University of Groningen

Functional Parallel Factor Analysis for Functions of One- and Two-dimensional Arguments

Choi, Ji Yeh; Hwang, Heungsun; Timmerman, Marieke

Published in:
Psychometrika

DOI:
[10.1007/s11336-017-9558-9](https://doi.org/10.1007/s11336-017-9558-9)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Choi, J. Y., Hwang, H., & Timmerman, M. (2018). Functional Parallel Factor Analysis for Functions of One- and Two-dimensional Arguments. *Psychometrika*, 83(1), 1-20. <https://doi.org/10.1007/s11336-017-9558-9>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

FUNCTIONAL PARALLEL FACTOR ANALYSIS FOR FUNCTIONS OF ONE- AND TWO-DIMENSIONAL ARGUMENTS

Ji YEH CHOI AND HEUNGSUN HWANG

MCGILL UNIVERSITY

MARIEKE E. TIMMERMAN

UNIVERSITY OF GRONINGEN

Parallel factor analysis (PARAFAC) is a useful multivariate method for decomposing three-way data that consist of three different types of entities simultaneously. This method estimates trilinear components, each of which is a low-dimensional representation of a set of entities, often called a mode, to explain the maximum variance of the data. Functional PARAFAC permits the entities in different modes to be smooth functions or curves, varying over a continuum, rather than a collection of unconnected responses. The existing functional PARAFAC methods handle functions of a one-dimensional argument (e.g., time) only. In this paper, we propose a new extension of functional PARAFAC for handling three-way data whose responses are sequenced along both a two-dimensional domain (e.g., a plane with x - and y -axis coordinates) and a one-dimensional argument. Technically, the proposed method combines PARAFAC with basis function expansion approximations, using a set of piecewise quadratic finite element basis functions for estimating two-dimensional smooth functions and a set of one-dimensional basis functions for estimating one-dimensional smooth functions. In a simulation study, the proposed method appeared to outperform the conventional PARAFAC. We apply the method to EEG data to demonstrate its empirical usefulness.

Key words: three-way data, parallel factor analysis, functional data analysis, spatial and temporal variation.

1. Introduction

Three-way data consist of three different types of entities simultaneously (e.g., subjects, variables, and occasions), each of which is also called a mode. A few examples of three-way data include neuroimaging data arranged as an array of brain locations by subjects by time points (e.g., Cox, 1996; Germond, Dojat, Taylor, & Garbay, 2000; Thirion & Fugeras, 2003), multivariate longitudinal data as an array of subjects by variables by occasions (e.g., Kroonenberg, 1987; Kuze et al., 1985; Oort, 2001), and gene expression data as an array of genes by subjects by time points (Holter, Maritan, Cieplak, Fedoroff, & Banavar, 2001; Xiao, Wang, & Khodursky, 2011).

Parallel factor analysis (PARAFAC) (Harshman, 1970), or equivalently canonical decomposition (CANDECOMP) (Carroll & Chang, 1970), is a well-known method for explaining variations in three-way data. PARAFAC aims to extract trilinear components, each of which is a low-dimensional representation of a mode, in such a way that they retain the maximum variance of the data. The number of the extracted components is the same across all modes. It generally provides unique solutions up to permutation, reflection and scaling of the components, under some mild conditions (e.g., Bro, 1997; Kruskal, 1977; Stegeman & Sidiropoulos, 2007) such that there exist no alternative ways to decompose the data that yield the same model fit. Thus, any post-processing technique (i.e., rotation) is unnecessary for the PARAFAC solution.

The authors wish to thank Jelena Ristic for her constructive comments on the analysis of EEG data.

The MATLAB code used in this paper is available upon request from the author Ji Yeh Choi.

Correspondence should be made to Ji Yeh Choi and Heungsun Hwang, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC H3A 1B1 Canada. Email: ji.yeh.choi@mail.mcgill.ca

PARAFAC obtains trilinear components from multivariate three-way data, where typically each mode's responses are unconnected measures on multiple variables or occasions. However, when a mode's responses can be sequenced along a continuum, such as time, frequency, or spatial location, they are no longer independent; it may be more natural to view them as a single connected entity or a function varying over the continuum. Conventional PARAFAC has not been well suited to extract trilinear components while treating entities in a mode as functions.

PARAFAC was recently extended to summarize three-way data whose responses were samples taken from smooth curves (e.g., Allen, 2013; Reis & Ferreira, 2002; Timmerman & Kiers, 2002). In these extensions, each mode's extracted components are considered functions of time and approximated as a linear combination of known basis functions, which is called the basis function expansion approach. Timmerman and Kiers (2002) decomposed three-way data of subjects by variables by occasions, such that the smooth component functions of occasions were approximated via this approach. It applied ordinary least squares to estimate the component functions, keeping the number of basis functions smaller than that of data points. Reis and Ferreira (2002) decomposed a three-way data considering a mode's components a function of time, similarly to Timmerman and Kiers (2002). However, they estimated the smoothing component functions via penalized least squares, adding roughness penalty terms to control for the amount of smoothness of a function. Overall, the existing methods decompose three-way data into trilinear components, taking into account the temporal variations in the entities of a certain mode(s) that are treated as functions of time. This enables to uncover and explicate the temporal dynamics of each component, thereby identifying the distinguishing characteristics of the various components. Moreover, by considering data functional rather than multivariate, it becomes more likely to arrive at stable estimates for components and to encounter less local minimum problems (Allen, 2013; Timmerman & Kiers, 2002).

Nevertheless, the existing methods are limited in that they are applicable only to functions of a single-dimensional argument, for example time. In practice, it is not uncommon that responses are evaluated as a function of a two-dimensional continuum, typically across different spatial locations. For instance, satellite and brain imaging data involve spatial properties, pertaining to a two-dimensional domain with x - and y -axis coordinates (e.g., Ramsay, 2002; Ramsay, Ramsay, & Sangalli, 2011; Sangalli, Ramsay, & Ramsay, 2013).

In this paper, we extend PARAFAC for the analysis of functions of multidimensional arguments. The proposed method aims to summarize three-way data that pertain both to functions varying over different x - and y -axis coordinates on a two-dimensional spatial domain, i.e., two-dimensional functions, and to one-dimensional functions. Specifically, the proposed method combines PARAFAC with basis function expansion approximations, where a set of piecewise quadratic finite element basis functions is for estimating two-dimensional smooth functions, and a set of one-dimensional basis functions is used for estimating one-dimensional smooth functions.

The paper is organized as follows. In Section 2, the technical underpinnings of the proposed method are provided in detail. An alternating regularized least squares algorithm is developed for parameter estimation. In Section 3, a simulation study is conducted to investigate the performance of the proposed method. In Section 4, the method is applied to real data to demonstrate its empirical usefulness. The final section summarizes the implications of the proposed method and discusses its prospective extensions.

