allocated to a separate cluster would later on move to a cluster with clinically diagnosed patients, allowing to detect healthy subjects at risk for the disease. These subjects and their associated RSNs could provide valuable information about disease progression. A better option, however, would be to construct a clustering method that groups the subjects based on changes in profiles over time in RSNs, which would imply an extension of C-ICA to longitudinal multi-subject fMRI data.

A final limitation of C-ICA pertains to the fact that only data from a single brain modality can be analysed. In this paper, we used rs-fMRI data as previous research clearly demonstrated that disruptions in RSNs are related to (mental) diseases and disease subtypes (Rombouts et al., 2005; Greicius et al., 2007; Greicius, 2008; Pievani et al., 2011; Kaiser et al., 2015). Of course, C-ICA can also be used, albeit maybe with some model adjustments, to analyse data from other brain modalities as long as can be expected that patient groups show – disease related – differences regarding the brain modality under study and ICA is able to uncover meaningful patterns from this brain modality. As patient groups also differ on other modalities, such as structural MRI (Plant et al., 2010; Cuingnet et al., 2011) and diffusion-weighted MRI (Nir et al., 2015), information from these modalities can be considered as complementary to the rs-fMRI information (Calhoun et al., 2006; Sui et al., 2011). As such, combining information from multiple modalities may help in uncovering the cluster structure underlying the data. Indeed, by incorporating several modalities in C-ICA, a more robust cluster structure can be estimated, yielding an improvement over a cluster structure that is based on a single modality only. To incorporate multiple modalities in the C-ICA method, one starting point is the joint ICA method (Calhoun et al., 2007), which enables the estimation of a single mixing matrix that is identical for both modalities and a source matrix that contains multi-modal components (i.e., features from different modalities). To cluster patients based on multi-modal data, a clusterwise version of joint ICA could be developed in which subjects are clustered based on similarities and differences in multi-modal components. As such, each subject cluster can be described by a different set of multi-modal components which enables the discovery of heterogeneity in multi-modal features between subject clusters. Based on these results, hypotheses could be generated regarding multi-modal biomarkers that then can be tested and confirmed in a follow-up study.

CRediT authorship contribution statement

Jeffrey Durieux: Conceptualization, Methodology, Software, validation, Formal analysis, data curation, Writing – original draft, Visualization. **Serge A.R.B. Rombouts:** Conceptualization, Writing – review & editing, Supervision, project administration, Funding acquisition. **Frank de Vos:** Resources, Data curation, Writing – review & editing. **Marisa Koini:** Resources, data curation, Writing – review & editing. **Tom F. Wilderjans:** Conceptualization, Methodology, Software, Validation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The research leading to the results reported in this paper was sponsored by a Talent Grant (Grant Number 406.16.563) from the Netherlands Organization for Scientific Research (NWO).

Appendix A. Independent component analysis for single-subject (rs-)fMRI data

In ICA for a single subject, a multivariate – observed – signal (e.g., the BOLD response for a set of voxels measured over time, see Fig. 1) is decomposed into a set of statistically independent components – also called spatial maps or RSNs – and their associated time courses. In the context of fMRI data, in most cases, independent components are sought for that are spatially independent (i.e., spatial ICA) as brain functioning is conceptualized as activity that is organized in spatially independent brain regions consisting of voxels that covary over time (Calhoun et al., 2009); sometimes, however, researchers prefer to uncover temporally independent components (i.e., temporal ICA). As such, ICA is able to separate systematic signal information from noise and irrelevant sources of variability that usually compromise the BOLD signal. Here, the systematic signal information refers to sets of correlated voxels which represent functionally connected brain regions, whereas irrelevant sources of variability refer to artefacts such as subtle head movements and cardiac pulsations. As a consequence, because of their ability to separate signal from noise, ICA based methods (Griffanti et al., 2014; Pruim et al., 2015) are often used in (rs-)fMRI studies as a pre-processing step in order to improve signal quality (for an example, see Feis et al., 2015). Moreover, ICA is also often employed to disclose the RSNs underlying subjects' fMRI data to study, for example, differences between subjects (groups) in brain functioning (e.g., comparing diseased patients to healthy controls in terms of RSNs).

