



Exploring individual and group differences in latent brain networks using cross-validated simultaneous component analysis



Nathaniel E. Helwig^{a,b,*}, Matthew A. Snodgrass^a

^a Department of Psychology, University of Minnesota, Minneapolis, MN, 55455, USA

^b School of Statistics, University of Minnesota, Minneapolis, MN, 55455, USA

ARTICLE INFO

Keywords:

Group component analysis
Multi-subject analysis
Multiway analysis
Parallel factor analysis
Parafac2
Tensor decomposition

ABSTRACT

Component models such as PCA and ICA are often used to reduce neuroimaging data into a smaller number of components, which are thought to reflect latent brain networks. When data from multiple subjects are available, the components are typically estimated simultaneously (i.e., for all subjects combined) using either tensor ICA or group ICA. As we demonstrate in this paper, neither of these approaches is ideal if one hopes to find latent brain networks that cross-validate to new samples of data. Specifically, we note that the tensor ICA model is too rigid to capture real-world heterogeneity in the component time courses, whereas the group ICA approach is too flexible to uniquely identify latent brain networks. For multi-subject component analysis, we recommend comparing a hierarchy of simultaneous component analysis (SCA) models. Our proposed model hierarchy includes a flexible variant of the SCA framework (the Parafac2 model), which is able to both (i) model heterogeneity in the component time courses, and (ii) uniquely identify latent brain networks. Furthermore, we propose cross-validation methods to tune the relevant model parameters, which reduces the potential of over-fitting the observed data. Using simulated and real data examples, we demonstrate the benefits of the proposed approach for finding credible components that reveal interpretable individual and group differences in latent brain networks.

1. Introduction

1.1. Background

Event-related (or task-based) neuroimaging studies record some brain signal in response to an event (or task) that is designed to elicit a particular neural response. For example, an event-related potential study records electroencephalography (EEG) data in response to some stimulus to understand the brain's processing (auditory and/or visual) of the stimulus. Unlike resting-state neuroimaging studies, event-related (or task-based) studies reveal brain networks that relate to the cognitive processes underlying the event/task of interest. This implies that such studies can be useful for establishing norms—and identifying abnormalities—in brain responses to different types of stimuli, which can improve our understanding of various psychological disorders, e.g., addiction, depression, bipolar disorder, schizophrenia, etc.

Typical neuroimaging studies collect data across time from multiple spatial locations (e.g., electrodes or voxels), and most studies collect data from several subjects who participate in multiple conditions of some experimental task. Furthermore, the subjects often participate in multiple

trials of each condition. In many cases, the multiple trials are treated as independent replications of the task, and the data are averaged across the repeated trials within each subject. However, in some cases there is interest in modeling the evolution of the brain response across the trials (e.g., see Williams et al., 2018). Regardless of whether the data are aggregated across trials, a typical neuroimaging study produces vast amounts of data—so some dimension reduction approach is typically needed to extract meaning from the signal.

1.2. Dimension reduction

Component analysis models such as Principal Components Analysis (PCA) and Independent Components Analysis (ICA) are frequently used to obtain lower-dimensional representations of neuroimaging data, which can be useful for identifying brain networks (McKeown et al., 1998; Calhoun et al., 2001). However, the PCA and ICA models are forms of bilinear models, which are designed to model data with two modes of variation, e.g., a time \times space data matrix from a single subject. To apply a bilinear model such as PCA or ICA to multi-subject neuroimaging data, it is necessary to either (i) fit a separate model to each subject's data, or

* Corresponding author. Department of Psychology and School of Statistics, University of Minnesota, Minneapolis, MN, 55455, USA.

E-mail addresses: helwig@umn.edu (N.E. Helwig), snodg031@umn.edu (M.A. Snodgrass).

(ii) fit a single model to the concatenated data (see Beckmann et al., 2005; Calhoun et al., 2009). The first approach is undesirable because there is no clear way to connect the solutions across the subjects, whereas the second approach is undesirable because the multiway structure of the data is not fully leveraged.

Tensor analysis models (see Kolda and Bader, 2009, for an overview), such as the Tucker Factor Analysis model (Tucker, 1966) and the Parallel Factors (or Canonical Polyadic Decomposition) model (Hitchcock, 1927; Harshman, 1970; Carroll and Chang, 1970), provide powerful extensions of bilinear models such as PCA and ICA. As a result, tensor models have the potential to leverage the multiway structure of neuroimaging data to extract brain networks that display systematic variation across time points, spatial locations (electrodes or voxels), and subjects. Such models could also be extended to higher-dimensional data (e.g., 4-mode or 5-mode data) if the subjects participate in multiple experimental conditions and/or trials. Note that tensor ICA (TICA) models have also been proposed (see Beckmann and Smith, 2005; De Vos et al., 2012). However, such extensions are only appropriate if the data have trilinear structure and the underlying components are statistically independent—which is a rather strict assumption (see Helwig and Hong, 2013; Stegeman, 2007).

1.3. Individual differences

Applications of tensor analysis models are on the rise in neuroimaging research, but the focus is typically on the Parafac/CPD model or the TICA model (e.g., see Andersen and Rayens, 2004; Beckmann and Smith, 2005; Cichocki, 2013; Cong et al., 2015; Mørup et al., 2006; Miwakeichi et al., 2004; Martínez-Montes et al., 2008). Note that these models implicitly leverage Cattell's (1944) *principle of parallel proportional profiles*, such that the neuroimaging data are assumed to display proportional variation across all three data modes (i.e., time, locations, and subjects). This implies that the underlying brain networks have the exact same time courses and spatial maps across subjects, which is a rather strict assumption. Note that assuming common spatial maps across subjects may be reasonable, given the spatial organization of the brain. However, assuming the same time courses for all subjects is only appropriate if the event/task elicits the exact same response—at the exact same time points—in all of the subjects.

With real data, there may exist differences in the timing and nature of different subject's neural responses to the event/task. For example, individual differences in the cognitive processing of the stimuli would result in differences in the timing and shape of features in EEG waveforms. As another example, individual differences in haemodynamic response functions would result in differences in the timing and shape of recorded functional magnetic imaging (fMRI) time courses. Typical applications of the Parafac/CPD model, as well as the TICA model, are too rigid to capture these sorts of individual differences—which are expected to exist in most real neuroimaging applications. Consequently, a more flexible framework is needed to capture the heterogeneity in the timing and nature of recorded brain signals from different subjects.

1.4. Proposed approach

In this paper, we discuss how the Parafac2 model (Harshman, 1972) and the Simultaneous Component Analysis (SCA) framework (Timmerman and Kiers, 2003) can be used to solve such problems. Specifically, we leverage recent advances in Parafac2 modeling with smoothness constraints (Helwig, 2017) to flexibly model individual differences in event-related neuroimaging data. We also extend these recent developments to create novel (smoothness constrained) extensions of the SCA framework for analyzing event-related neuroimaging data. The benefits of the Parafac2 model for neuroimaging research have been recently demonstrated (see Ferdowsi et al., 2015; Madsen et al., 2017; Makkiabadi et al., 2011). However, unlike these recent papers, our approach combines compression (through smoothness constraints) and cross-validation methods to automatically find the appropriate model

dimensionality and degree of smoothness of the data.

Using both simulated and real data, we illustrate the benefits of using cross-validated tensor models, particularly the Parafac2 model, for uniquely identifying brain networks in multi-subject neuroimaging data. Our results clearly demonstrate the dangers of ignoring individual differences in brain imaging data, as well as the interpretational issues inherent to decompositions with a rotational indeterminacy. Specifically, we show that (i) the Parafac/CPD model produces biased estimates of the spatial maps when individual differences exist in the underlying time courses, and (ii) tensor models with a rotational freedom can recover the data-generating tensor but cannot necessarily recover the true data-generating spatial maps (due to the rotational indeterminacy). We also show the robustness of the smoothness-constrained Parafac2 model (with CV tuning), which performed well across a wide variety of data-generating conditions.

2. Theory

2.1. Overview of component models

2.1.1. Bilinear models

Let $\mathbf{X} = \{x_{ij}\}_{I \times J}$ denote a time-by-space data matrix recorded from a single subject, where x_{ij} denotes the data recorded at the i -th time point ($i = 1, \dots, I$) from the j -th spatial location ($j = 1, \dots, J$). A bilinear model such as PCA (Pearson, 1901; Hotelling, 1933; Tipping and Bishop, 1999) or ICA (Comon, 1994; Bell and Sejnowski, 1995; McKeown et al., 1998; Hyvärinen et al., 2001) assumes that the data matrix can be decomposed into $R \leq \min(I, J)$ underlying factors (or components), which consist of an outer product of time courses and spatial maps (see Fig. 1a). The bilinear model depicted in Fig. 1a has a rotational indeterminacy, such that for any nonsingular $R \times R$ transformation matrix \mathbf{T} , we have that $\mathbf{A}\mathbf{B}' = \tilde{\mathbf{A}}\tilde{\mathbf{B}}'$, where $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}$ are the rotated time courses, $\tilde{\mathbf{B}} = \mathbf{B}(\mathbf{T}^{-1})'$ are the rotated spatial maps. This implies that, given any bilinear model solution, there are an infinite number of other possible solutions that fit the observed data equally well—but may produce different interpretations of the underlying factors. To obtain unique solutions, bilinear models must impose additional assumptions on the solution, such as orthogonality (PCA) or independence (ICA) of the factors/components. However, in practice, choosing the appropriate rotation can be difficult (see Browne, 2001, for an overview), and there is no guarantee that any standard rotation will find the true components.

2.1.2. Simultaneous component analysis

Let $\mathbf{X}_k = \{x_{ij(k)}\}_{I \times J}$ denote a time-by-space data matrix recorded from the k -th subject for $k = 1, \dots, K$. The Simultaneous Component Analysis (SCA) framework consists of a hierarchy of four models that can be used to factor analyze multivariate time series data collected from $K > 1$ subjects (Timmerman and Kiers, 2003). For multi-subject neuroimaging data, the SCA models assume that each subject has a unique time course, but the subjects share common spatial maps (see Fig. 1b). The four SCA models differ in terms of the assumed crossproduct structure of the subject specific time courses \mathbf{D}_k , see the Supplementary Online Materials (SOM). In all four cases, we can write $\mathbf{D}_k = \mathbf{A}_k\mathbf{C}_k$ where the \mathbf{C}_k are diagonal matrices (similar to Fig. 1d), and the constraints on the \mathbf{A}_k matrices will depend on the assumed version of the SCA model. In this paper, we explore the most flexible version of the SCA model (i.e., SCA-P), which makes no assumptions about the crossproduct structure of the \mathbf{D}_k weights. Note that in this case the SCA model is equivalent to the bilinear model in Fig. 1a applied to the stacked (i.e., row-wise concatenated) data matrices.