2. Methods

Let $\underline{\mathbf{Z}}$ ($N \times J \times S$) denote a three-way data array, where N , J , and S denote the numbers of entities in the first, second, and third modes, respectively. Suppose that entities in the first mode are multivariate (e.g., subjects) ($n = 1, \dots, N$), whereas those in the remaining modes

are functional. More specifically, entities in the second mode are assumed to be generated from smooth functions varying over a one-dimensional continuum (e.g., time) with the interval $[T_1, T_2]$, within which a set of J distinct time instants are observed ($t_j \in [T_1, T_2]$; $j = 1, \dots, J$). Entities in the third mode are assumed to be generated from two-dimensional smooth functions on a two-dimensional domain, denoted by Ω , with x - and y -axis coordinates. Within Ω , a set of S space locations are observed ($p_s = (x_s, y_s) \in \Omega$; $s = 1, \dots, S$). Let R denote the number of components ($r = 1, \dots, R$). Let $\mathbf{A} (N \times R)$, $\mathbf{B} (J \times R)$, and $\mathbf{C} (S \times R)$ denote the matrices of components for the first mode with elements a_{nr} , for the second mode with elements $b_r(t_j)$, and for the third mode with elements $c_r(p_s)$, respectively, where $b_r(t_j)$ indicates the evaluation of the r th one-dimensional smooth function $\mathbf{b}_r$ at the t_j th time instant, and $c_r(p_s)$ indicates the evaluation of the r th two-dimensional smooth function $\mathbf{c}_r$ at the p_s th spatial location.

The problem of the proposed method can be formulated as minimizing the following optimization criterion

$$\begin{aligned} \phi = & \sum_{n=1}^N \sum_{j=1}^J \sum_{s=1}^S \left(z_n(p_s, t_j) - \sum_{r=1}^R a_{nr} b_r(t_j) c_r(p_s) \right)^2 + \lambda \sum_{r=1}^R \int_{T_1}^{T_2} (D^2 \mathbf{b}_r)^2 dT \\ & + \rho \sum_{r=1}^R \int_{\Omega} (\Delta \mathbf{c}_r)^2 d\Omega \end{aligned} \quad (1)$$

with respect to a_{nr} , $\mathbf{b}_r$, and $\mathbf{c}_r$, where λ and ρ denote nonnegative smoothing parameters. Due to scaling indeterminacy, we may scale two of three component matrices by multiplying a scalar and compensate this scale by multiplying the inverse of the same scalar in the remaining component matrix (Bro, 1997). In addition, because the order of the components is not identified, the columns of the component matrices would be rearranged in descending order of variance explained (e.g.,

Kolda & Bader, 2009). In (1), $\sum_{r=1}^R \int_{T_1}^{T_2} (D^2 \mathbf{b}_r)^2 dT$ is called a roughness penalty that refers to the sum of the integrated squared second derivative of $\mathbf{b}_r$ over R components. This is indicative of the amount of roughness or curvature of one-dimensional functions, because the second derivative of a straight line is equal to zero and its value is expected to increase as a function becomes more variable. Similarly, $\sum_{r=1}^R \int_{\Omega} (\Delta \mathbf{c}_r)^2 d\Omega$ is used as a roughness penalty for the two-dimensional functions spanned on Ω . It involves a second-order partial derivative of $\mathbf{c}_r$ with respect to each x - and y -axis coordinates on Ω , where $\Delta \mathbf{c}_r$ is also referred to as the Laplace operator and computed as $\frac{\partial^2 \mathbf{c}_r}{\partial x^2} + \frac{\partial^2 \mathbf{c}_r}{\partial y^2}$. The Laplace operator remains invariant to both rotations and translations of the x - and y -axis coordinates (e.g., Folland, 1995; Rosenberg, 1997).

The smoothing parameters (λ and ρ) control for the relative importance of the roughness penalty terms in estimating $\mathbf{b}_r$ and $\mathbf{c}_r$. For example, when $\lambda = \rho = 0$, criterion (1) reduces to an ordinary least squares criterion. As the smoothing parameters become larger, more emphasis will be placed on the two penalty terms, leading the estimated functions to be smoother. The values of λ and ρ may be determined by using K -fold cross-validation (Hastie et al., 2009), which will be discussed later in this section.

Let $\Psi(t)$ denote an M by 1 vector of known basis functions evaluated at a time argument t . Let $\mathbf{w}_r$ denote an M by 1 vector of coefficients for the basis functions for the r th component. Each one-dimensional smooth function can then be approximated based on a basis function expansion, as follows:

$$\mathbf{b}_r(t) = \sum_{m=1}^M \psi_m(t) w_{mr} = \Psi(t)' \mathbf{w}_r. \quad (2)$$

Although various types of basis functions are available (Ramsay & Silverman, 2005, Chapter 3), the Fourier and spline basis functions are used most widely: the former for representing a stable periodic function, whereas the latter for approximating a non-periodic function that may fluctuate locally (Ramsay, Hooker, & Graves, 2009).

Two-dimensional smooth functions may be estimated by using the finite element method (FEM) (e.g., Pepper & Heinrich, 2006; Zienkiewicz, 1971; Zienkiewicz & Taylor, 2000) to simplify the original domain to a finite element function space of piecewise quadratic functions. Let $\xi(p) = (\xi_1, \dots, \xi_L)'$ denote an L by 1 vector of nodal basis functions evaluated at a spatial argument on Ω , where L indicates the total number of nodes. Let $\mathbf{f}_r = (f_r(\delta_1), \dots, f_r(\delta_L))'$ denote an L by 1 vector of values of c_r at L nodes $\{\delta_1, \dots, \delta_L\}$. In FEM, the original domain is partitioned into disjoint triangles, each of which consists of a set of six nodes (i.e., including three vertices and three edge midpoints in each triangle) and its associated set of six nodal basis functions. The quadratic polynomial over each triangle is represented as a linear combination of six nodal basis functions, each of which has 1 at a single node and 0 at the remaining five. Then, the r th two-dimensional smooth function spanned on Ω is constructed on the finite element space as follows:

$$\mathbf{c}_r(p) = \mathbf{f}_r' \xi(p). \quad (3)$$

An alternating regularized least squares (ARLS) algorithm (e.g., Hwang, 2009) is developed to minimize (1) with respect to a_{nr} , $\mathbf{b}_r$, and $\mathbf{c}_r$. In this algorithm, each of the unknown parameters is updated in a least squares sense, while the other parameters are fixed temporarily. Upon convergence, the so-called FIT percentage can be calculated as a measure of overall model fit, which indicates how much variance of the data is explained by a given model. The FIT percentage is given by $(1 - \phi/SS(\mathbf{X})) \times 100$ (%) (e.g., Bro & Kiers, 2003).

Given R , the algorithm alternates the following three steps until convergence.