Technically, ICA is a multivariate analysis technique that aims at finding a linear representation of non-Gaussian data in such a way that the statistical dependency between the underlying non-Gaussian source components is minimized. In the basic ICA model (Jutten and Herault, 1991; Comon, 1994; Hyvärinen et al., 2001), an (underlying) n-dimensional random vector of non-Gaussian independent source signals $\mathbf{s} = (s_1, \ldots, s_n)^T$ is recovered from an n-dimensional random vector of observed signal mixtures $\mathbf{x} = (x_1, \ldots, x_n)^T$. The observed mixed signals in $\mathbf{x}$ are obtained by a linear mixing by means of an $n \times n$ mixing matrix $\mathbf{A}$ (with elements a_{ij}) of the independent source signals in $\mathbf{s}$. Thus, the general ICA decomposition can be written as:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{6}$$

The unknown source signals $\mathbf{s}$ can be computed by multiplying the (pseudo-) inverse of the (unknown) mixing matrix $\mathbf{A}$ – called the unmixing matrix $\mathbf{W}$ with elements w_{ij} – with the observed mixed signals $\mathbf{x}$:

$$\mathbf{s} = \mathbf{A}^{\dagger}\mathbf{x} = \mathbf{W}\mathbf{x} \tag{7}$$

with † denoting the pseudo-inverse of a matrix.

Therefore, in order to find the underlying independent components $\mathbf{s}$, the unmixing matrix $\mathbf{W}$ has to be determined. To this end, one often searches for the matrix $\mathbf{W}$ that makes the components $\mathbf{s}$ as non-Gaussian as possible; due to the Central Limit Theorem (CLT), these maximally non-Gaussian components are also – as much as possible – statistically independent (for more information, see Hyvärinen et al., 2001). As several measures for nongaussianity can be adopted (e.g., measures based on kurtosis, negentropy, mutual information or cumulants), several classes of ICA algorithms to estimate the independent components have been proposed (for more information, see Hyvärinen et al., 2001). A popular implementation of ICA, known as fastICA, determines $\mathbf{W}$ with a fast fixed-point algorithm such that the estimated latent sources $\mathbf{s}$ are maximally non-Gaussian. To this end, fastICA maximizes an information theoretic quantity known as negentropy, which is a normalized version of differential entropy (Hyvärinen, 1999; Hyvärinen et al., 2001) that indicates how distant a distribution is from the Gaussian distribution. In particular, negentropy, which is always non-negative, equals 0 for a Gaussian distribution (Hyvärinen et al., 2001) and gets larger when the distribution becomes less and less Gaussian. As a consequence, maximizing the negentropy of projected data $\mathbf{W}\mathbf{x}$ gives source variables $\mathbf{s}$ that are maximally non-Gaussian (and due to the CLT also maximally statistically independent).

As estimating negentropy is computationally very difficult, often an approximation of negentropy is used and optimized, which is computationally less demanding. A commonly adopted approximation of negentropy (for a univariate random variable) is

$$J(s) \propto [E\{G(s)\} - E\{G(v)\}]^2 \tag{8}$$

here, v is a standardized Gaussian variable, s is the independent component that is sought for, E{} is the expectation operator and G is any nonquadratic contrast function. Useful choices for G are

$$G_1(s) = \frac{1}{a_1} \log \cosh a_1 s (\text{where } 1 \le a_1 \le 2) \tag{9}$$

and

$$G_2(s) = -\exp(-s^2/2) \tag{10}$$

Sometimes, the following more sophisticated approximation of negentropy is warranted:

$$J(s) \approx k_1(E\left\{G_1(s)\right\})^2 + k_2(E\left\{G_2(s)\right\})^2 - E\left\{G_2(v)\right\}^2 \tag{11}$$

with v and s defined as above, k_1 and k_2 being positive integers and G_1 and G_2 being respectively an odd and an even nonquadratic contrast function.