2.1.3. Parallel Factors-1 model

The Parallel Factors (Parafac) model was proposed by Harshman (1970) for factor analysis of multiway data, i.e., data collected across more than two modes of variation. Note that neuroimaging data from a

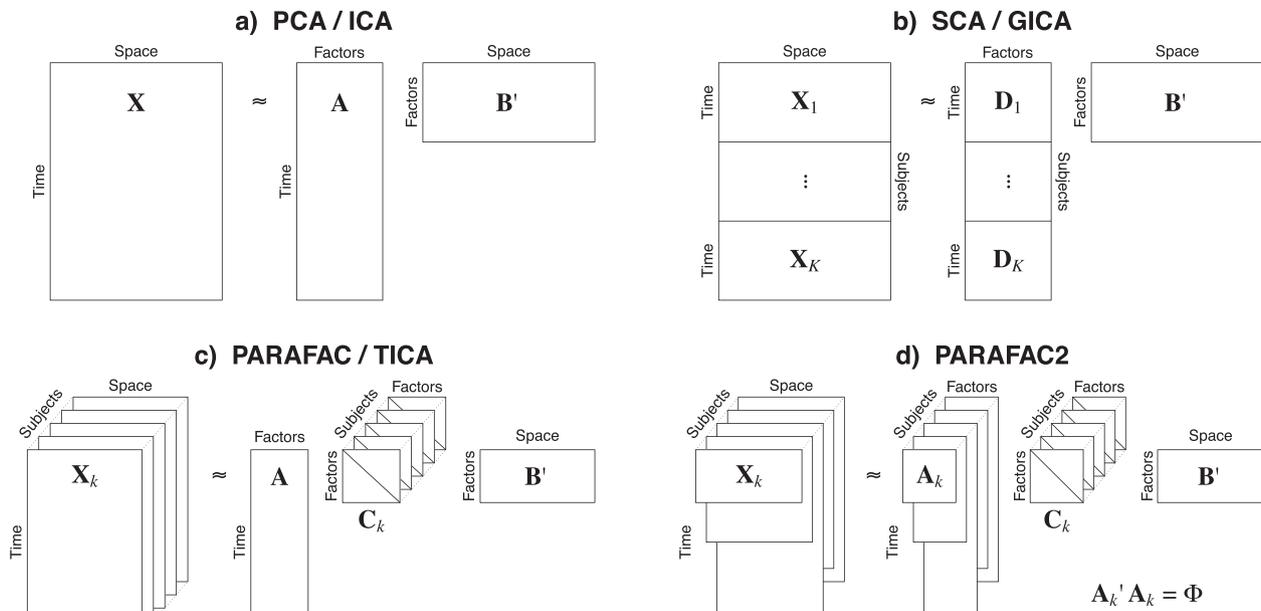


Fig. 1. Component Models. Visualizations of various component models for neuroimaging data. PCA = Principal Component Analysis, ICA = Independent Component Analysis, SCA = Simultaneous Component Analysis, GICA = Group ICA, TICA = Tensor ICA, PARAFAC = Parallel Factor Analysis, and PARAFAC2 = Parallel Factor Analysis-2.

single subject has two modes of variation (time by space), and the subjects add a third mode of variation to the data. For multi-subject neuroimaging data, the Parafac model assumes that each subject has the same time courses and spatial maps, which are weighted by subject specific scores (see Fig. 1c). Unlike the bilinear model in Fig. 1a, the Parafac model in Fig. 1c can produce solutions that are essentially unique, i.e., do not have a rotational indeterminacy (see Harshman, 1970; Harshman and Lundy, 1994; Kruskal, 1977; Sidiropoulos and Bro, 2000). So, if the assumed form of the Parafac model is correct, it is possible to uniquely find latent brain components using this approach (e.g., see Helwig and Hong, 2013; Mørup et al., 2006; Miwakeichi et al., 2004; Williams et al., 2018). However, the assumption that each subject has the same time courses will not be reasonable if between-subjects temporal heterogeneity exists in the underlying components—which may be the case in many real data applications.

2.1.4. Parallel Factors-2 model

The Parafac2 model (Harshman, 1972) is an extension of the Parafac model that has the flexibility to model data with heterogeneity in the factor weights. For multi-subject neuroimaging data, the Parafac2 model allows each subject to have unique time courses A_k , with the constraint that the time courses share a common crossproduct matrix $A_k' A_k = \Phi$ (see Fig. 1d). Similar to the Parafac model, the Parafac2 model can produce essentially unique solutions, i.e., solutions without a rotational indeterminacy (see Harshman, 1972; Harshman and Lundy, 1996; Ten Berge and Kiers, 1996; Kiers et al., 1999; Timmerman and Kiers, 2003; Helwig, 2013). Note that the Parafac2 model has the same form as the SCA model in Fig. 1b, with the additional assumption that $D_k = A_k C_k$ where the A_k matrices satisfy $A_k' A_k = \Phi$. Thus, by adding the “common crossproducts” constraint to the SCA model, it is possible to solve the rotational indeterminacy problem. Of course, in practice, the appropriateness of the common crossproduct constraint needs to be considered, e.g., by comparing the fit of the SCA-P and Parafac2 models.

2.1.5. Tensor and group ICA models

The ICA model can be extended to the multi-subject situation in (at least) two different ways: using tensor ICA (Beckmann and Smith, 2005) or group ICA (Beckmann et al., 2005). The TICA approach assumes the same model form as the Parafac model in Fig. 1c, but adds the assumption

that the spatial maps are statistically independent of one another. Note that this is a strict assumption that is not actually needed, given that the Parafac model already provides a unique decomposition (see Helwig and Hong, 2013). The GICA approach assumes the same model form as the SCA-P model in Fig. 1b, and solves the rotational indeterminacy using an ICA rotation of the spatial maps.

Note that both the TICA and GICA approaches assume the subjects share common spatial maps, which are assumed to be statistically independent of one another. The two approaches differ in their assumptions about the time courses, such that the GICA approach is more flexible. For even further flexibility, the “dual regression” (DR) approach can be applied to estimate subject specific spatial maps, which are based on the original group maps (see Beckmann et al., 2009). Note that because GICA is a form of SCA-P, the approach does not provide a unique solution, i.e., there is a rotational indeterminacy to the solution. As a result, the quality of the GICA solution (with or without DR) will depend on the appropriateness of the ICA rotation used to solve the indeterminacy.

2.2. Component models in practice

2.2.1. Overview

Given a chosen component model and number of factors R , the parameters are typically estimated using a least squares approach (Krijnen, 2006), see the SOM for details. For bilinear models (including SCA-P), a closed form solution exists in terms of the singular value decomposition (Eckart and Young, 1936) of the stacked matrix (Timmerman and Kiers, 2003), and then a post-hoc rotation can be applied to the spatial maps. For the Parafac and Parafac2 models, the parameters are typically estimated using an alternating least squares (ALS) algorithm with multiple random starts (Kiers et al., 1999; Faber et al., 2003; Tomasi and Bro, 2006). In default cases, the parameters are estimated using unconstrained least squares. However, imposing constraints (e.g., non-negativity or smoothness) on a mode’s weights can be useful, particularly when the data are noisy (Harshman and Lundy, 1994; Bro and De Jong, 1997; Timmerman and Kiers, 2002). Later in this subsection, we discuss how to incorporate smoothness constraints on the time courses, which can greatly improve the recovery of the underlying factors. But we first address the issue of choosing the appropriate dimensionality R for the given model.

2.2.2. Choosing the number of components

Selecting the appropriate dimensionality R in bilinear models has been a topic of research for decades (e.g., see Kaiser, 1960; Cattell, 1966; Horn, 1965; Revelle and Rocklin, 1979; Akaike, 1987). Although many different approaches have been proposed over the years, there does not exist a “perfect solution” that is ideal for all circumstances. The parametric approaches (e.g., likelihood ratio tests and information criteria) can perform well for data that meet their particular parametric assumptions, but will fail to provide correct results in general cases. In contrast, the scree line tests can provide useful heuristics, but fail to provide a theoretically rigorous rule for determining the number of components. As a result, selecting the appropriate number of components in a bilinear model is still an open topic of research (e.g., see Hong et al., 2006; Larsen and Warne, 2010; Warne and Larsen, 2014).

To choose the number of components for Parafac/Parafac2 models, it is typical to use the CORE CONSistency DIAGnostic (CORCONDIA) proposed by Bro and Kiers (2003). This criterion leverages the fact that the Parafac model (or Parafac2 model) is a special case of the 3-way Tucker factor analysis model (Tucker, 1966) applied to the data (or projected data for Parafac2, see Kamstrup-Nielsen et al., 2013). The CORCONDIA compares the difference between the least squares estimate of the Tucker “core” array to the Parafac implied “super-diagonal” core array. If the two are similar, the CORCONDIA value will be close to 100, which indicates that the assumed model is well supported by the observed sample of data. In contrast, if the CORCONDIA value is below 100, this indicates deviations from the assumed model structure. In practice, different CORCONDIA thresholds have been proposed for choosing the dimensionality, e.g., CORCONDIA ≥ 80 or ≥ 90 .

2.2.3. Incorporating smoothness constraints

Typical applications of component analysis models to spatiotemporal data do *not* take into consideration the spatial or temporal properties of the observed data. More specifically, the models do not take into account any information about the interrelationships between the levels of a given mode, e.g., that rows i and $i + 1$ of the A matrix (in Fig. 1a) correspond to sequential time points. If the true time courses change as a smooth function across time¹, then including this information in the model can greatly improve the parameter estimates—particularly when the signal-to-noise ratio is small (see Alsberg and Kvalheim, 1993; Reis and Ferreira, 2002; Timmerman and Kiers, 2002; Helwig, 2017). To incorporate smoothness information into the model, we can assume that the time course weights can be written as a linear combination of known basis functions with unknown coefficients. For example, a Parafac model with smoothness constraints on the time course weights assumes that $A = F\alpha$ where $F = \{f_\ell(i)\}_{I \times \nu}$ is a known basis function matrix² with ν denoting the degrees of freedom, and $\alpha = \{\alpha_{\ell r}\}_{\nu \times R}$ is an unknown matrix of coefficients.

The degrees of freedom parameter ν controls the smoothness of the A estimates, such that smaller values of ν relate to more smooth estimates. Note that setting $\nu = I$ (the number of time points) produces the same solution as would be obtained without the smoothness constraints imposed. As a result, the classic (unconstrained) solution can be considered a special case of this “smoothness constrained” approach (see the SOM for details). This implies that, in practice, one must determine a reasonable value of ν for the smoothing. Note that setting ν too small will introduce too much bias to the solution, whereas setting ν too large will

¹ The assumption that the true time courses are “smooth functions” is reasonable for applications of component models to spatiotemporal neuroimaging data. This is because, as will be described in the following paragraph, the unconstrained solution can be considered a special (“rough”) case of this assumption.

² The notation $f_\ell(i)$ denotes the ℓ -th basis function being evaluated at the i -th time point. In this paper, we use a cubic B-spline basis (de Boor, 2001), but other splines are possible (see Gu, 2013; Wahba, 1990).

introduce too much noise to the estimates (assuming noisy data). To find a balance between fitting and smoothing the data, it is typical to use ordinary cross-validation (OCV) applied to the tensor (see Timmerman and Kiers, 2002; Helwig, 2017). However, the OCV criterion is designed for situations where the error terms are “iid” (independent and identically distributed), which is not likely to be true with real neuroimaging data.

2.2.4. Cross-validation for SCA

In typical applications, both the number of components R and the degree of smoothness ν are determined from a single sample of K subjects. This has the potential to lead to *over-fitting* the data, i.e., choosing a value of R or ν that fails to cross-validate for new samples of data. To avoid over-fitting, we propose a novel k-fold cross-validation (CV) procedure to select both the model dimensionality R and the degree of smoothness ν . Specifically, we propose using k-fold CV of the CORCONDIA to determine the appropriate number of component R , and k-fold CV of the tensor mean-squared error (MSE) to select the degrees of smoothness ν . We note that k-fold CV has been used to tune R in multiway models (Louwerse et al., 1999); however, previous discussions have neglected the Parafac2 model, and have not used the cross-validated CORCONDIA for choosing R . The details of our proposed CV approach are discussed in the SOM, and the utility of the method is explored in the Simulation Study.

3. Methods

3.1. Simulation Study

3.1.1. Design

We designed a Simulation Study to compare the various multi-subject component analysis models discussed in Section 2.1, as well as the k-fold CV tuning methods proposed in Section 2.2.4. Our simulation data-generating components are motivated by our real data application to EEG data. Specifically, throughout the simulation, we assumed that $K = 30$ subjects participated in an event-related potential (ERP) experiment where the EEG data were sampled at $I = 256$ Hz from $J = 61$ electrodes. To ensure that the data-generating conditions have practical relevance, the group time courses and spatial maps for the $R = 2$ data-generating components (displayed in Fig. 2a) are derived from the real data results. As a part of the Simulation Study, we compared the methods under 24 different combinations of conditions, which represent all combinations of two factors: (i) the heterogeneity level, i.e., the amount of individual differences in the components (four levels, $\sigma \in \{0, 0.5, 1, 1.5\}$), and (ii) the signal to noise ratio (SNR; six levels: 0.1, 0.15, 0.25, 0.4, 0.6, 1).