Step 1 Update a_{nr} for fixed $\mathbf{b}_r$ and $\mathbf{c}_r$. In matrix notation, this is equivalent to minimizing

$$\begin{aligned} \phi_A &= SS(\mathbf{Z}_A - \mathbf{A}(\mathbf{C} \otimes \mathbf{B}))' \\ &= SS(\mathbf{Z}_A - \mathbf{A}\mathbf{Q}'), \end{aligned} \quad (4)$$

with respect to $\mathbf{A}$, where $\mathbf{Q} = (\mathbf{C} \otimes \mathbf{B})$, $\mathbf{Z}_A$ is a matrix with N rows and $(J \times S)$ columns, which is unfolded from $\underline{\mathbf{Z}}$ by arranging an N by J frontal matrix of $\underline{\mathbf{Z}}$ S times consecutively, and $\otimes$ is a column-wise Kronecker product of two matrices, called the Khatri–Rao product. Then, $\mathbf{A}$ is updated by

$$\hat{\mathbf{A}} = \mathbf{Z}_A \mathbf{Q} (\mathbf{Q}' \mathbf{Q})^{-1}. \quad (5)$$

Step 2 Update $\mathbf{b}_r$ for fixed a_{nr} and $\mathbf{c}_r$. From (2), let $\Psi = [\Psi(T_1), \dots, \Psi(T_2)]$ denote an M by J matrix of basis functions evaluated at $t_1 \dots t_J$ time instants given that $t_1 = T_1$ and $t_J = T_2$, and $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_R]$ denote an M by R matrix of coefficients for the basis functions. By combining (2) into (1), the optimization criterion can be rewritten as

$$\begin{aligned} \phi_B &= SS(\mathbf{Z}_B - \Psi' \mathbf{W} (\mathbf{C} \otimes \mathbf{A}))' + \lambda \text{tr}(\mathbf{W}' \Theta_T \mathbf{W}) \\ &= SS(\mathbf{Z}_B - \Psi' \mathbf{W} \mathbf{W}') + \lambda \text{tr}(\mathbf{W}' \Theta_T \mathbf{W}) \\ &= SS(\text{vec}(\mathbf{Z}_B) - \text{vec}(\Psi' \mathbf{W} \mathbf{W}')) + \lambda \text{tr}(\mathbf{W}' \Theta_T \mathbf{W}) \\ &= SS(\text{vec}(\mathbf{Z}_B) - (\mathbf{V} \otimes \Psi') \text{vec}(\mathbf{W})) + \lambda \text{vec}(\mathbf{W})' \text{vec}(\Theta_T \mathbf{W}) \\ &= SS(\text{vec}(\mathbf{Z}_B) - \Pi \text{vec}(\mathbf{W})) + \lambda \text{vec}(\mathbf{W})' (\mathbf{I}_R \otimes \Theta_T) \text{vec}(\mathbf{W}), \end{aligned} \quad (6)$$

with $\Theta_T = \int D^2\psi(t)D^2\psi(t)'dt$, where $\mathbf{V} = (\mathbf{C}|\otimes|\mathbf{A})$, $\mathbf{\Pi} = (\mathbf{V} \otimes \mathbf{\Psi}')$, $\mathbf{Z}_B$ is a J by $(N \times S)$ matrix placing the n th slice of $\underline{\mathbf{Z}}$ (i.e., a J by N matrix) consecutively, $\text{vec}(\mathbf{G})$ indicates a supervector formed by stacking all columns of $\mathbf{G}$ one below another, and $\mathbf{I}_R$ is an R by R identity matrix. Then, $\text{vec}(\hat{\mathbf{W}})$ is estimated by

$$\text{vec}(\hat{\mathbf{W}}) = [\mathbf{\Pi}'\mathbf{\Pi} + \lambda(\mathbf{I}_R \otimes \mathbf{\Theta}_T)]^{-1} \mathbf{\Pi}' \text{vec}(\mathbf{Z}_B), \quad (7)$$

The $\text{vec}(\hat{\mathbf{W}})$ is reshaped into $\mathbf{W}$. Afterward, $\mathbf{W}$ is inserted back to (2) to estimate evaluations of the one-dimensional smooth function $\mathbf{b}_r$ at T time instants.

Step 3 Update $\mathbf{c}_r$ for fixed a_{nr} and $\mathbf{b}_r$. Putting (3) in matrix notation, let $\mathbf{\Xi} = [\xi(p_1), \dots, \xi(p_S)]$ denote an L by S matrix of the evaluations of the L spatial basis functions in the S space locations, and $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_R]$ denote an L by R matrix of the coefficients in (3). By inserting (3) into (1), the optimization criterion becomes equivalent to

$$\begin{aligned} \phi_C &= \text{SS}(\mathbf{Z}_C - \mathbf{\Xi}'\mathbf{F}(\mathbf{B}|\otimes|\mathbf{A})') + \rho \text{tr}(\mathbf{F}'\mathbf{\Theta}_S\mathbf{F}) \\ &= \text{SS}(\text{vec}(\mathbf{Z}_C) - \text{vec}(\mathbf{\Xi}'\mathbf{F}\mathbf{U}')) + \rho \text{tr}(\mathbf{F}'\mathbf{\Theta}_S\mathbf{F}) \\ &= \text{SS}(\text{vec}(\mathbf{Z}_C) - (\mathbf{U} \otimes \mathbf{\Xi}') \text{vec}(\mathbf{F})) + \rho \text{vec}(\mathbf{F})' \text{vec}(\mathbf{\Theta}_S\mathbf{F}) \\ &= \text{SS}(\text{vec}(\mathbf{Z}_C) - \mathbf{\Gamma} \text{vec}(\mathbf{F})) + \rho \text{vec}(\mathbf{F})' (\mathbf{I}_R \otimes \mathbf{\Theta}_S) \text{vec}(\mathbf{F}), \end{aligned} \quad (8)$$

where $\mathbf{Z}_C$ is a S by NJ matricized version of $\underline{\mathbf{Z}}$, $\mathbf{U} = (\mathbf{B}|\otimes|\mathbf{A})$, $\mathbf{\Gamma} = (\mathbf{U} \otimes \mathbf{\Xi}')$, and $\mathbf{\Theta}_S = \int \left(\frac{\partial^2 \xi(p)}{\partial x^2} + \frac{\partial^2 \xi(p)}{\partial y^2} \right)^2 d\Omega$. Let $\mathbf{J}_0$ denote the L by L matrix of the integrals of the cross products of the L nodal basis functions (i.e., $\int \xi(p)\xi(p)'d\Omega$) and $\mathbf{J}_1$ denote the L by L matrix of the integrals of the sum of first partial derivatives of the nodal basis functions with respect to each coordinate (i.e., $\int (\xi_x \xi'_x + \xi_y \xi'_y) d\Omega$, where $\xi_x = \left(\frac{\partial \xi_1}{\partial x}, \dots, \frac{\partial \xi_L}{\partial x} \right)'$ and $\xi_y = \left(\frac{\partial \xi_1}{\partial y}, \dots, \frac{\partial \xi_L}{\partial y} \right)'$). Following the procedure used for achieving spatial smoothing via FEM, $\mathbf{\Theta}_S$ is approximated as $\mathbf{J}_1 \mathbf{J}_0^{-1} \mathbf{J}_1$ without solving for the second-order partial derivatives, whose derivation will be discussed in this section shortly. From (8), $\text{vec}(\hat{\mathbf{F}})$ is estimated by

$$\text{vec}(\hat{\mathbf{F}}) = [\mathbf{\Gamma}'\mathbf{\Gamma} + \rho(\mathbf{I}_R \otimes \mathbf{\Theta}_S)]^{-1} \mathbf{\Gamma}' \text{vec}(\mathbf{Z}_C), \quad (9)$$

The $\text{vec}(\hat{\mathbf{F}})$ is reshaped into $\mathbf{F}$. Afterward, $\mathbf{F}$ is inserted back to (2) to estimate evaluations of the two-dimensional smooth function $\mathbf{c}_r$ at S spatial locations.