In order to greatly reduce the numerical complexity of ICA, fastICA employs a pre-processing step known as whitening (i.e., data decorrelation). After this whitening step, the independent components necessarily lie in an orthogonal space. This implies that instead of an unrestricted mixing matrix $\mathbf{A}$ with all its parameters (see Eq. (6)), one only has to estimate an orthogonal mixing matrix $\tilde{\mathbf{A}}$ (note: $\tilde{\mathbf{A}}^T\tilde{\mathbf{A}} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix), which has way less parameters to estimate (Hyvärinen et al., 2001). Indeed, after whitening the data, the independent components can easily be estimated by finding the rotation matrix that maximizes the negentropy of $\mathbf{s}$. To uniquely define the independent components, also a unit norm constraint is imposed on each latent source $\mathbf{s}$. Note that, as fastICA with a small(er) Q also involves data reduction, the fastICA estimates of $\mathbf{A}$ and $\mathbf{s}$ yield a solution to the minimization of $\|\mathbf{x} - \mathbf{A}\mathbf{s}\|^2$ (with the rank of $\mathbf{A}\mathbf{s}$ being Q).

Appendix B. Ambiguities for the (C-)ICA model

The C-ICA model suffers from four sources of non-uniqueness/ambiguity (see Hyvärinen et al., 2001), with the first three also holding for the ICA model and the fourth one being specific for C-ICA. First, scaling ambiguity, which implies that a scaling of an independent component in $\mathbf{S}^r$ can be compensated by counter scaling the corresponding time course in $\mathbf{A}_i$ (and vice versa), resulting in the product $\mathbf{A}_i\mathbf{S}^r$ being unchanged. This is because in (C-)ICA both the mixing matrix and the independent components have to be estimated and only their product shows up in the ICA (and C-ICA) loss function. Here, any scalar multiplier applied to one of the sources s_q^r can be cancelled in $\mathbf{A}_i\mathbf{S}^r$ by dividing the row corresponding to that source in all associated $\mathbf{A}_i$'s by that scalar. As noted by Hyvärinen et al. (2001), this non-uniqueness can be accounted for during (C-)ICA estimation by enforcing that the independent components $\mathbf{S}$ all have unit variance (i.e., $E\left\{s_q^2\right\} = 1$). This, however, does not solve the sign ambiguity of the components (see further).

A second ambiguity is reflectional or sign ambiguity, which pertains to the possibility to change the sign of an estimated independent component. Indeed, multiplying one of the estimated components (in $\mathbf{S}^r$) by −1 does not affect the (C-)ICA model (6) as long as this is compensated for in the associated $\mathbf{A}_i$'s (i.e., multiplying the associated time course with −1). Note that reflectional ambiguity is a special case of scaling ambiguity (i.e., scaling with a factor of −1). Constraining the variance of the independent components to one, however, does not solve for reflectional ambiguity. A solution here could be to select

for each independent component that reflection of the component that yields the smallest number of negative elements.

Third, since both the components $\mathbf{S}^r$ and mixing matrix $\mathbf{A}_i$ are unknown, the order of the components in the (C-)ICA model can be freely interchanged. To see why this is possible, consider the basic ICA model from (6). Here it is possible to change the order of the components (and corresponding time courses) by introducing a permutation matrix $\mathbf{P}$ and its inverse $\mathbf{P}^{-1}$ into the basic ICA model:

$$\mathbf{x} = (\mathbf{AP})(\mathbf{P}^{-1}\mathbf{s}) = \mathbf{A}^*\mathbf{s}^* = \mathbf{As} \tag{12}$$

Here, $\mathbf{s}^* = \mathbf{P^{-1}s}$ denotes the latent source signals $\mathbf{s}$ permuted by the binary entries of $\mathbf{P}^{-1}$ and $\mathbf{A}^* = \mathbf{AP}$ contains the similarly permuted time courses. In other words, the order of the latent components in $\mathbf{s}$ (and corresponding vectors of $\mathbf{A}$) can be permuted without affecting the linear sum $\mathbf{x} = \mathbf{As}$. Note that by enforcing unit scaling on each element of $\mathbf{s}$, the variance of each independent component s_q cannot be used anymore to order the s_q's. As a way out, the variance in the associated time courses in $\mathbf{A}$ can still be used to somehow order the components.