Following Stegeman (2007) and Helwig and Hong (2013), we define the SNR as the root mean square of the signal divided by the root mean square of the noise (see the SOM). To simulate realistic ERP components with systematically different levels of heterogeneity, we used the following approach. Letting $F = \{f_\ell(i)\}_{I \times \nu}$ denote the B-spline basis function matrix with $\nu = 30$ degrees of freedom, each subject’s ERP components were defined as $A_k = F\alpha_k$ where $\alpha_k = \{\alpha_{\ell r}\}_{\nu \times R}$ is the k -th subject’s matrix of basis function coefficients. Letting $\alpha_{k,r}$ denote the r -th column of α_k , we generated the $\alpha_{k,r}$ vectors independently from a multivariate normal distribution. More specifically, we assumed that $\alpha_{k,r} \stackrel{iid}{\sim} N_\nu(\mu_r, \sigma^2 \Sigma_r)$ where μ_r and Σ_r denote the mean and covariance matrices,³ and $\sigma \in \{0, 0.5, 1, 1.5\}$ controls the heterogeneity⁴ of the ERPs (see Fig. 2b). Note that $\sigma = 0$ corresponds to $A_k = A$ for all k , so the

³ The μ_r and Σ_r parameters were created based on the real data results to ensure that the mean ERPs (see Fig. 2a) and individual differences in the ERPs (see Fig. 2b) resembled those of the real data.

⁴ The $\sigma = 1$ condition corresponds to the level of individual differences present in the real-data example.

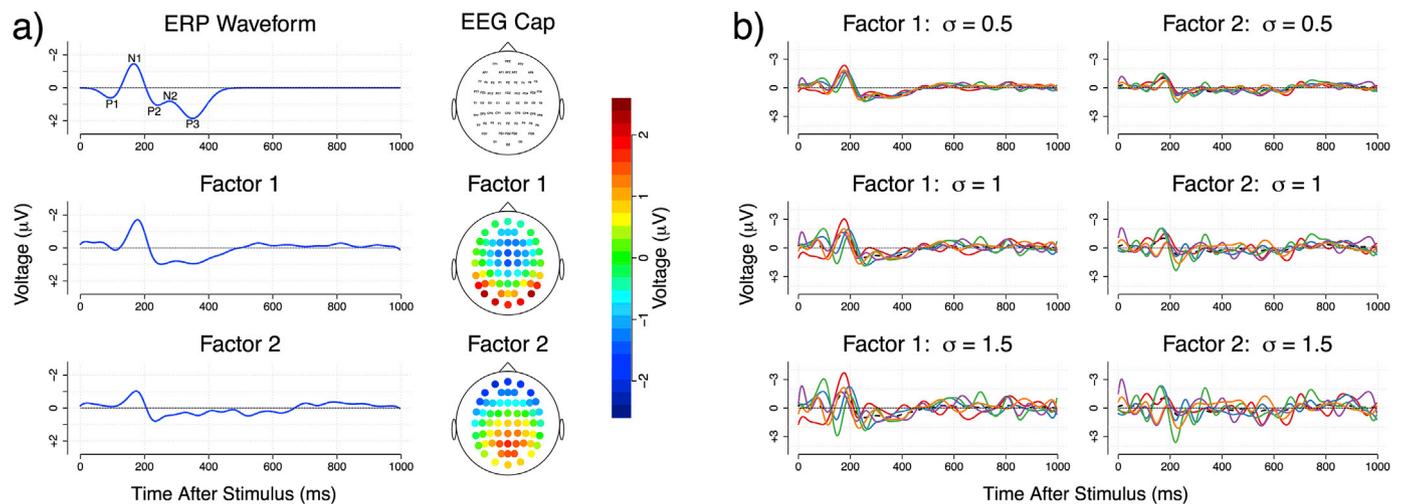


Fig. 2. Simulation Design Parameters. (a) The top row displays a prototypical event-related potential (ERP) waveform with five features (P1, N1, P2, N2, and P3), as well as the 61 channel electroencephalography (EEG) cap used in the simulation and real data examples. The middle and bottom rows show the data generating parameters for the simulation: the average Mode A (time) weights (i.e., $\mathbf{F}\mu_r$) and the Mode B (electrode) weights. (b) Simulated Mode A (time) weights corresponding to five random samples of the coefficients (i.e., $\mathbf{F}\alpha_{k,r}$) at the three different levels of heterogeneity. Note that when $\sigma = 0$ all of the subjects' ERPs are equal to their expected values, which are denoted with black dashed lines. Figures were created using the `eegkit` package (Helwig, 2018a) in R (R Core Team, 2019).

generated data follow the Parafac model form in this case. When $\sigma > 0$ the generated data meet the assumptions of the SCA-P model, but they do *not* meet the assumptions of the Parafac2 model—given that the cross-product constraint $\mathbf{A}_k^t \mathbf{A}_k = \Phi$ is not enforced as a part of the data generation.

3.1.2. Analyses

3.1.2.1. Data generation. For each of the 24 ($6 \text{ SNR} \times 4 \sigma$) cells of the simulation design, we generated 100 independent samples of data. The Mode A weights were generated as previously described, i.e., $\mathbf{A}_k = \mathbf{F}\alpha_k$ with $\alpha_{k,r} \stackrel{\text{iid}}{\sim} N_\nu(\mu_r, \sigma^2 \Sigma_r)$, the Mode B weights were fixed at the values displayed in Fig. 2a, and the Mode C weights were independently sampled from a uniform distribution on the interval [0.5, 1.5]. The errors were sampled independently from a standard normal distribution, and then were rescaled (i.e., multiplied by a constant) to achieve the needed SNR.

3.1.2.2. Model comparison. For each sample of generated data, we fit six models: (i) the Parafac model, (ii) the Parafac2 model, (iii) the SCA-P model with Varimax rotation (Kaiser, 1958), (iv) the SCA-P model with no rotation, which is the PCA solution, the (v) the GICA model using a FastICA rotation (Hyvärinen, 1999), and (vi) the GICA model (using FastICA) with dual regression. For the model comparison, the number of factors R and the degree of smoothness ν were fixed at the data generating values, i.e., $R = 2$ and $\nu = 30$, which makes it possible to compare the quality of the parameter estimates. The Parafac and Parafac2 models were fit to the data using a non-negativity constraint on the Mode C weights,⁵ and the \mathbf{C}_k weights for the SCA-P models were defined as the standard deviations of the columns of \mathbf{D}_k . For the Parafac and Parafac2 models, we used 10 random starts of the ALS algorithm. The models were fit using the `multiway` (Helwig, 2019) and `ica` (Helwig, 2018b) packages in R (R Core Team, 2019).

3.1.2.3. K-fold CV tuning. For each sample of generated data, we also

⁵ The non-negativity constraint ensures that the Mode C weights have a comparable interpretation in the Parafac, Parafac2, and SCA-P models, and has the added benefit of resolving the special sign indeterminacy of the Parafac2 model (Helwig, 2013).

explored the k-fold CV tuning methods proposed in Section 2.2.4. To select the model dimensionality, we evaluated the CV-CORCONDIA using $R = 1, \dots, 4$, and we compared three different thresholds for determining if the CORCONDIA is “close enough” to the optimal value of 100, i.e., the thresholds >80 , >85 , and >90 . To select the degree of smoothness ν , we evaluated the CV-MSE using 15 ν values ranging from 10 to 80 in increments of 5. We compared the results of our proposed selection rule (i.e., a threshold of $\delta = 0.2$ for the standardized CV-MSE change) to the results that would have been obtained using the optimal degree of smoothing (i.e., $\nu = 30$) and using no smoothing (or otherwise $\nu = 256$).

3.1.2.4. Performance metrics. The quality of each solution was quantified using various metrics to determine how well the method performed. We used the root mean squared error (RMSE) between the true and estimated tensors to quantify the quality of the tensor recovery, and we used the observed data R-squared to quantify the quality of the data recovery. Note that the R-squared value quantifies the difference between the estimated tensor and the observed data, whereas the RMSE value quantifies the difference between the estimated tensor and the true model tensor (i.e., without the error). Finally, to quantify the quality of the parameter recovery, we calculated the Tucker Congruence Coefficient (TCC) (Tucker, 1951) between the true and estimated weights, i.e., time courses and spatial maps. Values of the TCC closer to 1 indicate better agreement between the true and estimates parameters. See the SOM for further details on the performance metrics.

3.2. Application to EEG data

3.2.1. Data

To compare the performance of the various methods with real data, we use open-source EEG data collected from control and alcoholic subjects participating in a visual stimulus ERP study (Zhang et al., 1995). The data were collected in the Henri Begleiter Neurodynamics Laboratory at SUNY Downstate under the approval of the authors' home institutions. The dataset is freely available—thanks to Ingber (1997, 1998)—from the UCI Machine Learning repository (Dua and Graff, 2019). Our use of the open-source data adheres to the ethical standards of the authors' institution. The data consist of 61-channel ERPs recorded at 256 Hz for one second following the presentation of a visual stimulus from the Snodgrass and Vanderwart (1980) stimuli set. The data were collected from a total of 122 subjects (77 alcoholics and 45 controls) who were shown a single

stimulus (S1), which was followed by either a matching stimulus (S2m) or a non-matching stimulus (S2n). For our analyses, we focus on the subjects' initial response to the single stimulus, i.e., the S1 condition.

For each subject, multiple trials of the task were collected during each condition. Our analyses are based on a subset of 115 subjects (73 alcoholics and 42 controls) who have artifact free data for all electrodes on at least 20 trials of the S1 task. Note that we identified trials as having artifact (e.g., due to movement) if the absolute recorded voltage was greater than $100 \mu V$, which resulted in the exclusion of about 5% of the trials in the full dataset. The raw data were bandpass filtered using a window of 1–50 Hz, and then averaged across trials within each subject. To ensure an equal number of trials for each subject's mean data, we averaged the data across the first 20 trials for each subject. This resulted in a data tensor of dimension 256 time points \times 61 electrodes \times 115 subjects.

3.2.2. Analyses

The real data were analyzed using a similar procedure as was used in the Simulation Study. As a first step, the k-fold CV methods proposed in Section 2.2.4 were used to determine a reasonable number of components R and degree of smoothness ν for the data. The k-fold CV approach was applied in a similar fashion as before: (i) the CV-CORCONDIA was tuned with respect to $R = 1, \dots, 4$ to determine the number of components, and then (ii) the CV-MSE was tuned with respect to $\nu = (10, 15, \dots, 75, 80)$ to determine the degree of smoothness. Given the estimates of R and ν , we fit the six component models that were explored in the Simulation Study. Each model was fit using the same specifications that were used in the Simulation Study (see Section 3.1 for details).

4. Results

4.1. Simulation Study

4.1.1. Comparison of component models

Overview: The simulation results reveal noteworthy differences between the compared methods on the examined performance metrics, which are plotted in Fig. 3. We only plot the results for two (of the four) examined levels of heterogeneity, i.e., no heterogeneity ($\sigma = 0$) versus moderate heterogeneity ($\sigma = 1$), which is enough to make our key points.

Observed Data Recovery: For all levels of heterogeneity conditions, we found that the observed R-squared values tended to be ordered such that Parafac < Parafac2 < SCAP = GICA < GICA-DR. This ordering was expected, given that it represents the ordering of the methods with respect to the flexibility of the assumed model structure (from least to most flexible). With no heterogeneity present in the component time courses (Fig. 3 left), the differences between the R-squared values for the different methods was small. This was anticipated given that all of the methods are flexible enough to model the data when $\sigma = 0$. In contrast, when heterogeneity is present in the latent time courses (Fig. 3 right), the R-squared difference between Parafac and other methods is more substantial. This is not surprising, given that the Parafac model is not flexible enough to capture the individual differences that exist in the time courses when $\sigma = 1$.