For the derivation of $\mathbf{\Theta}_S$ in (8), we focus on estimating the smooth function $\mathbf{c}_r$ itself. For the moment, let h_{sr} indicate the s th row element in the r th column of $\mathbf{C}$ while treating the elements as multivariate. Given the discrete observations of $\mathbf{C}$, the r th two-dimensional smooth function would be updated by finding the function $\mathbf{c}_r$ that minimizes

$$\sum_{s=1}^S \{h_{sr} - \mathbf{c}_r(p_s)\}^2 + \rho \int (\Delta \mathbf{c}_r)^2 d\Omega, \quad (10)$$

Let $\mathcal{O}^2(\Omega)$ denote a functional space that consists of continuous functions on Ω whose second-order partial derivatives are square integrable, and $\mathcal{O}_\bullet^2(\Omega)$ denote a subset of $\mathcal{O}^2(\Omega)$ consisting of continuous functions whose derivatives are zero on the boundary of Ω . By introducing an

auxiliary function g ($g \in O^2(\Omega)$) and using Green's theorem or integration by parts, the problem of updating $\mathbf{c}_r$ is reformulated to finding $(\mathbf{c}_r, g) \in O^2_\bullet(\Omega) \times O^2(\Omega)$ that satisfy

$$\begin{aligned} \sum_{s=1}^S \mathbf{c}_r(p_s) u(p_s) - \rho \int (\nabla g \nabla u) &= \sum_{s=1}^S u(p_s) h_{sr} \\ \int g v + \int (\nabla \mathbf{c}_r \nabla v) &= 0, \end{aligned} \quad (11)$$

for every pair of functions (u, v) spanned on $O^2_\bullet(\Omega)$ and $O^2(\Omega)$, respectively. In (11), ∇ indicates the gradient operator. This reformulation of (11) guarantees a unique solution, imposing the constraint of having no curvature on the boundary of the spatial domain (Sangalli, Ramsay, & Ramsay, 2013). Hence, incorporating this spatial smoothing via FEM would keep the uniqueness property of PARAFAC intact.

Based on the fact that an arbitrary two-dimensional function in the finite element space can be expressed as a linear combination of nodal basis functions [i.e., as provided in (3)], functions in (11) can be evaluated in the finite element space. Let $\sum_{s=1}^S \mathbf{c}_r(p_s) u(p_s) = \mathbf{u}^* \mathbf{\Lambda} \mathbf{f}'_r$, where $\mathbf{u}^* = (\mathbf{u}(\delta_1), \dots, \mathbf{u}(\delta_L))'$ and $\mathbf{\Lambda}$ is an L by L diagonal matrix whose l th diagonal entry is 1 if the l th node is a data point and 0 otherwise. Let $\mathbf{h}_r = [h_{1r}, \dots, h_{Sr}]'$ denote an S by 1 vector that contain the r th column of discrete observations of $\mathbf{C}$. Equation (11) can be rewritten as the following linear system

$$\begin{aligned} \mathbf{u}^* \mathbf{\Lambda} \mathbf{f}'_r - \rho \mathbf{u}^* \mathbf{J}_1 \mathbf{g}^* &= \mathbf{u}^* \mathbf{h}_r \\ \mathbf{v}^* \mathbf{J}_0 \mathbf{g}^* + \mathbf{v}^* \mathbf{J}_1 \mathbf{f}'_r &= \mathbf{0}, \end{aligned} \quad (12)$$

From the second equation in (12), we would get $\mathbf{g}^* = -\mathbf{J}_0^{-1} \mathbf{J}_1 \mathbf{f}'_r$. Inserting the resultant $\mathbf{g}^*$ to the first equation in (12), we can approximate $\mathbf{\Theta}_S$ as $\mathbf{J}_1 \mathbf{J}_0^{-1} \mathbf{J}_1$. More details of Green's theorem are provided in Sangalli, Ramsay, and Ramsay (2013).

The ARLS algorithm can utilize the same procedures that have been used for generating initial values for $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ as in conventional PARAFAC, for instance a direct trilinear decomposition, singular vector decomposition, and orthogonalization of random values (e.g., Bro, 1997; Sands & Young, 1980). Alternatively, we may run a small number of iterations of the PARAFAC algorithm and employ the resultant solutions as starting values for $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. Although the ARLS algorithm is monotonically convergent, it will be useful to repeat the algorithm with a large number of different sets of initial values for the parameters (e.g., 100 sets) in order to safeguard against convergence to a local minimum.

The values of λ and ρ may be chosen automatically by applying K -fold cross-validation (Hastie et al., 2009). In K -fold cross-validation, we divide the entire dataset into K subsamples. We use one subsample as a test set, while using all of the remaining $K - 1$ subsamples as a training set. The training set is used for parameter estimation, and the resultant estimates are applied to the test set in order to compute prediction error. This process is repeated K times by using each subsample consecutively and only once as a test set. We then calculate the cross-validation estimate of prediction error by averaging prediction errors over all K test sets. The values of λ and ρ associated with the smallest cross-validation estimate of prediction error can be selected as the final ones.

Owing to the inclusion of the smoothing parameters, the proposed method is rather insensitive to the number of basis functions adopted. The values of λ and ρ can be adjusted automatically depending on how many basis functions are used. For example, when a large number of basis

functions are used, a smaller value of the smoothing parameter tends to be selected via K -fold cross-validation, whereas when a small number of basis functions is used, a larger value of the smoothing parameter tends to be selected.

The number of components (R) needs to be specified prior to applying the ARLS algorithm. We may determine the number of R by employing two approaches. One is by looking at the percentage of variance unexplained against different values of components. We may choose R such that adding another component doesn't decrease much of the percentage of the unexplained variance. Another measure is called core consistency diagnostic (CORCONDIA), introduced by Bro and Kiers (2003). The CORCONDIA, expressed in percentage, bespeaks the degree of trilinearity of the data. A value greater than 90% indicates that the PARAFAC model with a given R fits to the data very well, whereas a value smaller than 85% indicates that the PARAFAC model is inappropriate with too many components being extracted.

