Extracting host-specific developmental signatures from longitudinal microbiome data

Balázs Erdős[1,*], Christos Chatzis[1,2], Jonathan Thorsen[3,4], Jakob Stokholm[3,5], Age K. Smilde[6], Morten A. Rasmussen[3,5], and Evrim Acar[1,*]

[1]Department of Data Science and Knowledge Discovery, Simula Metropolitan Center for Digital Engineering, Oslo, Norway
[2]Faculty of Technology, Art and Design, Oslo Metropolitan University, Oslo, Norway
[3]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Copenhagen University Hospital - Herlev and Gentofte, Copenhagen, Denmark
[4]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[5]Department of Food Science, University of Copenhagen, Copenhagen, Denmark
[6]Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, the Netherlands
[*]Corresponding author: erdos.blz@gmail.com (B.E.), evrim@simula.no (E.A.)

Abstract

Motivation: Longitudinal microbiome studies provide critical insights into microbial community dynamics and their relation to host health. Tensor decompositions offer a powerful framework for the unsupervised analysis of such data, yielding interpretable low-dimensional temporal patterns. However, existing approaches based on the CANDECOMP/PARAFAC (CP) model assume common temporal dynamics for all subjects and therefore cannot capture subject-specific trajectories. **Results:** To address this limitation, we propose to use the PARAFAC2 model, a more flexible tensor decomposition method to explicitly model subject-specific variations, such as shifts and delays in temporal patterns. Through systematic comparisons on simulated and real-world datasets—including studies of infant gut maturation and dietary interventions—we demonstrate that PARAFAC2 outperforms CP in capturing subject-specific temporal trajectories, and enables the discovery of biologically relevant patterns that are overlooked by CP. Furthermore, we introduce replicability as a robust criterion for selecting the number of components and demonstrate that the extracted patterns are replicable. **Availability and implementation:** Code and data for reproducing the results of this paper are available at `https://github.com/blzserdos/parafac2_microbiome` and archived at `https://doi.org/10.5281/zenodo.17674126`.

Introduction

The human microbiome exhibits complex temporal dynamics that vary substantially both within and across individuals, making longitudinal analysis crucial for understanding its evolution through developmental stages (e.g., through childhood) or following interventions, as well as its link to host health. Advances in high-throughput sequencing now allow for dense temporal sampling, offering new opportunities to link microbial dynamics with host phenotypes. However, analyzing such data poses distinct computational challenges that call for specialized methods (Kodikara et al., 2022). Traditional dimensionality reduction methods, such as principal coordinates analysis (PCoA) (Goodrich et al., 2014), operate at the sample level and fail to account for the inherent temporal structure and within-subject correlations present in longitudinal data.

To address these limitations, tensor decomposition-based approaches have recently emerged for the analysis of longitudinal microbiome data. Tensor decompositions, which are extensions of matrix factorizations to multi-way arrays (i.e., higher-order tensors), factorize multi-way data into a set of interpretable components, with successful applications across diverse fields such as chemometrics,

neuroscience, and computational phenotyping (Kolda and Bader, 2009; Hunyadi et al., 2017; Becker et al., 2023). In microbiome research, existing tools such as compositional tensor factorization (CTF) (Martino et al., 2021), parafac4microbiome (van der Ploeg et al., 2025), microTensor (Ma and Li, 2023), and temporal tensor decomposition (TEMPTED) (Shi et al., 2024) have primarily relied on the CANDECOMP/PARAFAC (CP) tensor decomposition (Harshman, 1970; Carroll and Chang, 1970), which represents the data as a sum of components, each defined by taxa loadings (a microbial signature), subject loadings (its expression across individuals), and time loadings (its temporal progression).

However, a critical limitation of these CP-based methods is the assumption of shared time loadings across all subjects. This assumption restricts these methods to modeling only population-level dynamics and prevents them from capturing subject-specific temporal variation that is commonly observed in microbiome studies (Zhou et al., 2024). This limitation is particularly relevant when individual temporal variation may reflect sources of important biological information. For instance, in studies of infant gut microbiome maturation, subjects may follow the same underlying biological maturation process, but at a different pace, representing the underlying biological heterogeneity or environmental exposures (Vatanen et al., 2019; Hoskinson et al., 2023). Similarly, in intervention studies involving antibiotics or dietary changes, subjects often exhibit divergent responses in the timing and magnitude of microbial adaptation (even within the same treatment group), underscoring the need for models that can capture these personalized dynamics (Dethlefsen and Relman, 2011; Tanes et al., 2021).