Latent Tensor Recovery: Unlike the R-squared (which measures observed data fit), the RMSE measures how well the estimated tensor fits the unobserved data-generating tensor. As a result, the RMSE patterns in Fig. 3 reveal some interesting insights that are not discernible from the

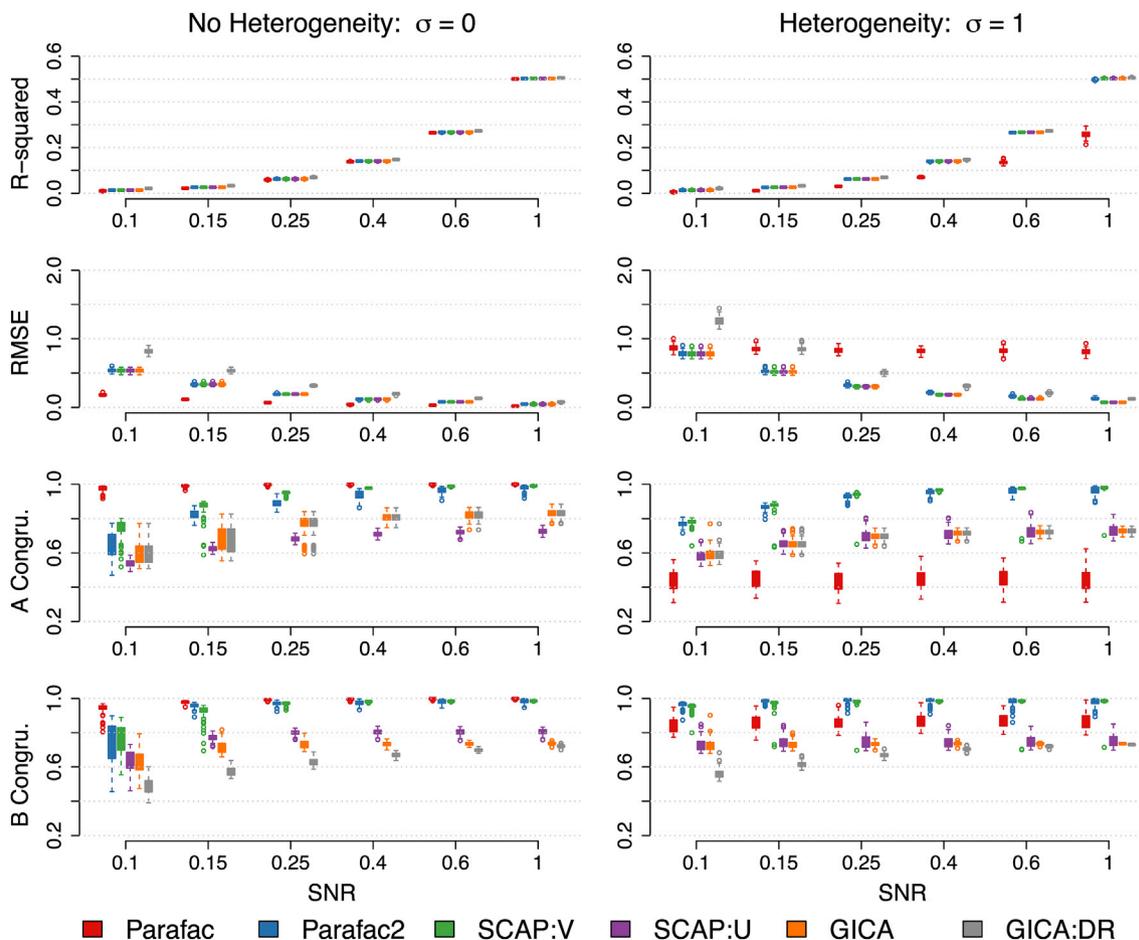


Fig. 3. Simulation Performance Results. The subplots display different performance metrics (rows) at different heterogeneity levels (columns). Within each subplot, the x-axis displays the results at the six different signal-to-noise ratios (SNRs). The six boxplots within each SNR denote the results from the six different methods.

observed data R-squared values. Focusing first on the $\sigma = 0$ condition, we see that the RMSE values are approximately ordered such that Parafac < Parafac2 = SCAP = GICA < GICA-DR. For all methods, the RMSE values tend to decrease as the SNR increases, which is expected in this case. For small SNR values, the Parafac solution produces noticeably smaller (i.e., better) RMSE values than the other solutions, whereas the GICA-DR solution produces noticeably larger (i.e., worse) RMSE values than the Parafac2, SCAP, and GICA methods.

The results show a somewhat different pattern when heterogeneity is present in the component time courses. In the $\sigma = 1$ case, the Parafac RMSE does not decrease as the SNR increases, given that the Parafac model is not flexible enough to model the heterogeneity in the data. Also, when $\sigma = 1$ we see that the Parafac2 model shows a slightly larger RMSE than the SCA-P and GICA models. This is not surprising given that the generated data follow the SCA-P model form—not the Parafac2 model form. What is surprising is that the Parafac2 model is able to perform nearly as well as the SCA-P model, even though the data-generating tensor does not meet the Parafac2 model assumptions.

Parameter Recovery: The bottom two rows of Fig. 3 show the TCC between the true parameters (i.e., time courses and spatial maps) and their estimates obtained using the six different component analysis methods. As a reminder, TCC values closer to 1 indicate better agreement between the parameters and the estimates. With no heterogeneity present in the component time courses ($\sigma = 0$), the Parafac model performs noticeably better than the other methods, particularly at the small SNRs. In contrast, when there is heterogeneity ($\sigma = 1$), the Parafac model shows a noteworthy reduction in performance, particularly with respect to the time course estimation. These results were expected, given that the Parafac model is appropriate for the $\sigma = 0$ condition, but not the $\sigma = 1$ condition. Unlike the Parafac results, the Parafac2 parameter recovery improves as the SNR increases for both the homogeneous and heterogeneous conditions. It is interesting to note that the Parafac2 model can well-recover the parameters when $\sigma = 1$, even though these data do not meet the Parafac2 model assumptions.

Comparing the results for the other methods provides a clear example of the important implications of the rotational indeterminacy problem. As a reminder, the two SCA-P models and the GICA model have the exact same fit measures (i.e., R-squared and RMSE), given that these three solutions are different rotations of the model in Fig. 1b. As the SNR increases, the Varimax rotation (coincidentally) produces rotated time courses and spatial maps that find the true data-generating parameters. In contrast, the PCA orientation (i.e., unrotated SCA-P) and the Fast ICA orientation (i.e., GICA) both fail to find the true data-generating parameters. This is because the true data generating spatial maps are moderately correlated (≈ 0.33), so the orthogonality and independence assumptions are not reasonable. Finally, it is worth noting that the GICA-DR approach resulted in worse spatial map estimates than using GICA without DR, but the difference between the approaches diminished as the SNR increased. This result was expected given that the true data-generating spatial maps were common across subjects.

Algorithm Runtimes: The model fitting times are ordered such that the SCA variants are the most efficient (≈ 0.02 sec), the Parafac model is next (≈ 0.1 sec), and the Parafac2 model is least efficient (≈ 0.2 sec). Note that these runtimes are in the expected order, given that SCA has a closed-form solution, whereas the Parafac and Parafac2 models rely on iterative algorithms. Tuning the Parafac2 model (using k-fold CV) for the number of factors R (≈ 2.9 – 3 s) and the degree of smoothness ν (≈ 4.5 – 5 s) was the major computational burden of the Parafac2 analysis. However, in practice, the tuning and fitting can be accomplished rather efficiently, given that the k-fold CV can be implemented in parallel on a cluster.

4.1.2. Choosing the number of factors

The results of our k-fold CV-CORCONDIA approach (for selecting the number of factors) are plotted in the top row of Fig. 4. As a reminder, we evaluated the CV-CORCONDIA for $R \in \{1, \dots, 4\}$ factors with thresholds

of >80 , >85 , and >90 . Across all 24 cells of the simulation design, the dimensionality of $R = 4$ was never selected. As a result, Fig. 4 only displays the selection results for $R \in \{1, 2, 3\}$. As expected, the results reveal that the number of selected factors tends to decrease as the CORCONDIA threshold increases. Furthermore, as expected, the selection results improve as the SNR increases. For all SNRs, the correct dimensionality ($R = 2$) is chosen most frequently. For the smaller SNRs, the threshold has a more noteworthy effect, such that over-selection of R is more common with a threshold of >80 . However, for larger SNRs, the different thresholds provide rather consistent results, and select the correct dimensionality of $R = 2$ for a majority of simulation replications.

4.1.3. Choosing the degree of smoothness

The results of our k-fold CV-MSE approach (for selecting the degree of smoothness) are plotted in the bottom three rows of Fig. 4. As a reminder, we evaluated the CV-MSE for $\nu \in \{10, 15, \dots, 75, 80\}$ degrees of freedom with a threshold of $\delta = 0.2$ standardized units for the CV-MSE change. We compared the performance metrics using our estimated ν to those obtained from the optimal (data-generating) ν and those obtained without smoothness constraints. The results depicted in Fig. 4 clearly demonstrate the effectiveness of our proposed k-fold CV-MSE approach. In particular, note that the performance metrics for the estimated ν value are (i) very similar to those of the optimal ν value, and (ii) substantially better than those obtained without including the smoothness constraints—particularly for small SNRs. Thus, these results highlight the benefits of including smoothness constraints in component models, and illustrate the potential of data-driven methods for automatic tuning of the degree of smoothness.

4.2. EEG application

4.2.1. Dimensionality and smoothness

The k-fold CV results for tuning the number of factors R and the degree of smoothness ν for the real EEG data are plotted in Fig. 5. As in the Simulation Study, we first tuned the Parafac2 model with respect to R using the k-fold CV-CORCONDIA proposed in Section 2.2.4. The results in Fig. 5 (left) clearly reveal that $R = 2$ factors should be preferred: note that $R = 2$ would be selected using any of the thresholds explored in the Simulation Study. Fixing $R = 2$, we tuned the model with respect to ν using the k-fold CV-MSE proposed in Section 2.2.4. The results in Fig. 5 (right) illustrate that the CV-MSE standardized differences seem to stabilize for $\nu \geq 30$, which is the value of ν selected by our heuristic. Consequently, throughout the remainder of the example, the results are explored using $R = 2$ factors and $\nu = 30$ degrees of freedom.

4.2.2. Model fit comparison

The Parafac model explained about 30% of the variation in the data, which was substantially smaller than the R-squared values for the other methods. This suggests that the Parafac model is not flexible enough to capture the heterogeneity that exists in the component time courses. The Parafac2, SCA-P, and GICA models all displayed similar R-squared values (0.633 for Parafac2 vs. 0.641 for others), which were more than double the R-squared value for the Parafac solution. The small difference in R-squared between the Parafac2 solution and the SCA-P/GICA solutions suggests that the Parafac2 model assumptions are reasonable for these data. In other words, constraining the time courses to have a common crossproduct matrix did not noticeably reduce the model's ability to fit the observed data. But the constraints have practical implications for interpreting the underlying components, as will be discussed in the next subsection.