3. A Simulation Study

We conducted a simulation study to evaluate how well the proposed method recovered the parameter values of the component matrices. In this study, the structure of the three-way data was prescribed as 10 subjects by 150 time points by 100 spatial locations. Data were generated given the PARAFAC model in (4), i.e., $\mathbf{Z}_A = \mathbf{A}(\mathbf{C} \otimes \mathbf{B})' + \mathbf{E}$, where $\mathbf{E}$ was drawn from a normal distribution with mean zero and standard deviation σ . The number of components was fixed at two ($R = 2$), whereas the standard deviation of $\mathbf{E}$ was to vary at three different levels ($\sigma = .5, 1$, and 3). An N by R matrix of components for the first mode ($\mathbf{A}$) was generated from a normal distribution with mean of $.5$ and standard deviation of 1.5 . A J by R matrix of components for the second mode ($\mathbf{B}$) was generated by first specifying two underlying one-dimensional smooth functions $\mathbf{b}^{(1)}(t)$ and $\mathbf{b}^{(2)}(t)$, defined over the interval $[T_1, T_2]$, as follows:

$$\begin{aligned} \mathbf{b}^{(1)}(t) &= \cos\left(\frac{2\pi}{40}\right)(t+1) + 4 \sin\left(\frac{2\pi}{70}\right)(t+4) \\ \mathbf{b}^{(2)}(t) &= 10 \cos\left(\frac{2\pi}{50}\right)(t+2) + 5 \sin\left(\frac{2\pi}{80}\right)(t+5), \end{aligned} \quad (13)$$

where the interval is $[0, 150]$ and the superscript indicates the r th component ($r = 1, 2$). Then, $\mathbf{B}$ was constructed by evaluating the one-dimensional smooth functions at 150 time instants. An S by R matrix of components for the third mode ($\mathbf{C}$) was generated in a similar manner, except that $\mathbf{C}$ was assumed to arise from two-dimensional smooth functions spanned on a two-dimensional domain Ω . For this, the boundary of the spatial domain on Ω was set as the same as the shape introduced in Ramsay (2002). Note that this spatial domain has been widely adopted in spatial data analysis because it is considered more general than a simple rectangular domain, imposing irregular shapes over the domain (e.g., Ramsay et al., 2011; Sangalli et al., 2013; Scott-Hayward, MacKenzie, Donovan, Walker, & Ashe, 2014; Wood, Bravington, & Hedley, 2008). We prescribed two functions at $R = 2$ varying along Ω , as graphically illustrated in Figure 1. Figure 1a, b displays the first and second component's underlying two-dimensional smooth functions, respectively. Finally, the values of $\mathbf{C}$ were determined by evaluating the values at 100 different spatial locations, indicated by filled triangle in Figure 2.

We generated 500 simulated datasets at each level of σ , thus yielding a total of 1500 datasets. At each level, the smoothing parameters were determined based on fivefold cross-validation ($K = 5$), considering 17 alternative values changing from 10^{-12} to 10^4 by a factor of 10. To lessen the computational burden, the cross-validation procedure was applied to the first simulated

FIGURE 1.

The prescribed underlying two-dimensional smooth functions on the two-dimensional domain for $R = 2$. **a**, **b** The first and second components, respectively.

dataset only to choose the values of the smoothing parameters, and then, the same values were used for the remaining datasets.

We used a B-spline basis system to represent the extracted components for the second mode (**B**) as functions of time. To this end, two elements should be predetermined: the number of knots that divide the entire interval of a function into subintervals, and the order of a polynomial for defining a spline function at any subinterval. We created order four B-spline basis functions with 50 equally spaced knots over the interval $[0, 150]$. To approximate the extracted components for the third mode (**C**) as functions of space, 1152 disjoint triangles were created on a two-dimensional domain Ω . Constructing basis functions was carried out using Dr. James Ramsay's MATLAB codes which are available at <https://www.psych.mcgill.ca/misc/fda>.

For comparison, the same simulated datasets were also analyzed by the conventional PARAFAC, assuming that all responses of the data were multivariate. To evaluate the parameter recovery capability, we examined the average relative biases and mean square errors (MSE) of the estimates of the trilinear components for evaluating parameter recovery of the proposed method. We computed the relative bias of an estimate as $\text{relative bias} = \frac{\mathbb{E}[\omega_{\text{est}}] - \theta}{\theta} \times 100$, and the

FIGURE 2.

The prescribed boundary of $p_s = (x_s, y_s)$ ($s = 1, \dots, S$) on a spatial two-dimensional domain. *Filled triangle* indicate 100 observed spatial points.

MSE as $\text{MSE} = \mathbb{E}[(\omega_{\text{est}} - \theta)^2]$, where θ and ω_{est} denote a parameter and its estimate, respectively. The two measures were applied to all of N , J , and S elements in the trilinear components. Then, we computed the average of the resultant values over the number of entities in its corresponding mode to obtain the average relative biases and MSE.

Tables 1 and 2 provide the average relative biases and MSE values of the proposed method and conventional PARAFAC at the three different levels of σ , respectively. As shown in the table, for the proposed method, the average relative biases of the estimates of **A**, **B**, and **C** obtained from the proposed method were smaller than 10% in absolute value, which may be regarded as an acceptable level of relative bias (e.g., Bollen et al., 2007; Lei, 2009). Moreover, the average MSE values of the proposed method were small and quite close to zero across all modes in different conditions. At each mode, although there were slight increases in the relative biases and MSE as error variance increased, the differences appeared rather negligible. It suggests that the proposed method accurately recovered the true parameters even with the degree of error variances increased.

TABLE 1.
The average relative bias (Rbias) and mean square error (MSE) of obtained from the proposed method.

σ	Mode	Rbias (%)	MSE
.5	A	5.1642	0.0306
	B	2.2416	0.0101
	C	4.1803	0.0459
1	A	5.9164	0.0342
	B	2.7461	0.0254
	C	3.3231	0.0497
3	A	5.7284	0.0326
	B	2.8074	0.0298
	C	4.7679	0.0514

TABLE 2.
The average relative bias (Rbias) and mean square error (MSE) obtained from the conventional PARAFAC.

σ	Mode	Rbias (%)	MSE
.5	A	36.6586	0.1318
	B	60.2883	0.3972
	C	49.3831	0.1465
1	A	42.7088	0.1676
	B	62.2214	0.4450
	C	51.7594	0.2367
3	A	51.7594	0.2421
	B	76.3210	0.5562
	C	57.7594	0.4124

On the other hand, the average relative biases and MSE from the conventional PARAFAC approach were much larger than those from the proposed method across all conditions. This indicates that the proposed method tended to yield estimates that were closer to the true parameter values than the conventional PARAFAC, in the presence of three-way data whose responses were considered smooth functions.

With the proposed method, we would graphically visualize the extent to which the estimates of the r th smooth functions differ from the true functions over their corresponding domain. It enables to inspect which time instants over $[T_1, T_2]$ or area on spatial domain Ω may show greater discrepancy in comparison to the true functions. For this, for each component in the second mode, at each level of σ , we plotted the true smooth function over the entire time interval $[T_1, T_2]$ as well as the estimates of each smooth function (i.e., $\hat{\mathbf{b}}_r$) averaged over the 500 simulated datasets. Figure 3 displays the true smooth function and the averaged estimates of $\hat{\mathbf{b}}_r$ at $\sigma = 3$. Although the discrepancy seemed to be somewhat greater in time instants with curves, overall, the proposed method recovered the true functions sufficiently well. Note that the averaged estimates of the r th smooth function at the remaining two levels of σ were not presented here because a similar pattern or even smaller degree of discrepancy was observed. For each component in the third mode, we computed and plotted the MSE of the estimates of $\hat{\mathbf{c}}_r$ over the spatial domain Ω at $\sigma = 3$ in Figure 4. As shown in the figures, their MSE values remained small over the entire spatial domain. This was the case for the other levels of σ as well.