A fourth ambiguity, which solely applies to the C-ICA model (and not to ICA), is that the cluster indices of the $\mathbf{S}^r$'s can be permuted freely. Thus, not only the independent components $\mathbf{S}$ can be permuted into $Q!$ ways (see third ambiguity) but also the clustered source signals $\mathbf{S}^r$ can be permuted into $R!$ different ways. Note that this ambiguity, which is often indicated as the labelling problem, pertains to almost all clustering procedures.

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders : DSM-5, fifth ed. American Psychiatric Association, Arlington, VA.

Arabie, P., Hubert, L., 1996. Advances in cluster analysis relevant to marketing research. In: Gaul, W., Pfeifer, D. (Eds.), From Data to Knowledge: Theoretical and Practical Aspects of Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–19.

Ashby, F., 2011. Statistical Analysis of FMRI Data. MIT Press.

Beckmann, C.F., DeLuca, M., Devlin, J.T., Smith, S.M., 2005. Investigations into resting-state connectivity using independent component analysis. Philos. Trans. R. Soc. B 360 (1457), 1001–1013.

Beckmann, C.F., Mackay, C.E., Filippini, N., Smith, S.M., 2009. Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. NeuroImage 47 (Suppl 1), S148.

Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging 23 (2), 137–152. http://dx.doi.org/10.1109/TMI.2003.822821.

Beckmann, C.F., Smith, S.M., 2005. Tensorial extensions of independent component analysis for multisubject {FMRI} analysis. NeuroImage 25 (1), 294–311.

Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind seperation and blind deconvolution. Neural Comput. 7, 1129–1159.

Bougeard, S., Abdi, H., Saporta, G., Niang, N., 2018. Clusterwise analysis for multiblock component methods. Adv. Data Anal. Classif. 12 (2), 285–313.

Brusco, M.J., 2006. A repetitive branch-and-bound algorithm for minimum within-cluster sums of squares partitioning. Psychometrika 71, 347–363.

Brusco, M.J., Cradit, J.D., 2005. ConPar: a method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses. J. Math. Psych. 49 (2), 142–154.

Calhoun, V.D., Adalı, T., 2006. Unmixing fMRI with independent component analysis. IEEE Eng. Med. Biol. Mag. 25 (2), 79–90.

Calhoun, V.D., Adalı, T., Giuliani, N., Pekar, J., Kiehl, K., Pearlson, G., 2006. Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data. Hum. Brain Mapp. 27 (1), 47–62.

Calhoun, V.D., Adali, T., McGinty, V., Pekar, J.J., Watson, T.D., Pearlson, G., 2001a. fMRI activation in a visual-perception task: Network of areas detected using the general linear model and independent components analysis. NeuroImage 14 (5), 1080–1088.

Calhoun, V.D., Adalı, T., Pearlson, G.D., Pekar, J.J., 2001b. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp. 14 (3), 140–151.

Calhoun, V.D., Liu, J., Adalı, T., 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage 45 (1), S163–S172.

Calhoun, V.D., Silva, R., Liu, J., 2007. Identification of multimodal MRI and EEG biomarkers using joint-ICA and divergence criteria. In: 2007 IEEE Workshop on Machine Learning for Signal Processing. IEEE, pp. 151–156.

Cariou, V., Wilderjans, T.F., 2018. Consumer segmentation in multi-attribute product evaluation by means of non-negatively constrained CLV3W. Food Qual. Pref. 67, 18–26.

Ceulemans, E., Kiers, H.A.L., 2006. Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. Br. J. Math. Stat. Psychol. 59, 133–150.

Ceulemans, E., Van Mechelen, I., Leenen, I., 2007. The local minima problem in hierarchical classes analysis: An evaluation of a simulated annealing algorithm and various multistart procedures. Psychometrika 72 (3), 377.

Comon, P., 1994. Independent component analysis, a new concept? Signal Process. 36 (3), 287–314.