4.2.3. Interpreting the components

The meaning of the components is typically inferred from the patterns of the estimated spatial maps and/or time courses. When applying the Parafac2 or SCA model, each subject has a unique time course matrix, so the interpretation of the components is typically inferred from the spatial

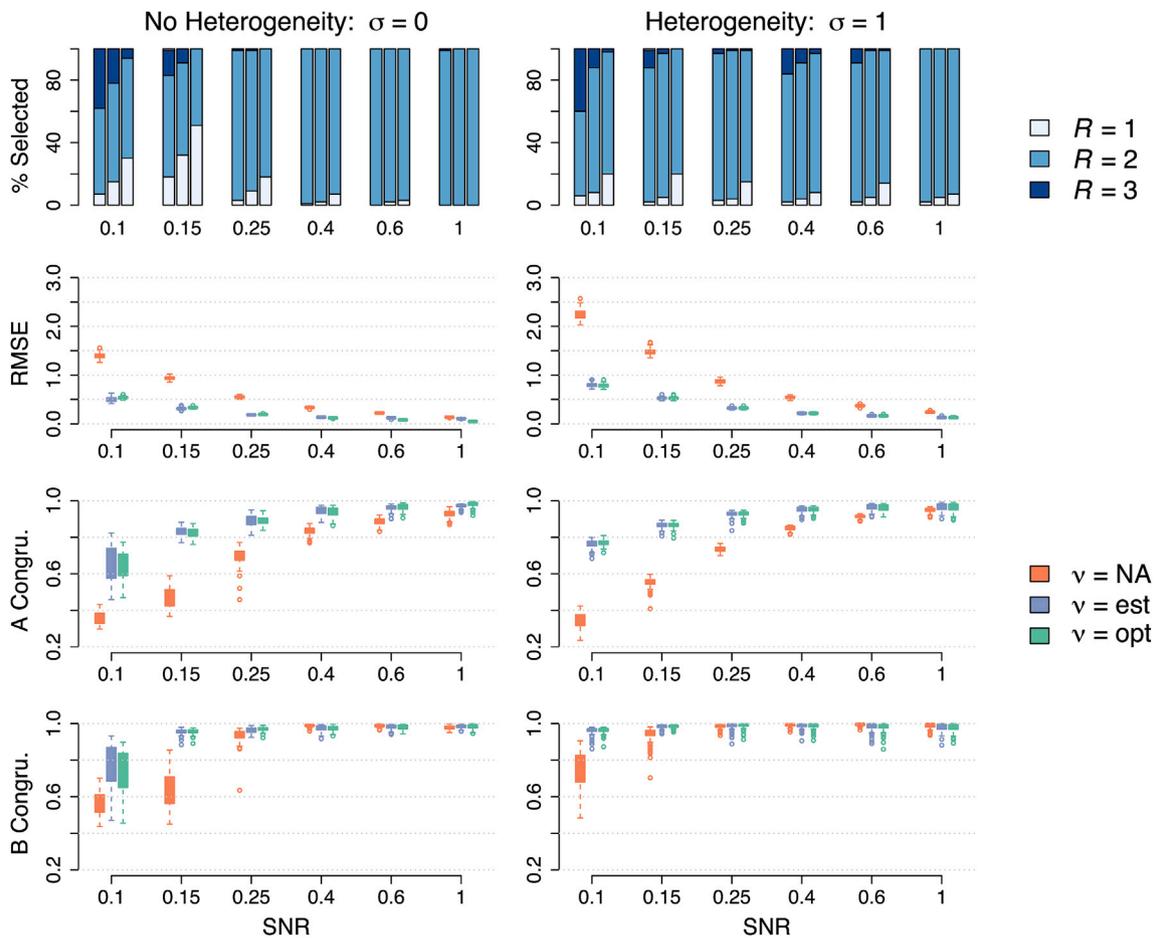


Fig. 4. Simulation Tuning Results. Top: The top row shows the results from our k-fold cross validation (CV) method to choose the number of components. Within each SNR, the 3 bars denote the three different CORCONDIA selection rules (>80 , >85 , >90), and the colors within each bar denote the percentage of times a given number of components ($R = 1, 2, 3$) was selected. Others: Rows 2–4 show the results from our k-fold CV method to choose the degree of smoothness. These plots display the tensor and parameter recovery using three different methods: (i) no smoothing, (ii) smoothing using the degrees of freedom ν estimated from our proposed approach, and (iii) smoothing using the optimal (i.e., data-generating) degrees of freedom ν .

maps. In Fig. 6 we plot the estimated spatial maps for the Parafac2 solution, the Varimax rotated SCA-P solution, the unrotated SCA-P solution, and the GICA solution. Note that the Parafac2 and Varimax rotated solutions produce a similar interpretation: Factor 1 shows voltage differences between the temporal-parietal versus frontal-central electrodes, whereas Factor 2 shows voltage differences between the central-parietal versus frontal electrodes. Furthermore, note that the unrotated SCA-P and GICA solutions produce a similar interpretation as one another, but this interpretation differs from the previous one: Factor 1 is parietal versus frontal component, and Factor 2 is a central versus peripheral

component. Which interpretation should we choose? This conundrum demonstrates the practical consequences of the rotational indeterminacy problem—as well as the power of Parafac2’s intrinsic axis property (i.e., lack of a rotational indeterminacy).

Without considering the Parafac2 solution, one might (incorrectly) conclude that the second interpretation should be preferred, e.g., because two of the three SCA-P rotations resulted in this solution. However, in light of the Parafac2 solution, it is clear that the Varimax rotated SCA-P solution (or the Parafac2 solution) should be preferred for interpretational purposes. This is because the Parafac2 model fits nearly as well as

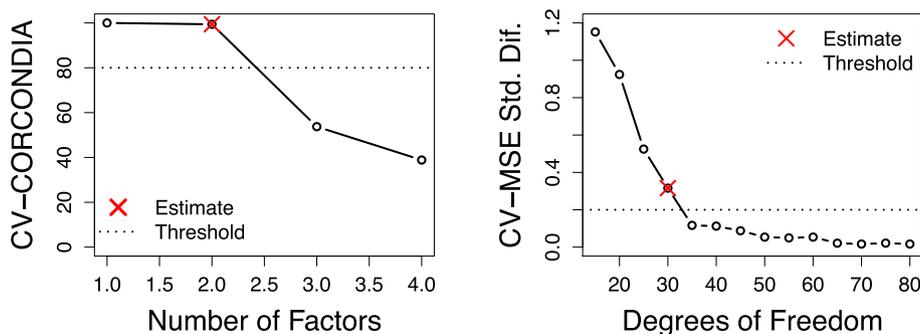


Fig. 5. Real Data Tuning Results. The cross-validation values of the CORCONDIA (left) and the MSE standardized difference (right) for the real EEG data application. The results suggest that $R = 2$ factors and $\nu = 30$ degrees of freedom are reasonable.

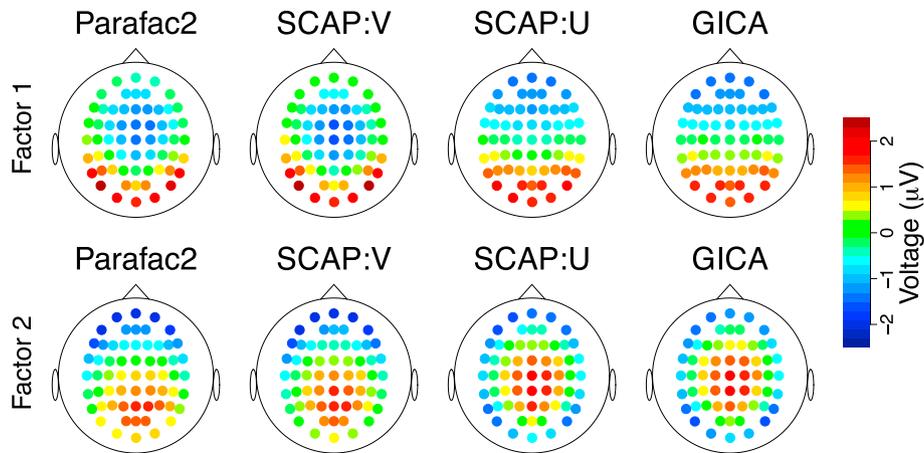


Fig. 6. Real Data Spatial Maps. The estimated spatial maps for factors 1 and 2 (rows) obtained via four different methods (columns). All four solutions have similar R-squared values (0.633 for Parafac2 vs. 0.641 for others), which suggests that the Parafac2 model is reasonable. Figures were created using the `eeegkit` package (Helwig, 2018a) in R (R Core Team, 2019).

the SCA-P model, and produces a unique interpretation of the underlying components—due to Parafac2’s intrinsic axis property. Based on the voltage difference patterns in the Parafac2 solution, we refer to Factors 1 and 2 as the *Ventral* and *Dorsal* factors, respectively. Historically, the ventral and dorsal streams have been distinguished as the “what” (i.e., object processing) versus the “where” (i.e., spatial processing) parts of the visual system (see Schneider, 1969; Mishkin et al., 1983). More recently, the distinction between the ventral and dorsal streams is often described as the “what” (i.e., vision for perception) versus the “how” (i.e., vision for action) parts of the visual system (see Goodale and Milner, 1992; Milner and Goodale, 2008).

4.2.4. Alcoholic versus control differences

The R-squared values previously reported were calculated ignoring group membership. Conditioning on the group, we find that the model fits the control subjects’ data (67.0% variance accounted for, VAF) slightly better than the alcoholic subjects’ data (59.6% VAF). Interestingly, if we calculate the factorwise R-squared separately for each group, we find some noteworthy differences. For the alcoholic subjects, the two factors have nearly equal influence such that the Dorsal Factor (30.5% VAF) is slightly more influential than the Ventral Factor (29.1% VAF). In contrast, for the control subjects we find that the Ventral Factor (39.3% VAF) is more influential than the Dorsal Factor (27.7% VAF). It is interesting that the alcoholic subjects have less coherence in the Ventral Factor (as evidenced by the reduced R-squared), which seems to imply an abnormality in the “what” processing of the information.

Examining the features of the ERP waveforms in Fig. 7, we observe some interesting differences in the timing and forms of the expected ERP trajectories for the alcoholic and control subjects. For the Ventral Factor, both groups have an obvious P1 feature that peaks about 105 ms after the stimulus; however, for the Dorsal Factor the P1 feature peaks about 19 ms later for the alcoholic subjects (109 ms) compared to the control subjects (90 ms). Note that the Dorsal P1 peak occurs slightly after the Ventral P1 peak for the alcoholic subjects, whereas the opposite is true for the controls. The P1 feature of the ERP is typically thought to reflect the “cost of attention” (Luck et al., 1994), so the delay of the Dorsal P1 peak in the alcoholic subjects may reflect a deficit in early spatial (or action) processing related to the cost of shifting one’s attention to a new stimulus.

The most prominent peak in each factor is the N1 feature, which is noticeably delayed in the alcoholic subjects. For the control subjects, the Dorsal N1 peak (172 ms) and the Ventral N1 peak (176 ms) occur slightly before the corresponding N1 peaks in the alcoholic subjects (Dorsal and Ventral: 180 ms). It is interesting to note that the alcoholics have a noteworthy delay in the N1 peak of both factors, even though the P1 peak was only delayed for the Dorsal Factor. Furthermore, the magnitude of

the N1 peak in the alcoholic subjects is reduced for both factors, providing further evidence that the alcoholics have an information processing abnormality early in the visual processing stream. Note that the amplitude of the N1 peak has been found to be moderated by attention (e.g., see Haider et al., 1964; Eason et al., 1969; Luck et al., 2000; Wascher et al., 2009). Thus, these differences provide further evidence that the visual processing abnormalities in alcoholics may relate to attentional deficits/costs early in the visual processing stream.

For both factors, the most noteworthy difference between the ERPs of the alcoholics and controls occurs about 200–450 ms after the stimulus. During this time window, the alcoholics display an overall reduction in the magnitude of the ERP, as well as a difference in the shape of the ERP waveform. For the Ventral Factor, the alcoholic ERP seems to lack the typical N2 and P3 features that are evident in the control ERP. Note that the reduction in amplitude of the P3 component of alcoholic subjects has been well-documented in the literature (e.g., see Porjesz et al., 1980, 1987; Porjesz and Begleiter, 1990a, b). Intriguingly, the Dorsal Factor of the alcoholic subjects seems to have prolonged N2 and P3 components, which peak substantially later than in the control subjects (≈ 80 –180 ms delay). Note that delays in the N2 and P3 features have been noted in previous visual stimulus ERP studies of alcoholics (see Emmerson et al., 1987; Fein and Chang, 2006; Fein and Andrew, 2011). This suggests that the “missing” N2 and P3 in the Ventral component may be related to a delayed/prolonged Dorsal stream processing of the visual stimulus.