4. An Empirical Application

The present example is a subset of the electroencephalography (EEG) data collected from an object recognition task (Zhang, Begleiter, Porjesz, Wang, & Litke, 1995) (available at <https://archive.ics.uci.edu/ml/datasets/EEG+Database>). PARAFAC has been recognized as a practically useful approach to extracting components from EEG data in that it resolves the inherent drawback of other matrix decomposition techniques, which allow for a two-dimensional matrix structure only (e.g., Cichocki, Zdunek, Phan, & Amari, 2009; Miwakeichi et al., 2004; Mørup, 2005; Weis, Romer, Haardt, Jannek, & Husar, 2009). In fact, EEG data are often recorded constituting more than two different types of entities simultaneously, and thus, they are naturally well fitted by decomposition techniques for a three-way data array (e.g., Cong et al., 2015; Nazarpour, Sanei, Shoker, & Chambers, 2006; Mizuhara, Yonehara, & Gofuku, 2007). In this regard, we have chosen EEG data to demonstrate the applicability of the proposed method. In the example, each subject was fitted with an electrode cap and the amplified signals were recoded from 64

FIGURE 3.

The averaged estimates of each component's smooth function, plotted over the time interval $[T_1, T_2]$, at $\sigma = 3$. **a, b** The first and second components, respectively. *Solid lines in blue* refer to the prescribed true smooth functions, while *dashed red lines* indicate the estimates of each component's smooth function averaged across 500 simulated datasets (Color figure online).

scalp electrodes. Figure 5 displays a two-dimensional topographical map of the electrodes with their labels from “Guidelines for Standard Electrode Position Nomenclature” by the American Electroencephalographic Society (Sharbrough et al., 1991). To measure voltage values from each electrode, a reference should be selected in advance, and in this example the reference electrode was placed to Cz.

In the example, a delayed match-to-sample task was employed. The stimuli of the task were 90 pictures of familiar objects. One of the 90 pictures was presented to each subject as the first stimulus (S1) with its verbal label, and he/she was asked to encode S1. After a delay of 300ms (milliseconds), another picture was presented as the second stimuli (S2) to indicate whether S2 corresponds to S1 in terms of semantic category. The subject was asked to press a mouse key in

FIGURE 4.

The mean square error (MSE) of the estimates of each component's smooth function computed over the spatial domain Ω at $\sigma = 3$. **a, b** The first and second components, respectively.

one hand if S2 was the same as S1 (i.e., matching condition) and to press a mouse key in the other hand if S2 differed from S1 (i.e., non-matching condition). The completion of the recognition task was considered a single trial. At each trial, an epoch of 1000 ms following the second stimulus presentation was recorded with the sampling rate of 256 Hz. Because hertz (Hz) is equivalent to cycles per second, it implies the total number of time points per trial is 256. The trials were presented with an inter-trial interval of 3.2 s.

For our analysis, we concentrated on a single subject's EEG recordings obtained under 10 trials of the recognition tasks (all being classified into matching condition). The amplified signals were averaged over the 10 trials prior to the analysis. The aim of our analysis was to investigate which frequency bands (i.e., wave activity) were likely to be interrelated with temporal profiles and brain regions. More specifically, the proposed method was employed to extract a number of components that would summarize the EEG recordings in terms of frequency, time, and space, while accounting for temporal and spatial variations in the data.

To formulate a set of three-way data from a subject's EEG recordings, the spectral power of the EEG recordings was quantified by using a complex wavelet transformation. The spectral power $\mathbf{z}(s, f, t)$ of the s th electrode ($s = 1, \dots, 64$) at the f th frequency ($f = 2, \dots, 40$ with

FIGURE 5.

A two-dimensional topographical map of the electrodes with their labels.

an increment of two) and the t th time point ($t = 1, \dots, 256$) was given by the squared norm of the convolution of a Morlet wavelet with the EEG amplified signal $v(s, t)$, as follows:

$$\mathbf{z}(s, f, t) = |w(t, f) * v(d, f)|^2, \quad (14)$$

where the complex Morlet wavelet $w(t, f)$ is defined by $w(t, f) = 2\sqrt{\pi}\sigma_b \exp(-(t/\sigma_b)^2) \exp(i2\pi tf)$ with σ_b being the bandwidth parameter. In the example, the width of the wavelet was set to 7 and this specification would compute the bandwidth parameter (following $7 = 2\pi\sigma_b f$) (Miwakeichi et al., 2004; Mørup, Hansen, Herrmann, Parnas, & Arnfre, 2006). This wavelet transformation yielded a three-way data of 20 frequencies by 256 time points by 64 spatial locations. Since the analysis includes two different types of entities (i.e., time points and electrodes in the scalp) that are better regarded as samples of functions (Tian, 2010; Valdes-Sosa, 2004), the proposed method seems preferable over conventional PARAFAC.

We used a B-spline basis system to represent the extracted components for the second mode (**B**) as functions of time and created order four B-spline basis functions with 70 equally spaced knots over the interval $[0, 256]$. To approximate the extracted components for the third mode (**C**) as functions of space, 362 disjoint triangles were constructed on the boundary of two-dimensional topographical map of the scalp, as provided in Figure 6. As the s th point moves toward the boundary, much denser triangles are generated, suggesting that this would be better suited for indexing electrodes that are close to the boundaries. The disjoint triangles were used for constructing piecewise quadratic functions via FEM.

We should determine the number of components (R) and the values of the smoothing parameters (λ and ρ). We applied fivefold cross-validation by varying each of λ and ρ from 10^{-12} to 10^4 by a factor of 10, at each of different values of R that changed from 1 to 6. For each R , the final value was chosen at which the cross-validation estimate of prediction error was minimized. For example, when $R = 2$, the smoothing parameters were determined as $\lambda = 10^{-6}$ and $\rho = 10^0$, and when $R = 3$, they were determined as $\lambda = 10^{-8}$ and $\rho = 10^0$. The computation for the

FIGURE 6.
Creating disjoint triangles on a two-dimensional topographical map of the scalp.

TABLE 3.
Percentage of unexplained variance and CORCONDIA values against different values of components.

R	% of unexplained variance	CORCONDIA
1	16.2813	100.00
2	8.6374	99.71
3	7.2041	96.62
4	6.4341	85.25
5	5.7905	48.51
6	5.6349	4.56

entire cross-validation procedures took 2.21 h on a desktop computer with 3 GHz Intel CPU and 4 GB memory. Given the values of λ and ρ for different values of R , we examined the changes in percentage of unexplained variance and CORCONDIA values. As shown in Table 3, we chose $R = 3$ because no substantial change occurred in the unexplained variance percentage beyond $R = 3$, and its CORCONDIA value (greater than 90%) suggested that the PARAFAC model would be appropriate for summarizing the data.

The components for the first mode (**A**) describe frequency characteristics of the EEG data. The first identified component consisted of waves that were particularly amplified at 2, 4, 6, and 8 Hertz (Hz). The frequency bands that entail such Hz are known as Delta (<4 Hz) and Theta bands (4–8 Hz). The second component was highly amplified at 12, 14, 16, 18, 20, 22, and 24 Hz, whose rhythms are classified as Alpha (8–15 Hz) and Beta bands (16–30 Hz). The third component was predominant at 34, 36, 38, and 40 Hz, belonging to the range of Gamma-band (>32 Hz) oscillations. These frequency bands have been known as the common types of rhythms that become salient when involving memory tasks in human (e.g., Altamura, 2010; Doppelmayr, Klimesch, Schwaiger, Stadler, & Roehm, 2000; Klimesch, 1999).