Unlike the CP model, PARAFAC2 (Harshman, 1972) offers a more flexible alternative by relaxing the strict assumption of multilinearity when analyzing multi-way data by letting the factors in one mode change across tensor slices. This allows it to model shape changes, such as the shifts and delays that characterize individualized temporal trajectories. The power of this flexible approach has been well established in other domains. In chemometrics, for instance, PARAFAC2 is used to resolve and quantify chemical compounds in mixtures where chromatographic profiles exhibit retention time shifts between runs (Amigo et al., 2008). In neuroscience, it successfully identified changes in brain connectivity consistent with prior pathological knowledge of Alzheimer's disease (Spyrou et al., 2019), and also captured subject-specific temporal profiles from the analysis of multi-subject functional neuroimaging data (Mørup et al., 2025). It has been leveraged to construct personalized disease trajectories from electronic health records, capturing each patient's phenotypic evolution across their clinical encounters (Perros et al., 2019). More recently, in single-cell genomics, PARAFAC2 has provided a framework for integrating data from perturbational experiments, capturing how cellular responses vary across different conditions and treatments (Ramirez et al., 2025).

In this work, we adopt the PARAFAC2 decomposition for the analysis of longitudinal microbiome data, allowing us to model subject-specific temporal trajectories. We evaluate and compare CP and PARAFAC2 on both simulated and real-world longitudinal microbiome datasets, including studies on infant gut maturation and dietary interventions. We show that PARAFAC2 more effectively captures subject-specific temporal dynamics, which in turn enables the discovery of biologically relevant patterns that are overlooked by the more restrictive CP model. Finally, we illustrate and discuss the use of replicability analyses in model selection to improve the robustness of the extracted patterns.

Materials and methods

Tensor decompositions

CANDECOMP/PARAFAC (CP)

The CP model (Harshman, 1970; Carroll and Chang, 1970; Hitchcock, 1927) approximates a higher-order tensor as the sum of rank-one tensors. An R-component CP model of a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is defined as $\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$, where $\circ$ denotes the vector outer product, $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ correspond to the r-th column of factor matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, and $\mathbf{C} \in \mathbb{R}^{K \times R}$, respectively. In matrix notation, using the CP model, the kth frontal slice of $\mathcal{X}$, denoted $\mathbf{X}_k = \mathcal{X}_{:,:,k}$, is approximated as:

$$\mathbf{X}_k \approx \mathbf{A}\mathrm{diag}(\mathbf{c}_{k,:})\mathbf{B}^\top, \tag{1}$$

where $\mathrm{diag}(\mathbf{c}_{k,:})$ is an $R \times R$ diagonal matrix with the kth row of $\mathbf{C}$ on the diagonal, and $^\top$ denotes the matrix transpose. The CP model is unique up to permutation and scaling ambiguities under mild conditions (Kruskal, 1977; Kolda and Bader, 2009), ensuring reliable interpretation of the resulting factors. More specifically, columns of the factor matrices can be permuted in the same way across modes (permutation ambiguity), and can be scaled as long as the component norm stays the same, e.g.,

$\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = \gamma_1 \mathbf{a}_r \circ \gamma_2 \mathbf{b}_r \circ \frac{1}{\gamma_1 \gamma_2} \mathbf{c}_r$ (scaling ambiguity). Under these ambiguities, the interpretation of the CP factors remains unchanged. If $\mathcal{X}$ represents a *taxa* by *time* by *subjects* tensor, each column of $\mathbf{A}$ (taxa loadings), i.e., $\mathbf{a}_r$, may reveal a microbial signature with the corresponding temporal profile captured in the columns of $\mathbf{B}$ (time loadings), i.e., $\mathbf{b}_r$. Finally, columns of $\mathbf{C}$ (subject loadings), i.e., $\mathbf{c}_r$, indicate the extent to which the compositional changes over time are present in each subject (Figure 1).

PARAFAC2

Unlike the CP model, the PARAFAC2 model (Harshman, 1972) allows for the patterns in $\mathbf{B}_k$ to vary across the third mode (Figure 1):

$$\mathbf{X}_k \approx \mathbf{A}\text{diag}(\mathbf{c}_{k,:})\mathbf{B}_k^\top, \quad \{\mathbf{B}_k\}_{k=1}^K \in \mathcal{P}, \tag{2}$$

where $\mathcal{P} = \left\{ \{\mathbf{B}_k\}_{k=1}^K \mid \mathbf{B}_{k_1}^\top \mathbf{B}_{k_1} = \mathbf{B}_{k_2}^\top \mathbf{B}_{k_2}, \forall k_1, k_2 \leq K \right\}$ is the set defining the constant cross-product constraint of PARAFAC2, which ensures unique decomposition up to scaling and permutation ambiguities under certain conditions (Kiers et al., 1999). The added flexibility leads to the additional challenge of sign ambiguity in PARAFAC2. Reformulating (2) as $\mathbf{X}_k \approx \sum_{r=1}^R c_{k,r}(\mathbf{a_r} \circ [\mathbf{b_k}]_\mathbf{r})$, where $c_{k,r} = \mathbf{C}(k,r)$, $\mathbf{a}_r = \mathbf{A}(:,r)$, and $[\mathbf{b}_k]_r = \mathbf{B}_k(:,r)$, we observe that $c_{k,r}$ may arbitrarily flip signs together with $[\mathbf{b}_k]_r$ (Harshman, 1972). We resolve this ambiguity by imposing non-negativity on $\mathbf{C}$ (Kiers et al., 1999). Given a *taxa* by *time* by *subjects* tensor $\mathcal{X}$, PARAFAC2 relaxes the strict assumption of the CP model by allowing temporal patterns $\mathbf{B}_k$ to vary across subjects k. In contrast, CP requires identical temporal patterns for all subjects, i.e, $\mathbf{B}_1 = \mathbf{B}_2 = \cdots = \mathbf{B}_K = \mathbf{B}$ (Figure 1). Therefore, the PARAFAC2 model may be better suited for the analysis of datasets in which subjects exhibit heterogeneous temporal patterns.