5. Discussion

5.1. Synthesis of findings

The simulation results clearly demonstrate the potential of cross-validated tensor models, particularly Parafac2, for component analysis of multi-subject data. Note that the Parafac2 model performed well (with respect to both tensor and parameter recovery) across a variety of SNRs and heterogeneity levels. In contrast, the Parafac model only performed well when the data contained no heterogeneity in the component time courses. The SCA-P and GICA models were able to well-recover the tensor but not necessarily the parameters—due to the rotational indeterminacy. The SCA-P model with Varimax rotation worked well in this case, but cannot be expected to perform similarly well for other configurations of the spatial maps (i.e., Varimax has no guarantees to find the “true” components). The important point to take away is that the Parafac2 model was able to well-recover the Mode B (space) weights even when the data were generated from the SCA-P model. This suggests that the Parafac2 model can be useful for finding an ideal orientation (i.e., rotation) of the SCA-P solution—even when the Parafac2 model is not the

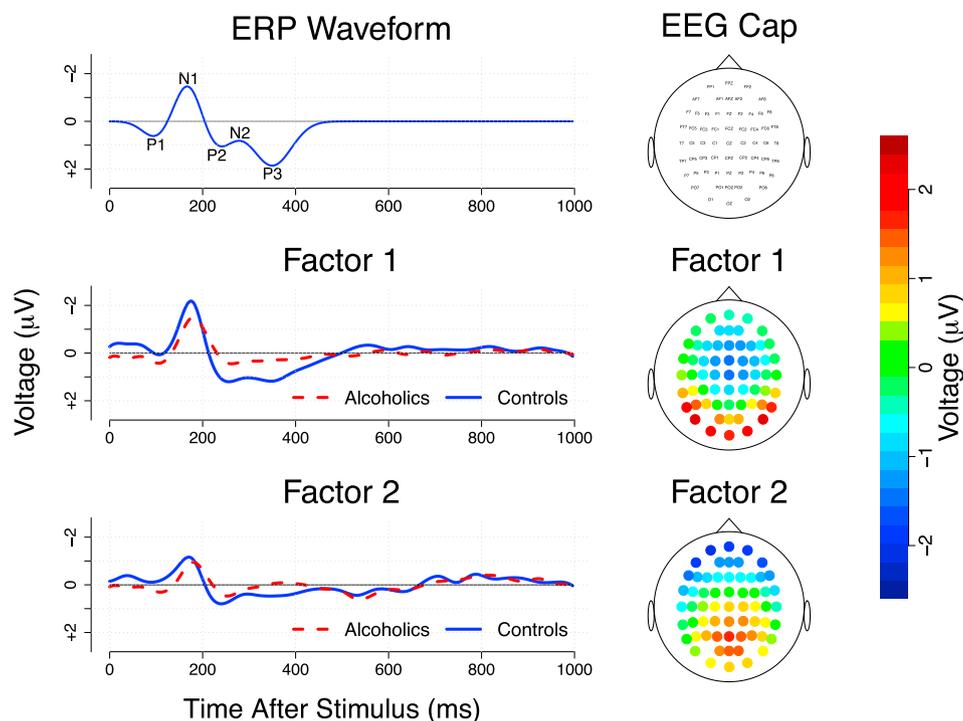


Fig. 7. Real Data Parafac2 Solution. The top row is identical to the top row in Fig. 2a. The middle and bottom rows show the estimated parameters: the average times courses (i.e., $A_k C_k$) for each group, as well as the Mode B (spatial) weights common to both groups. Figures were created using the `eegkit` package (Helwig, 2018a) in R (R Core Team, 2019).

correct model for the data.

The real data example offered a novel look at a classic dataset by comparing different multi-subject component analysis models (Parafac, Parafac2, SCA-P, GICA). The real data results reveal that the Parafac2 model seems appropriate for the data, given that (i) the Parafac model—which is less flexible—explained substantially less of the data variation, and (ii) the SCA-P model—which is more flexible—explained a similar amount of the data variation. Furthermore, the real data results reveal the practical significance of Parafac2’s uniqueness property. In particular, the SCA-P solution with $R = 2$ factors produced a nearly identical fit as the Parafac2 solution, but the SCA-P solution could produce several possible interpretations of the underlying brain networks (see Fig. 6). In contrast, the Parafac2 solution produced a unique result, where the $R = 2$ factors could be interpreted as the classic Ventral and Dorsal components of the visual processing stream.

The rotational indeterminacy of the SCA-P model poses a serious dilemma, given that different rotations—including different ICA algorithms—can produce different interpretations of the results. Although ICA inspired rotations are often considered in neuroimaging research, many other possible factor analytic rotations could be considered (see Browne, 2001; Bernaards and Jennrich, 2005). For any particular application, it is unclear which rotation should be preferred. This implies that two researchers (using two different rotations) could arrive at different interpretations of the underlying brain components, which is unsatisfactory. In contrast, we have demonstrated that the Parafac2 model can produce unique solutions that are useful for interpreting underlying brain components. Thus, when the Parafac2 and SCA-P models produce a similar fit to the data, the Parafac2 model should be preferred due to its intrinsic axis property.

5.2. Implications for task-based fMRI studies

Although we used EEG data in our application, our results have important implications for component analysis of other types of functional neuroimaging data, such as fMRI data. In past applications, the

TICA model has been the preferred approach for component analysis of multi-subject (task-based) fMRI data (Beckmann and Smith, 2005). Previous studies have noted that the TICA model can produce biased estimates of spatial maps when the true components are spatially dependent, and have recommended the Parafac model for component analysis of multi-subject task-based fMRI data (Stegeman, 2007; Helwig and Hong, 2013). However, these past studies failed to consider how un-modeled heterogeneity in the component time courses can negatively affect estimates of spatial maps. Our results clearly reveal that ignoring subject-specific heterogeneity in the component time courses can produce biased estimates of the latent time courses and spatial maps.

This implies that applying the Parafac (or TICA) model to data with temporal heterogeneity can be expected to produce misleading results. In particular, one should expect to over-factor the solution because “extra” factors will be needed to capture the un-modeled temporal heterogeneity. Substantial evidence suggests that there exists non-ignorable individual differences in haemodynamic response functions (e.g., see Handwerker et al., 2004). Consequently, the Parafac (or TICA) model may be too rigid for the analysis of typical task-based fMRI data, even when the task has the exact same stimulus function for each subject. To accommodate temporal heterogeneity in task-based fMRI studies, one might consider using the GICA model instead of the TICA model. However, our results suggest that the GICA model may produce misleading interpretations of the results, due to the rotational indeterminacy, and that the Parafac2 model has the potential to produce more valid results.

To illustrate this point, we used the `fmri` package (Tabelow and Polzehl, 2011) in R to simulate hypothetical fMRI data from a visual stimulus task. Like the EEG data example, the simulated fMRI data are comprised of $R = 2$ factors that are spatially dependent (see Fig. 8a). To introduce heterogeneity into the data, we randomly jittered the HRF parameters that define the expected BOLD signal for each factor (see Fig. 8b), which produces individual differences that do not conform to the Parafac2 model. We then analyzed the resulting *noise free* tensor to determine how well the various methods can find the data generating components. The entire data generating process was repeated 100 times

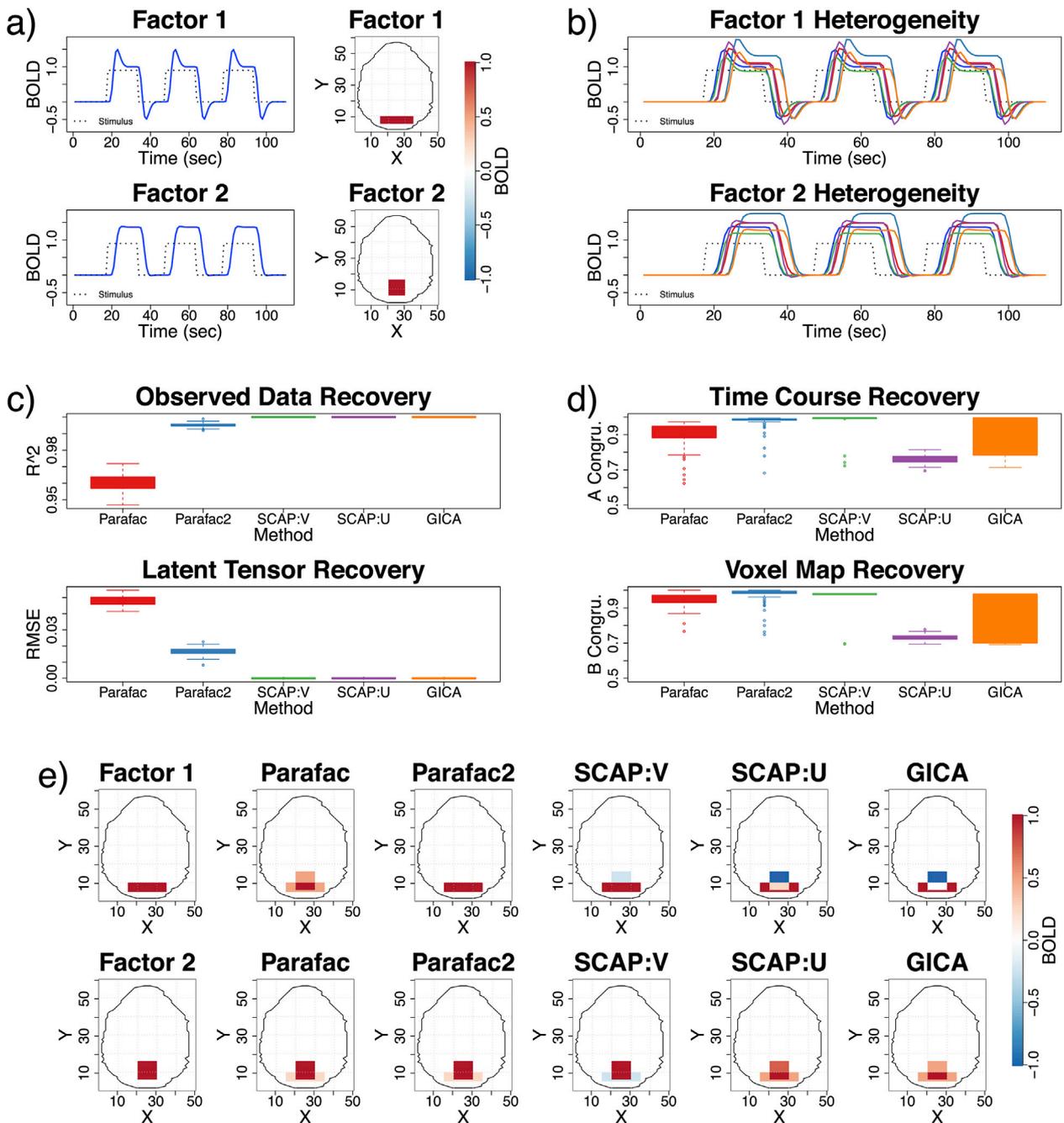


Fig. 8. Simulated fMRI Example. (a) Data generating time courses and voxel maps. (b) Example of simulated heterogeneity in latent time courses. (c) Recovery of data (R-squared) and tensor (RMSE) across 100 replications. (d) Recovery of parameters across 100 replications. (e) Examples of estimated voxel maps.

to examine the sensitivity of the solution for each method. Note that we did not include the smoothness constraints for the models, as there is no need to smooth the noiseless data.