Figures 7 and 8 exhibit the components for the second and third modes, respectively. Figure 7 displays the extracted components for the second mode (**B**). It shows how the temporal profile

FIGURE 7.

The estimated temporal components under $R = 3$. **a-c** The first, second, and third components, respectively.

has evolved with its association with the corresponding frequency-band information. The first component was best related to the Delta- and Theta-band oscillations, encompassing a temporal profile with increased activities around the 100th time point. It indicates that these band oscillations had occurred at a quite early stage (i.e., 390ms after the stimulus presentation). In contrast, the second component related to the Alpha- and Beta-band oscillations was dominant across most of the range of 256 time points. For the third component related to the Gamma-band oscillations, the peak latency was observed at relatively later time points corresponding to 820ms out of an epoch of 1000ms. This suggests that the Gamma-band oscillations tended to boost their activities at a bit later time points than the other two types of frequencies.

Lastly, Figure 8 displays the extracted components for the third mode (C). It describes topographic distributions of EEG power on the electrodes. The first component related to the Delta- and Theta-band oscillations was most likely localized at electrodes placed near frontotemporal regions. The topographic distributions of the second component that were related to the Alpha- and Beta-band oscillations were maximal over PO_Z, PO₃, PO₄ and electrodes near frontal regions (i.e., AF₇, AF₃, AF_Z, AF₄, AF₈, FP₁, FP_Z, and FP₂). The spatial locations of the third component related to the Gamma-band oscillations were similar to those of the first component, showing increased activations over AF₈, AF₄, F₄, F₆, F₈, FP₂, FC₅, and P₃ electrodes. These regions identified by our analysis were consistent with the previous visual memory researches (e.g., Braver et al., 1997; Krause et al., 2000; Sauseng, Klimesch, Schabus, & Doppelmayr, 2005; Zhang et al., 1995).

5. Concluding Remarks

We generalized PARAFAC to the framework of functional data analysis, especially with functions of more than a single argument. This method extracts trilinear components whose elements can be treated as a function of either a one- or two-dimensional continuum. The empirical usefulness of the method was demonstrated through the analysis of both simulated and real data. The simulation study showed that the method recovered population trilinear components sufficiently well. The proposed method was also applied to the analysis of real EEG data. It was able to describe frequency–time–space characteristics of the brain’s spontaneous electrical activity measured from EEG recordings.

We may further extend the proposed method to improve the scope of its applicability. First, the proposed method can be combined with other decomposition techniques such as Tucker3 models (Tucker, 1966; Kroonenberg & De Leeuw, 1980). The Tucker3 models attempt to obtain not only a low-dimensional representation of each mode, but also a reduced three-dimensional block, called a core array, which relates the extracted components of all modes to one another. The extension to the Tucker analysis may be of particular use when there exist some cross-interactions among the extracted components of different modes (Kroonenberg, Harshman, & Murakami, 2009). To formulate the core array in the Tucker3 models as function(s), we may consider incorporating the smoothing method introduced by Bernardi, Sangalli, Mazza, and Ramsay (2016), which estimates so-called spatiotemporal smooth functions while accounting for the interdependencies between space and time.

Second, the proposed method currently considers each smoothing parameter (λ and ρ) scalar, assuming that the roughness or curvature of the estimated functions remains constant across the entire interval of the functions. However, a smooth function may have different degrees of curvature over its interval, rendering the function to vary locally in shape. Hence, the proposed method may perform poorly when smooth functions are expected to have rapid fluctuations in some parts and little variability in other parts. To address this potential issue, we may incorporate Bayesian inferences via Monte Carlo Markov chain (MCMC) techniques, wherein a smoothing parameter

is allowed to vary across the interval of the estimated functions to adjust their smoothness locally (e.g., Ruppert & Carroll, 2000; Wood, Jiang, & Tanner, 2002).

Third, the proposed method focuses thus far on functions of a one- and two-dimensional type of continuum. This may be extended to a three-dimensional continuum (i.e., volume), on which an ordered triplet of axes (i.e., a three-dimensional Cartesian coordinate system) is used for representing three-dimensional functions. Although in principle, it will be straightforward to extend the proposed method to handle functions of a high-dimensional continuum, some computationally efficient approach for constructing basis functions with tetrahedrons in a three-dimensional domain needs to be developed. To generate such tetrahedrons efficiently, it may be worthwhile to consider integrating the proposed method to a three-dimensional hidden Markov model (HMM) to infer underlying distribution of responses in the three-dimensional coordinate system (e.g., Li, Joshi, & Wang, 2004).

Lastly, when entities in a mode are comprised of subjects, the proposed method assumes that all subjects are collected from a single population. However, this may not be suitable when subjects are compounded from a number of heterogeneous subgroups that exhibit distinctive response patterns. Examples of so-called cluster-level heterogeneity abound in various areas including developmental psychology (e.g., Dauvier, Chevalier, & Blaye, 2012) and marketing (e.g., Kamakura & Russell, 1989). To take into account such cluster-level heterogeneity, we may generalize the proposed method into the framework of clusterwise PARAFAC (Wilderjans & Ceulemans, 2013), so that a different set of components can be extracted from each cluster of subjects, while treating entities in the remaining modes as smooth functions.

References

- Allen, G. (2013). Multi-way functional principal components analysis. In *Computational advances in multi-sensor adaptive processing (CAMSAP), 2013 IEEE 5th International Workshop on*. IEEE.
- Altamura, M., Goldberg, T. E., Elvevåg, B., Holroyd, T., Carver, F. W., Weinberger, D. R., et al. (2010). Prefrontal cortex modulation during anticipation of working memory demands as revealed by magnetoencephalography. *Journal of Biomedical Imaging*. doi:10.1155/2010/840416.
- Bernardi, M. S., Sangalli, L. M., Mazza, G., & Ramsay, J. O. (2016). A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. *Stochastic Environmental Research and Risk Assessment*. doi:10.1007/s00477-016-1237-3.
- Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent variable models under misspecification: Two-stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods & Research*, 36(1), 48–86.
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage*, 5(1), 49–62.
- Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2), 149–171.
- Bro, R., & Kiers, H. A. L. (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, 17(5), 274–286.
- 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(3), 283–319.
- Cichocki, A., Zdunek, R., Phan, A. H., & Amari, S. (2009). *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation*. New York: Wiley.
- Cong, F., Lin, Q. H., Kuang, L. D., Gong, X. F., Astikainen, P., & Ristaniemi, T. (2015). Tensor decomposition of EEG signals: A brief review. *Journal of Neuroscience Methods*, 248, 59–69.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 162–173.
- Dauvier, B., Chevalier, N., & Blaye, A. (2012). Using finite mixture of GLMs to explore variability in children's flexibility in a task-switching paradigm. *Cognitive Development*, 27(4), 440–454.
- Doppelmayr, M., Klimesch, W., Schwaiger, J., Stadler, W., & Roehm, D. (2000). The time locked theta response reflects interindividual differences in human memory performance. *Neuroscience Letters*, 278(3), 141–144.
- Folland, G. B. (1995). *Introduction to partial differential equations*. Princeton, NJ: Princeton University Press.
- Germond, L., Dojat, M., Taylor, C., & Garbay, C. (2000). A cooperative framework for segmentation of MRI brain scans. *Artificial Intelligence in Medicine*, 20(1), 77–93.
- Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis. In *UCLA Working Papers in Phonetics* (Vol. 16, pp. 1–84).
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer.