Crone, E.A., Elzinga, B.M., 2015. Changing brains: how longitudinal functional magnetic resonance imaging studies can inform us about cognitive and social-affective growth trajectories. WIREs Cogn. Sci. 6 (1), 53–63.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.-O., Chupin, M., Benali, H., Colliot, O., Alzheimer's Disease Neuroimaging Initiative, 2011. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage 56 (2), 766–781.

de Leeuw, J., 1994. Block-relaxation algorithms in statistics. In: Bock, H.-H., Lenski, W., Richter, M.M. (Eds.), Information Systems and Data Analysis. Springer-Verlag, Berlin, pp. 308–325.

De Ridder, D., Kittler, J., Duin, R.P., 2000. Probabilistic PCA and ICA subspace mixture models for image segmentation. In: BMVC. Citeseer, pp. 1–10.

De Roover, K., Ceulemans, E., Timmerman, M.E., 2012a. How to perform multiblock component analysis in practice. Behav. Res. Methods 44, 41–56.

De Roover, K., Ceulemans, E., Timmerman, M.E., Vansteelandt, K., Stouten, J., Onghena, P., 2012b. Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. Psychol. Methods 17, 100–119.

De Roover, K., Vermunt, J.K., Timmerman, M.E., Ceulemans, E., 2017. Mixture simultaneous factor analysis for capturing differences in latent variables between higher level units of multilevel data. Struct. Equ. Model.: Multidiscip. J. 24 (4), 506–523.

De Soete, G., Carroll, J.D., 1994. K-means clustering in a low-dimensional euclidean space. In: Diday, E., Lechevallier, Y., Schader, M., Bertrand, P., Burtschy, B. (Eds.), New Approaches in Classification and Data Analysis. Springer Verlag, Heidelberg, pp. 212–219.

de Vos, F., Koini, M., Schouten, T.M., Seiler, S., van der Grond, J., Lechner, A., Schmidt, R., de Rooij, M., Rombouts, S.A., 2018. A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease. NeuroImage 167, 62–72.

Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R.N., Zebley, B., Oathes, D.J., Etkin, A., et al., 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat. Med. 23 (1), 28–38.

Durieux, J., Wilderjans, T.F., 2019. Partitioning subjects based on high-dimensional fMRI data: comparison of several clustering methods and studying the influence of ICA data reduction in big data. Behaviormetrika 1–41.

Erhardt, E.B., Rachakonda, S., Bedrick, E.J., Allen, E.A., Adalı, T., Calhoun, V.D., 2011. Comparison of multi-subject ICA methods for analysis of fMRI data. Hum. Brain Mapp. 32 (12), 2075–2095.

Esposito, F., Scarabino, T., Hyvärinen, A., Himberg, J., ormisano, E.F., Comani, S., Tedeschi, G., Goebel, R., Seifritz, E., Salle, F.D., 2005. Independent component analysis of fMRI group studies by self-organizing clustering. NeuroImage 25 (1), 193–205.

Feis, R.A., Smith, S.M., Filippini, N., Douaud, G., Dopper, E.G.P., Heise, V., Trachtenberg, A.J., van Swieten, J.C., van Buchem, M.A., Rombouts, S.A.R.B., Mackay, C.E., 2015. ICA-based artifact removal diminishes scan site differences in multi-center resting-state fMRI. Front. Neurosci. 9.

García, V., Mollineda, R.A., Sánchez, J.S., 2009. Index of balanced accuracy: A performance measure for skewed class distributions. In: Iberian Conference on Pattern Recognition and Image Analysis. Springer, pp. 441–448.

Gili, M., Serra, L., Perri, R., Giove, F., Maraviglia, B., Caltagirone, C., Bozzali, M., 2011. Regional brain atrophy and functional disconnection across Alzheimer's disease evolution. J. Neurol. Neurosurg. Psychiatry 82, 58–66.