The results reveal that the SCA-P and GICA models fit the data perfectly (as expected), such that they produce R-squared values of one and RMSE values of zero (see Fig. 8c). Compared to the SCA-P and GICA models, the Parafac2 model shows a slight reduction in fit, whereas the Parafac model shows a more noteworthy reduction in fit (see Fig. 8c). Note that to obtain a similar fit as the other methods, the Parafac model would require extracting “extra” factors, which would result in a misinterpretation of the data. Focusing on the parameter recovery (see Fig. 8d), we see that the Parafac model struggles to correctly recover the time courses *and* voxel maps, which demonstrates the interpretational consequences of ignoring the temporal heterogeneity. The Parafac2

model performs best at recovering the voxel maps, with the Varimax rotated SCA-P model also performing reasonably well.⁶

The unrotated SCA-P model (i.e., PCA solution) produces a systematic bias that results in a substantial misinterpretation of the results: the first factor has bipolar (positive and negative) activations, and the second factor is a blend of the two data generating factors (see Fig. 8e). The GICA solution is rather unstable, such that the method produces results that range between the unrotated and Varimax rotated SCA-P solutions. In approximately 2/3 of the replications, the GICA approach produces a

⁶ As in the real data example, the data generating components have a similar influence on the solution (in terms of the explained variation), which is why the Varimax rotation performs well.

solution similar to the Varimax rotated SCA-P model, whereas the other replications produce a solution more similar to the unrotated SCA-P model (see Fig. 8e). This highlights the dangers of relying solely on the GICA rotation of the SCA-P solution, which can produce misleading results when the factors are spatially dependent. In particular, the bipolar spatial maps produced by the unrotated SCA-P and GICA models (see Fig. 8e) might be misclassified as “noise components” (Griffanti et al., 2017) or misinterpreted as genuine components.

5.3. Applications to functional connectivity

Our results also have important implications for the application of component analysis models to resting-state fMRI data. In resting-state fMRI studies, component models are often used to provide lower-dimensional representations of the data, which can be used to understand functional connectivity patterns. In a recent study, Madsen et al. (2017) demonstrated that the GICA model and the Parafac2 model outperformed other component analysis strategies in terms of predicting between voxel covariances for new data. Interestingly, these authors noted that the GICA model often produces more interpretable components, whereas the Parafac2 model produces more stable results. Note that the better interpretability of the GICA solution seems to reflect the fact that the GICA model produces sparser components, which are more comparable to past ICA based results. But the authors do note that both the GICA and Parafac2 models “identify several components that are restricted within certain brain regions, allowing easier interpretation” (Madsen et al., 2017, p. 895).

When interpreting the results of Madsen et al. (2017), it is important to note that their study focuses on how well the various component models can predict between voxel covariances for new data. These authors did not address how well the component models can recover the true latent brain networks that generate the observed functional connectivity patterns between brain regions. The fact that the GICA and Parafac2 models both performed well with respect to predicting voxel-wise covariances suggests that the Parafac2 model structure may be reasonable to assume for resting-state fMRI. However, the result does *not* imply that the two methods perform similarly in terms of recovering the true voxel maps. This idea is analogous to our finding that the two methods can produce similar fit statistics (i.e., R-squared and RMSE), while producing rather different estimates of the model parameters (e.g., time courses and voxel maps). Note that this point is also made by Madsen et al., who discuss the interpretational differences between the GICA and Parafac2 results.

In practice, the imperative question is whether the Parafac2 model or the GICA model (or some other rotation of the SCA-P solution) is most appropriate for the given data. Due to the longstanding tradition of using ICA based methods in neuroimaging, it is likely that the GICA approach will produce components that are more familiar. However, this is an indication of the stability (or reliability) of the ICA algorithm as applied to neuroimaging data, and not necessarily an indication of the validity of the resulting components. To determine which approach produces a more genuine look at the functional organization of the brain, it is first necessary to determine what it means for a component to be “valid” in the first place. One reasonable possibility is to validate components via their ability to explain meaningful individual differences in behavioral constructs (Smith et al., 2015). Consequently, future functional connectivity studies should consider comparing the Parafac2 and GICA (and other SCA) models to determine which approach results in components with better predictive validity for external constructs of interest.

5.4. Extensions and future directions

In this paper, we applied smoothness constraints to a single mode of the data (i.e., the time mode). However, in some applications it may be desirable to smooth more than one mode of the data (e.g., see Leibovici,

2010; Choi et al., 2018). For example, in the application discussed in this paper, we could have assumed that the spatial maps were smooth functions of some 3D electrode coordinates. Such an extension could be quite useful in situations with a large number of spatial locations (e.g., high-density EEG or fMRI). Note that if the smooth functions can be represented with a relatively small ν , the smoothness constraints can greatly improve the computational efficiency of the model fitting.

As another possibility, we could consider a single-subject application where the K levels of Mode C correspond to K trials from the same subject (instead of data from K different subjects). In this case, it may be reasonable to assume that the Mode C (trial) weights are unknown smooth functions of the (sequential) trial index. For example, in a learning experiment it may be interesting to model the evolution of the subject’s spatiotemporal response across the trials of the experimental task. The smoothness constraints that we applied to the time courses could be easily extended to these scenarios, or other higher-way scenarios that involve smoothing multiple data modes.

In the current study—and the mentioned extensions—the fundamental difficulty is tuning the model, i.e., choosing the appropriate degrees of freedom for the latent functions, and choosing the model dimensionality. Our results reveal that data-driven metrics, such as k-fold cross-validation of the CORCONDIA and MSE, can provide useful heuristics for model selection. However, real-world model selection often requires careful comparison of competing models, where thought is given to a balance between the model’s fit, complexity, and interpretational value. More future work is needed to develop and explore robust model selection methods for tensor decompositions.

6. Conclusions

This paper discusses and compares various multi-subject component analysis models for analyzing task-based (or event-related) neuroimaging data. We recommend comparing a hierarchy of simultaneous component analysis models to determine a reasonable model for the data. Furthermore, we recommend using cross-validation to tune the number of factors and degree of smoothness of the time latent time courses. Our Monte Carlo simulation results highlight the benefits of the Parafac2 model, which outperformed the competing methods—even though the data were generated from the more general SCA-P model. Also, our real data results demonstrate that the Parafac2 model can produce interpretable solutions that reveal informative individual and group differences in latent brain networks.

Conflicts of interest

The authors have no conflicting interests.

Acknowledgements

NEH was supported by NIH grants 1U01MH108150-01A1 and 1R01MH112583-01A1 and a Single Semester Leave award from the College of Liberal Arts at the University of Minnesota. MAS was supported by a fellowship from the Graduate Research Partnership Program in the Department of Psychology at the University of Minnesota.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.116019>.

References

- Akaike, H., 1987. Factor analysis and AIC. *Psychometrika* 52, 317–332. <https://doi.org/10.1007/BF02294359>.
- Alsing, B., Kvalheim, O., 1993. Compression of n th-order data arrays by B-splines. part 1: theory. *J. Chemom.* 7, 61–73.