- Holter, N. S., Maritan, A., Cieplak, M., Fedoroff, N. V., & Banavar, J. R. (2001). Dynamic modeling of gene expression data. *Proceedings of the National Academy of Sciences*, 98(4), 1693–1698.
- Hwang, H. (2009). Regularized generalized structured component analysis. *Psychometrika*, 74(3), 517–530.
- Kamakura, W. A., & Russell, G. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26, 379–390.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29(2), 169–195.
- Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3), 455–500.
- Krause, C. M., Sillanmäki, L., Koivisto, M., Häggqvist, A., Saarela, C., Revonsuo, A., et al. (2000). Effects of electromagnetic field emitted by cellular phones on the EEG during a memory task. *Neuroreport*, 11(4), 761–764.
- Kroonenberg, P. M. (1987). Multivariate and longitudinal data on growing children. Solutions using a three-mode principal component analysis and some comparison results with other approaches. In J. Janssen, F. Marcotorchino, & J. M. Proth (Eds.), *Data analysis. The ins and outs of solving real problems* (pp. 89–112). New York: Plenum.
- Kroonenberg, P. M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1), 69–97.
- Kroonenberg, P. M., Harshman, R. A., & Murakami, T. (2009). Analysing three-way profile data using the Parafac and Tucker3 models illustrated with views on parenting. *Applied Multivariate Research*, 13(1), 5–41.
- Kruskal, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Its Applications*, 18(2), 95–138.
- Kuze, T., Goto, M., Ninomiya, K., Asano, K., Miyazawa, S., Munekata, H., et al. (1985). A longitudinal study on development of adolescents' social attitudes. *Japanese Psychological Research*, 27(4), 195–205.
- Lei, P. W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity*, 43(3), 495–507.
- Li, J., Joshi, D., & Wang, J. Z. (2004). Stochastic modeling of volume images with a 3-D hidden Markov model. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*. IEEE.
- Miwakeichi, F., Martinez-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(3), 1035–1045.
- Mizuhara, H., Yonehara, E., & Gofuku, A. (2007). Parallel factor analysis can decompose human EEG during a finger movement task. In *SICE, 2007 Annual Conference*. IEEE.
- Mørup, M. (2005). *Analysis of brain data—Using multi-way array models on the EEG*. M.Sc. Thesis. Technical University of Denmark.
- Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., & Arnfred, S. M. (2006). Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, 29(3), 938–947.
- Nazarpour, K., Sanei, S., Shoker, L., & Chambers, J. A. (2006). Parallel space-time-frequency decomposition of EEG signals for brain computer interfacing. In *Signal processing conference, 2006 14th European*. IEEE.
- Oort, F. J. (2001). Three-mode models for multivariate longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 54(1), 49–78.
- Pepper, D. W., & Heinrich, J. C. (2006). *The finite element method: Basic concepts and applications*. New York: Taylor & Francis.
- Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*. New York: Springer.
- Ramsay, J. O., Ramsay, T. O., & Sangalli, L. M. (2011). Spatial functional data analysis. *Recent advances in functional data analysis and related topics* (pp. 269–275). Berlin: Springer.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.
- Ramsay, T. O. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 307–319.
- Reis, M. M., & Ferreira, M. (2002). PARAFAC with splines: A case study. *Journal of Chemometrics*, 16(8–10), 444–450.
- Rosenberg, S. (1997). *The Laplacian on a Riemannian manifold: An introduction to analysis on manifolds*. Cambridge: Cambridge University Press.
- Ruppert, D., & Carroll, R. J. (2000). Spatially-adaptive penalties for spline fitting. *Australian & New Zealand Journal of Statistics*, 42(2), 205–223.
- Sands, R., & Young, F. W. (1980). Component models for three-way data: An alternating least squares algorithm with optimal scaling features. *Psychometrika*, 45(1), 39–67.
- Sangalli, L. M., Ramsay, J. O., & Ramsay, T. O. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4), 681–703.
- Sauseng, P., Klimesch, W., Schabus, M., & Doppelmayr, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International Journal of Psychophysiology*, 57(2), 97–103.
- Scott-Hayward, L. A. S., MacKenzie, M. L., Donovan, C. R., Walker, C. G., & Ashe, E. (2014). Complex region spatial smoother (CRSS). *Journal of Computational and Graphical Statistics*, 23(2), 340–360.
- Sharbrough, F., Chatrian, G. E., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8(2), 200–202.
- Stegeman, A., & Sidiropoulos, N. D. (2007). On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra and Its Applications*, 420(2), 540–552.
- Thirion, B., & Fugeras, O. (2003). Dynamical components analysis of fMRI data through kernel PCA. *NeuroImage*, 20(1), 34–49.

- Tian, T. S. (2010). Functional data analysis in brain imaging studies. *Frontiers in Psychology*, 1, 1–11.
- Timmerman, M. E., & Kiers, H. A. L. (2002). Three-way component analysis with smoothness constraints. *Computational Statistics & Data Analysis*, 40(3), 447–470.
- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3), 279–311.
- Valdes-Sosa, P. A. (2004). Spatio-temporal autoregressive models defined over brain manifolds. *Neuroinformatics*, 2(2), 239–250.
- Weis, M., Romer, F., Haardt, M., Jannek, D., & Husar, P. (2009). *Multi-dimensional space-time-frequency component analysis of event related EEG data using closed-form PARAFAC*. Paper presented at the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing.
- Wilderjans, T. F., & Ceulemans, E. (2013). Clusterwise Parafac to identify heterogeneity in three-way data. *Chemometrics and Intelligent Laboratory Systems*, 129, 87–97.
- Wood, S. N., Bravington, M. V., & Hedley, S. L. (2008). Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 931–955.
- Wood, S. A., Jiang, W., & Tanner, M. (2002). Bayesian mixture of splines for spatially adaptive nonparametric regression. *Biometrika*, 89(3), 513–528.
- Xiao, G., Wang, X., & Khodursky, A. B. (2011). Modeling three-dimensional chromosome structures using gene expression data. *Journal of the American Statistical Association*, 106(493), 61–72.
- Zhang, X. L., Begleiter, H., Porjesz, B., Wang, W., & Litke, A. (1995). Event related potentials during object recognition tasks. *Brain Research Bulletin*, 38(6), 531–538.
- Zienkiewicz, O. C. (1971). *The finite element method in engineering science*. London: McGraw-Hill.
- Zienkiewicz, O. C., & Taylor, R. L. (2000). *The finite element method*. Oxford: Butterworth-Heinemann.

Manuscript Received: 4 NOV 2015

Final Version Received: 8 NOV 2016