Gour, N., Felician, O., Didic, M., Koric, L., Gueriot, C., Chanoine, V., Confort-Gouny, S., Guye, M., Ceccaldi, M., Ranjeva, J.P., 2014. Functional connectivity changes differ in early and late-onset alzheimer's disease. Hum. Brain Mapp. 35 (7), 2978–2994. http://dx.doi.org/10.1002/hbm.22379.

Greicius, M.D., 2008. Resting-state functional connectivity in neuropsychiatric disorders. Curr. Opin. Neurol. 21 (4), 424–430.

Greicius, M.D., Flores, B.H., Menon, V., Glover, G.H., Solvason, H.B., Kenna, H., Reiss, A.L., Schatzberg, A.F., 2007. Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. Biol. Psychiat. 62 (5), 429–437.

Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default-mode network activity distinguishes alzheimer's disease from healthy aging: Evidence from functional MRI.. Proc. Natl. Acad. Sci. USA 101 (13), 4637–4642.

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C.F., Auerbach, E.J., Douaud, G., Sexton, C.E., Zsoldos, E., Ebmeier, K.P., Filippini, N., Mackay, C.E., Moeller, S., Xu, J., Yacoub, E., Baselli, G., Ugurbil, K., Miller, K.L., Smith, S.M., 2014. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. NeuroImage 95 (Supplement C), 232–247.

Grigg, O., Grady, C.L., 2010. Task-related effects on the temporal and spatial dynamics of resting-state functional connectivity in the default network. PLoS One 5 (10), e13311.

Guo, Y., Pagnoni, G., 2008. A unified framework for group independent component analysis for multi-subject fMRI data.. NeuroImage 42, 1078–1093.

Harshman, R.L., 1972. PARAFAC2. Mathematical and technical notes. In: UCLA Working Papers in Phonetics, Vol. 22. pp. 30–44.

Harshman, R.A., Lundy, M.E., 1994. PARAFAC: Parallel factor analysis. Comput. Statist. Data Anal. 18, 39–72.

Helwig, N.E., 2018. Ica: Independent component analysis. [Computer software manual]. (R package version 1.0-2).

Himberg, J., Hyvärinen, A., Esposito, F., 2004. Validating the independent components of neuroimaging time series via clustering and visualization. NeuroImage 22 (3), 1214–1222.

Hubert, L., Arabie, P., 1985. Comparing partitions. J. Classification 2 (1), 193–218.

Hyvärinen, A., 1997. One-unit contrast functions for independent component analysis: A statistical analysis. In: Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop. IEEE, pp. 388–397.

Hyvärinen, A., 1999. Fast and robust fixed-point algorithm for independent component analysis. IEEE Trans. Neur. Net. 10 (3), 626–634. http://dx.doi.org/10.1109/72.761722.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley and Sons, New York.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841.

Jenkinson, M., Beckmann, C., Behrens, T., Woolrich, M., Smith, S., 2012. FSL. NeuroImage 62, 782–790.

Jutten, C., Herault, J., 1991. Blind separation of sources, Part 1: An adaptive algorithm based on neuromimetic architecture. Signal Process. 24, 1–10.

Kaiser, R.H., Andrews-Hanna, J.R., Wager, T.D., Pizzagalli, D.A., 2015. Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. JAMA Psychiatry 72 (6), 603–611.

Kuhn, H.W., 1955. The Hungarian method for the assignment problem. Nav. Res. Logist. Q. 2 (1–2), 83–97.

Lee, T.-W., Lewicki, M.S., Sejnowski, T.J., 2000. ICA mixture models for unsupervised classification of non-Gaussian classes and automatic context switching in blind signal separation. IEEE Trans. Pattern Anal. Mach. Intell. 22 (10), 1078–1089.

Lorenzo-Seva, U., Ten Berge, J.M., 2006. Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology 2 (2), 57–64.

Lynall, M.-E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., Bullmore, E., 2010. Functional connectivity and brain networks in schizophrenia. J. Neurosci. 30 (28), 9477–9487.

Milligan, G.W., Soon, S.C., Sokol, L.M., 1983. The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure. IEEE Trans. Pattern Anal. Mach. Intell. 5, 40–47.