- Andersen, A.H., Rayens, W.S., 2004. Structure-seeking multilinear methods for the analysis of fmri data. *Neuroimage* 22, 728–739. <https://doi.org/10.1016/j.neuroimage.2004.02.026>.
- Beckmann, C.F., DeLuca, M., Devlin, J., Smith, S., 2005. Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Biol. Sci.* 360, 1001–1013.
- Beckmann, C.F., Mackay, C., Filippini, N., Smith, S.M., 2009. Group comparison of resting-state fmri data using multi-subject ica and dual regression. *Neuroimage* 47, S148.
- Beckmann, C.F., Smith, S.M., 2005. Tensorial extensions of independent component analysis for multisubject FMRI analysis. *Neuroimage* 25, 294–311.
- Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.
- Bernaards, C.A., Jennrich, R.I., 2005. Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educ. Psychol. Meas.* 65, 676–696.
- Bro, R., De Jong, S., 1997. A fast non-negativity-constrained least squares algorithm. *J. Chemom.* 11, 393–401.
- Bro, R., Kiers, H.A.L., 2003. A new efficient method for determining the number of components in parafac models. *J. Chemom.* 17, 274–286.
- Browne, M.W., 2001. An overview of analytic rotation in exploratory factor analysis. *Multivar. Behav. Res.* 36, 111–150.
- Calhoun, V., Adali, T., Pearlson, G., Pekar, J., 2001. Spatial and temporal independent component analysis of functional mri data containing a pair of task-related waveforms. *Hum. Brain Mapp.* 13, 43–53.
- Calhoun, V., Liu, J., Adali, T., 2009. A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. *Neuroimage* 45, S163–S172.
- Carroll, J.D., Chang, J.J., 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika* 35, 283–319.
- Cattell, R.B., 1944. “Parallel proportional profiles” and other principles for determining the choice of factors by rotation. *Psychometrika* 9, 267–283.
- Cattell, R.B., 1966. The scree test for the number of factors. *Multivar. Behav. Res.* 1, 245–276. https://doi.org/10.1207/s15327906mbr0102_10.
- Choi, J.Y., Hwang, H., Timmerman, M.E., 2018. Functional parallel factor analysis for functions of one- and two-dimensional arguments. *Psychometrika* 83, 1–20. <https://doi.org/10.1007/s11336-017-9558-9>.
- Cichocki, A., 2013. Tensor Decompositions: A New Concept in Brain Data Analysis? *CoRR* abs/1305.0395. URL: <http://arxiv.org/abs/1305.0395>.
- Comon, P., 1994. Independent component analysis, a new concept? *Signal Process.* 36, 287–314.
- Cong, F., Lin, Q.H., Kuang, L.D., Gong, X.F., Astikainen, J., Ristaniemi, T., 2015. Tensor decomposition of eeg signals: a brief review. *J. Neurosci. Methods* 248, 59–69. <https://doi.org/10.1016/j.jneumeth.2015.03.018>.
- de Boor, C., 2001. In: *A Practical Guide to Splines*, Revised ed. Springer-Verlag, New York.
- De Vos, M., Nion, D., Van Huffel, S., De Lathauwer, L., 2012. A combination of parallel factor and independent component analysis. *Signal Process.* 92, 2990–2999.
- Dua, D., Graff, C., 2019. UCI machine learning repository. URL: <http://archive.ics.uci.edu/ml>.
- Eason, R.G., Harter, M.R., White, C.T., 1969. Effects of attention and arousal on visually evoked cortical potentials and reaction time in man. *Physiol. Behav.* 4, 283–289. [https://doi.org/10.1016/0031-9384\(69\)90176-0](https://doi.org/10.1016/0031-9384(69)90176-0).
- Eckart, C., Young, G., 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1, 211–218.
- Emmerson, R.Y., Dustman, R.E., Shearer, D.E., Chamberlin, H.M., 1987. Eeg, visually evoked and event related potentials in young abstinent alcoholics. *Alcohol* 4, 241–248. [https://doi.org/10.1016/0741-8329\(87\)90018-8](https://doi.org/10.1016/0741-8329(87)90018-8).
- Faber, N.K.M., Bro, R., Hopke, P.K., 2003. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review. *Chemometr. Intell. Lab. Syst.* 65, 119–137.
- Fein, G., Andrew, C., 2011. Event-related potentials during visual target detection in treatment-naïve active alcoholics. *Alcohol Clin. Exp. Res.* 35, 1171–1179. <https://doi.org/10.1111/j.1530-0277.2011.01450.x>.
- Fein, G., Chang, M., 2006. Visual p300s in long-term abstinent chronic alcoholics. *Alcohol Clin. Exp. Res.* 30, 2000–2007. <https://doi.org/10.1111/j.1530-0277.2006.00246.x>.
- Ferdowsi, S., Abolghasemi, V., Sanei, S., 2015. A new informed tensor factorization approach to eeg-fmri fusion. *J. Neurosci. Methods* 254, 27–35. <https://doi.org/10.1016/j.jneumeth.2015.07.018>.
- Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.
- Griffanti, L., Douaud, G., Bijsterbosch, J., Evangelisti, S., Alfaro-Almagro, F., Glasser, M.F., Duff, E.P., Fitzgibbon, S., Westphal, R., Carone, D., Beckmann, C.F., Smith, S.M., 2017. Hand classification of fmri ica noise components. *Neuroimage* 154, 188–205. <https://doi.org/10.1016/j.neuroimage.2016.12.036>.
- Gu, C., 2013. In: *Smoother Spline ANOVA Models*, second ed. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4614-5369-7>.
- Haider, M., Spong, P., Lindsley, D.B., 1964. Attention, vigilance, and cortical evoked-potentials in humans. *Science* 145, 180–182. <https://doi.org/10.1126/science.145.3628.180>.
- Handwerker, D.A., Ollinger, J.M., D’Esposito, M., 2004. Variation of bold hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21, 1639–1651. <https://doi.org/10.1016/j.neuroimage.2003.11.029>.
- Harshman, R.A., 1970. Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multimodal factor analysis. *UCLA Work. Pap. Phonetics* 16, 1–84.
- Harshman, R.A., 1972. PARAFAC2: mathematical and technical notes. *UCLA Work. Pap. Phonetics* 22, 30–44.
- Harshman, R.A., Lundy, M.E., 1994. PARAFAC: parallel factor analysis. *Comput. Stat. Data Anal.* 18, 39–72.
- Harshman, R.A., Lundy, M.E., 1996. Uniqueness proof for a family of models sharing features of Tucker’s three-mode factor analysis and PARAFAC/CANDECOMP. *Psychometrika* 61, 133–154.
- Helwig, N.E., 2013. The special sign indeterminacy of the direct-fitting Parafac2 model: some implications, cautions, and recommendations for Simultaneous Component Analysis. *Psychometrika* 78, 725–739.
- Helwig, N.E., 2017. Estimating latent trends in multivariate longitudinal data via Parafac2 with functional and structural constraints. *Biom. J.* 15, 783–803. <https://doi.org/10.1002/bimj.201600045>.
- Helwig, N.E., 2018a. eegkit: toolkit for electroencephalography data. <http://CRAN.R-project.org/package=eegkit.R> package version 1.0-4.
- Helwig, N.E., 2018b. ica: independent component analysis. <http://CRAN.R-project.org/package=ica>. R.package version 1.0-2.
- Helwig, N.E., 2019. multiway: component models for multi-way data. <http://CRAN.R-project.org/package=multiway>. R.package version 1.0-6.
- Helwig, N.E., Hong, S., 2013. A critique of Tensor Probabilistic Independent Component Analysis: implications and recommendations for multi-subject FMRI data analysis. *J. Neurosci. Methods* 213, 263–273.
- Hitchcock, F.L., 1927. The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.* 6, 164–189. <https://doi.org/10.1002/sapm192761164>.
- Hong, S., Mitchell, S.K., Harshman, R., 2006. Bootstrap scree tests: a Monte Carlo simulation and applications to published data. *Br. J. Math. Stat. Psychol.* 59, 35–57.
- Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. <https://doi.org/10.1007/BF02289447>.
- Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24 (417–441), 498–520.
- Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10, 626–634.
- Hyvärinen, A., Karhunen, J., Oja, E., 2001. *Independent Component Analysis*. John Wiley & Sons, Inc., New York.
- Ingber, L., 1997. Statistical mechanics of neocortical interactions: canonical momenta indicators of electroencephalography. *Phys. Rev. E* 55, 4578–4593.
- Ingber, L., 1998. Statistical mechanics of neocortical interactions: training and testing canonical momenta indicators of eeg. *Math. Comput. Model.* 27, 33–64.
- Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23, 187–200.
- Kaiser, H.F., 1960. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* 20, 141–151. <https://doi.org/10.1177/001316446002000116>.
- Kamstrup-Nielsen, M.H., Johnsen, L.G., Bro, R., 2013. Core consistency diagnostic in PARAFAC2. *J. Chemom.* 27, 99–105. <https://doi.org/10.1002/cem.2497>.
- Kiers, H.A.L., Ten Berge, J.M.F., Bro, R., 1999. PARAFAC2–Part I. A direct fitting algorithm for the PARAFAC2 model. *J. Chemom.* 13, 275–294.
- Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. *SIAM Rev.* 51, 455–500.
- Krijnen, W.P., 2006. Convergence of the sequence of parameters generated by alternating least squares algorithms. *Comput. Stat. Data Anal.* 51, 481–489. <https://doi.org/10.1016/j.csda.2005.09.003>.
- Kruskal, J.B., 1977. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* 18, 95–138.
- Larsen, R., Warne, R.T., 2010. Estimating confidence intervals for eigenvalues in exploratory factor analysis. *Behav. Res. Methods* 42, 871–876. <https://doi.org/10.3758/BRM.42.3.871>.
- Leibovici, D.G., 2010. Spatio-temporal multiway decompositions using principal tensor analysis on k-modes: the R package PTAk. *J. Stat. Softw.* 34, 1–34. URL: <http://www.jstatsoft.org/v34/i10/>.
- Louwerson, D.J., Smilde, A.K., Kiers, H.A.L., 1999. Cross-validation of multiway component models. *J. Chemom.* 13, 491–510. [https://doi.org/10.1002/\(SICI\)1099-128X\(199909/10\)13:5<491::AID-CEM537>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1099-128X(199909/10)13:5<491::AID-CEM537>3.0.CO;2-2).
- Luck, S.J., Hillyard, S.A., Mouloua, M., Woldorff, M.G., Clark, V.P., Hawkins, H.L., 1994. Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological evidence for early selection. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 887–904. <https://doi.org/10.1037/0096-1523.20.4.887>.
- Luck, S.J., Woodman, G.F., Vogel, E.K., 2000. Event-related potential studies of attention. *Trends Cogn. Sci.* 4, 432–440. [https://doi.org/10.1016/S1364-6613\(00\)01545-X](https://doi.org/10.1016/S1364-6613(00)01545-X).
- Madsen, K.H., Churchill, N.W., Mørup, M., 2017. Quantifying functional connectivity in multi-subject fmri data using component models. *Hum. Brain Mapp.* 38, 882–899. <https://doi.org/10.1002/hbm.23425>.
- Makkiabadi, B., Jarchi, D., Sanei, S., 2011. Blind separation and localization of correlated p300 subcomponents from single trial recordings using extended parafac2 tensor model. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6955–6958. <https://doi.org/10.1109/IEMBS.2011.6091758>.
- Martínez-Montes, E., Sánchez-Bornot, J.M., Valdés-Sosa, P.A., 2008. Penalized PARAFAC analysis of spontaneous EEG recordings. *Stat. Sin.* 18, 1449–1464.
- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fmri data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6, 160–188.
- Milner, A.D., Goodale, M.A., 2008. Two visual systems re-viewed. *Neuropsychologia* 46, 774–785. <https://doi.org/10.1016/j.neuropsychologia.2007.10.005>.
- Mishkin, M., Ungerleider, L.G., Macko, K.A., 1983. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417. [https://doi.org/10.1016/0166-2236\(83\)90190-X](https://doi.org/10.1016/0166-2236(83)90190-X).

- Miwakeichi, F., Martínez-Montes, E., Valdés-Sosa, P.A., Nishiyama, N., Mizuhara, H., Yamaguchi, Y., 2004. Decomposing eeg data into space–time–frequency components using parallel factor analysis. *Neuroimage* 22, 1035–1045.
- Mørup, M., Hansen, L., Herrmann, C., Parnas, J., Arnfred, S., 2006. Parallel factor analysis as an exploratory tool for wavelet transformed event-related eeg. *Neuroimage* 29, 938–947.
- Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* 2, 559–572.
- Porjesz, B., Begleiter, H., 1990a. Event-related potentials for individuals at risk for alcoholism. *Alcohol* 7, 465–469.
- Porjesz, B., Begleiter, H., 1990b. Neuroelectric processes in individuals at risk for alcoholism. *Alcohol* 25, 251–256.
- Porjesz, B., Begleiter, H., Bihari, B., Kissin, B., 1987. The N2 component of the event-related brain potential in abstinent alcoholics. *Electroencephalogr. Clin. Neurophysiol.* 66, 121–131.
- Porjesz, B., Begleiter, H., Garozzo, R., 1980. Visual evoked potential correlates of information deficits in chronic alcoholics. In: Begleiter, H. (Ed.), *Biological Effects of Alcohol*. Plenum Press, pp. 603–623.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>. R version 3.6.0.
- Reis, M.M., Ferreira, M., 2002. PARAFAC with splines: a case study. *J. Chemom.* 18, 444–450.
- Revelle, W., Rocklin, T., 1979. Very simple structure: an alternative procedure for estimating the optimal number of interpretable factors. *Multivar. Behav. Res.* 14, 403–414. https://doi.org/10.1207/s15327906mbr1404_2.
- Schneider, G.E., 1969. Two visual systems. *Science* 163, 895–902. <https://doi.org/10.1126/science.163.3870.895>.
- Sidiropoulos, N.D., Bro, R., 2000. On the uniqueness of multilinear decomposition of n -way arrays. *J. Chemom.* 14, 229–239. [10.1002/1099-128X\(200005/06\)14:3<229::AID-CEM587>3.0.CO;2-N](https://doi.org/10.1002/1099-128X(200005/06)14:3<229::AID-CEM587>3.0.CO;2-N).
- Smith, S.M., Nichols, T.E., Vidaurre, D., Winkler, A.M., Behrens, T.E.J., Glasser, M.F., Ugurbil, K., Barch, D.M., Van Essen, D.C., Miller, K.L., 2015. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* 18, 1565–1567. <https://doi.org/10.1038/nn.4125>.
- Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: norms for the naming agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215.
- Stegeman, A., 2007. Degeneracy in Candecomp/Parafac and Indscal explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika* 72, 601–619.
- Tabelow, K., Polzehl, J., 2011. Statistical parametric maps for functional mri experiments in r: the package fmri. *J. Stat. Softw.* 44, 1–21. <https://doi.org/10.18637/jss.v044.i11>.
- Ten Berge, J.M.F., Kiers, H.A.L., 1996. Some uniqueness results for PARAFAC2. *Psychometrika* 61, 123–132.
- Timmerman, M.E., Kiers, H.A.L., 2002. Three-way component analysis with smoothness constraints. *Comput. Stat. Data Anal.* 40, 447–470.
- Timmerman, M.E., Kiers, H.A.L., 2003. Four simultaneous component analysis models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika* 68, 105–121.
- Tipping, M.E., Bishop, C.M., 1999. Mixtures of probabilistic principal component analysers. *Neural Comput.* 11, 443–482.
- Tomasi, G., Bro, R., 2006. A comparison of algorithms for fitting the PARAFAC model. *Comput. Stat. Data Anal.* 50, 1700–1734.
- Tucker, L.R., 1951. A Method for Synthesis of Factor Analysis Studies. Educational Testing Service, Princeton, NJ. Technical Report 984.
- Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31, 279–311.
- Wahba, G., 1990. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Philadelphia.
- Warne, R.T., Larsen, R., 2014. Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. *Psychol. Test Assess. Model.* 56, 104–123.
- Wascher, E., Hoffmann, S., Sanger, J., Grosjean, M., 2009. Visuo-spatial processing and the n1 component of the erp. *Psychophysiology* 46, 1270–1277. <https://doi.org/10.1111/j.1469-8986.2009.00874.x>.
- Williams, A.H., Kim, T.H., Wang, F., Vyas, S., Ryu, S.I., Shenoy, K.V., Schnitzer, M., Kolda, T.G., Ganguli, S., 2018. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron* 98, 1099–1115. <https://doi.org/10.1016/j.neuron.2018.05.015> e8.
- Zhang, X.L., Begleiter, H., Porjesz, B., Wang, W., Litke, A., 1995. Event related potentials during object recognition tasks. *Brain Res. Bull.* 38, 531–538.