Millsap, R.E., Meredith, W., 1988. Component analysis in cross-sectional and longitudinal data. Psychometrika 53, 123–134.

Nickerson, L.D., Smith, S.M., Öngür, D., Beckmann, C.F., 2017. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. Front. Neurosci. 11, 115.

Nir, T.M., Villalon-Reina, J.E., Prasad, G., Jahanshad, N., Joshi, S.H., Toga, A.W., Bernstein, M.A., Jack, Jr., C.R., Weiner, M.W., Thompson, P.M., Alzheimer's Disease Neuroimaging Initiative, 2015. Diffusion weighted imaging-based maximum density path analysis and classification of Alzheimer's disease. Neurobiol. Aging 36, S132–S140.

Olde Dubbelink, K.T., Stoffers, D., Deijen, J.B., Twisk, J.W., Stam, C.J., Hillebrand, A., Berendse, H.W., 2013. Resting-state functional connectivity as a marker of disease progression in Parkinson's disease: A longitudinal MEG study. NeuroImage: Clin. 2 (Supplement C), 612–619.

Olejnik, S., Algina, J., 2003. Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychol. Methods 8 (4), 434.

Pievani, M., de Haan, W., Wu, T., Seeley, W.W., Frisoni, G.B., 2011. Functional network disruption in the degenerative dementias. Lancet Neurol. 10 (9), 829–843.

Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage 50 (1), 162–174.

Pruim, R.H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. CA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. NeuroImage 112 (Supplement C), 267–277.

Rombouts, S.A.R.B., Barkhof, F., Goekoop, R., Stam, C.J., Scheltens, P., 2005. Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study. Hum. Brain Mapp. 26 (4), 231–239.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. NeuroImage 90 (Supplement C), 449–468.

Seiler, S., Schmidt, H., Lechner, A., Benke, T., Sanin, G., Ransmayr, G., Lehner, R., Dal-Bianco, P., Santer, P., Linortner, P., Schmidt, R., 2012. Driving cessation and dementia: results of the prospective registry on dementia in Austria (PRODEM). PLoS One 7 (12), e52710.

Silva, R.F., Plis, S.M., Adalı, T., Calhoun, V.D., 2014. A statistically motivated framework for simulation of stochastic data fusion models applied to multimodal neuroimaging. NeuroImage 102, 92–117. http://dx.doi.org/10.1016/j.neuroimage.2014.04.035, Retrieved from https://www.sciencedirect.com/science/article/pii/S1053811914003048. Multimodal Data Fusion.

Smilde, A.K., Kiers, H.A.L., Bijlsma, S., Rubingh, C.M., van Erk, M.J., 2009. Matrix correlations for high-dimensional data: the modified RV-coefficient. Bioinformatics 25 (3), 401–405.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, S208–S219.

Staffaroni, A.M., Brown, J.A., Casaletto, K.B., Elahi, F.M., Deng, J., Neuhaus, J., Cobigo, Y., Mumford, P.S., Walters, S., Saloner, R., Karydas, A., Coppola, G., Rosen, H.J., Miller, B.L., Seeley, W.W., Kramer, J.H., 2018. The longitudinal trajectory of default mode network connectivity in healthy older adults varies as a function of age and is associated with changes in episodic memory and processing speed. J. Neurosci. 38 (11), 2809–2817.

Stegeman, A., 2007. Comparing Independent Component Analysis and the PARAFAC Model for Artificial Multi-Subject fMRI Data. Unpublished Technical Report, Univ. of Groningen.

Stegeman, A., Mooijaart, A., 2008. Independent Component Analysis with Errors by Least Squares Covariance Fitting. Unpublished Report, Retrieved from http://www.alwinstegeman.nl/docs/Stegeman_Mooijaart_ICA.pdf.

Steinley, D., Brusco, M.J., Henson, R., 2012. Principal cluster axes: A projection pursuit index for the preservation of cluster structures in the presence of data reduction. Multivar. Behav. Res. 47 (3), 463–492.

Sui, J., Pearlson, G., Caprihan, A., Adalı, T., Kiehl, K.A., Liu, J., Yamamoto, J., Calhoun, V.D., 2011. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. NeuroImage 57 (3), 839–855.

ten Berge, J.M.F., Kiers, H.A.L., Van der Stel, V., 1992. Simultaneous components analysis. Stat. Appl. 4, 377–392.

Timmerman, M.E., Ceulemans, E., Kiers, H.A.L., Vichi, M., 2010. Factorial and reduced k-means reconsidered. Comput. Statist. Data Anal. 54, 1858–1871.

Timmerman, M.E., Kiers, H.A.L., 2003. Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. Psychometrika 68, 105–121.

Tucker, L., 1951. A Method for Synthesis of Factor Analysis Studies. Report No. 984, Department of the Army, Washington, DC.

Van de Ven, V.G., Formisano, E., Prvulovic, D., Roeder, C.H., Linden, D.E.J., 2004. Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. Hum. Brain Mapp. 22, 165–178.

Vichi, M., Kiers, H.A.L., 2001. Factorial k-means analysis for two-way data. Comput. Statist. Data Anal. 37, 49–64.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., Rosseel, Y., 2011. neurosim: An R package for generating fMRI data. J. Stat. Softw. 44 (10), 1–18.

Wilderjans, T.F., Cariou, V., 2016. CLV3w: A clustering around latent variables approach to detect panel disagreement in three-way conventional sensory profiling data. Food Qual. Pref. 47, 45–53.

Wilderjans, T.F., Ceulemans, E., 2013. Clusterwise PARAFAC to identify heterogeneity in three-way data. Chemometr. Intell. Lab. Syst. 129, 87–97.

Wilderjans, T.F., Ceulemans, E., Kuppens, P., 2012a. Clusterwise HICLAS: A generic modeling strategy to trace similarities and differences in multiblock binary data. Behav. Res. Methods 44 (2), 532–545.

Wilderjans, T.F., Ceulemans, E., Meers, K., 2013. CHull: A generic convex-hull-based model selection method. Behav. Res. Methods 45 (1), 1–15.

Wilderjans, T.F., Ceulemans, E., Van Mechelen, I., 2008. The CHIC model: A global model for coupled binary data. Psychometrika 73 (4), 729–751.

Wilderjans, T.F., Ceulemans, E., Van Mechelen, I., 2009. Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes. Comput. Statist. Data Anal. 53, 1086–1098.

Wilderjans, T.F., Ceulemans, E., Van Mechelen, I., 2012b. The SIMCLAS model: Simultaneous analysis of coupled binary data matrices with noise heterogeneity between and within data blocks. Psychometrika 77 (4), 724–740.

Wilderjans, T.F., Ceulemans, E., Van Mechelen, I., van den Berg, R.A., 2011. Simultaneous analysis of coupled data matrices subject to different amounts of noise. Br. J. Math. Stat. Psychol. 64, 277–290.

Wilderjans, T.F., Depril, D., Van Mechelen, I., 2012c. Block-relaxation approaches for fitting the INDCLUS model. J. Classification 29, 277–296.

Wilderjans, T.F., Vande Gaer, E., Kiers, H.A., Van Mechelen, I., Ceulemans, E., 2017. Principal covariates clusterwise regression (PCCR): Accounting for multicollinearity and population heterogeneity in hierarchically organized data. Psychometrika 82 (1), 86–111.

Yang, Z., Zuo, X.-N., Wang, P., Li, Z., LaConte, S.M., Bandettini, P.A., Hu, X.P., 2012. Generalized RAICAR: discover homogeneous subject (sub) groups by reproducibility of their intrinsic connectivity networks. Neuroimage 63 (1), 403–414.

Zhang, H.-Y., Wang, S.-J., Liu, B., Ma, Z.-L., Yang, M., Zhang, Z.-J., Teng, G.-J., 2010. Resting brain connectivity: changes during the progress of Alzheimer disease. Radiology 256 (2), 598–606.

Zhu, X., Hunter, D.R., 2019. Clustering via finite nonparametric ICA mixture models. Adv. Data Anal. Classif. 13 (1), 65–87.