Taxa, time, and subject loadings of the CP and PARAFAC2 models are visualized in Figure 1, either using all three modes (panel i), or in two modes (panel ii), by scaling the time loadings with their corresponding subject loading, referred to as scaled time loadings, i.e. $c_{k,r}[\mathbf{b}_k]_r \quad \forall k \leq K$.

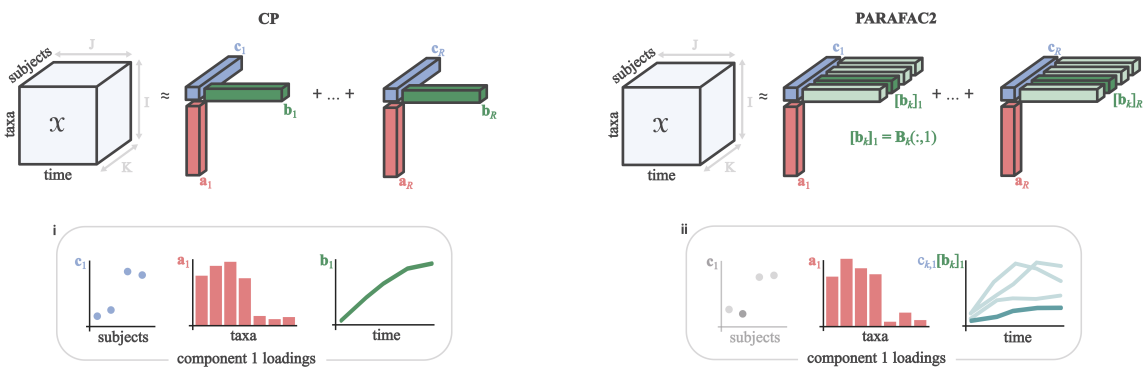

Figure 1: Illustration of an R-component CP (left) and PARAFAC2 (right) model of longitudinal microbiome data arranged as a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$. PARAFAC2 decomposes the tensor $\mathcal{X}$ into taxa factors $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_R]$, subject factors $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_R]$, and host-specific time factors $\mathbf{B}_k = [\mathbf{b}_{k,1}, \ldots, \mathbf{b}_{k,R}]$, $\forall k \leq K$, where K is the number of subjects. PARAFAC2 allows the $\mathbf{B}_k$ factor matrices to vary across tensor slices under the constraint that their cross product is constant. This is more flexible than CP, which finds common time factors $\mathbf{B}_1 = \mathbf{B}_2 = \cdots = \mathbf{B}_K = \mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_R]$ describing all subjects. Component r comprises taxa loadings $\mathbf{a}_r$ characterizing a microbial signature, time loadings $\mathbf{b}_r$ in the case of CP and $[\mathbf{b}_k]_r = \mathbf{B}_k(:,r)$ in the case of PARAFAC2 that capture how the signature evolves over time, and subject loadings $\mathbf{c}_r$ that reflect the strength of the microbial signature over time in each individual. Example visualization of a component across its three modes (i), and in two modes, after deriving *scaled time loadings* by scaling the time loadings with their corresponding subject loading, i.e. $c_{k,r}[\mathbf{b}_k]_r \quad \forall k \leq K$ (ii). Scaled time loadings in the case of CP are calculated analogously as $c_{k,r}\mathbf{b}_r \quad \forall k \leq K$.

Model selection

Selecting the appropriate number of model components R is crucial for ensuring that the resulting model is both biologically meaningful and valid. Several approaches have been proposed to determine

the number of components (Bro and Kiers, 2003; Kolda and Bader, 2009; Fog Froriep Halberg et al., 2024; Rošťáková and Rosipal, 2022); nevertheless, this task remains an active topic of research (Adali et al., 2022). In this study, we determine R based on the replicability and interpretability of the extracted components (Adali et al., 2022; Rošťáková and Rosipal, 2022). Replicability refers to the consistency of identifying similar patterns across random subsamples of the dataset and extends the concept of split-half analysis (Harshman and De Sarbo, 1984). We adopt a replicability-based model selection approach—by subsampling along the subject mode—to ensure that the extracted patterns are robust and representative of the underlying study population, as previously done to select the number of components for CP models (Yan et al., 2024). Let $\mathcal{S} = \{k\}_{k=1}^{K}$ be the set of all slice indices in the subject mode of the complete data tensor. As an example, consider the subsampled tensors, $\mathcal{X}^{(1)} \in \mathbb{R}^{I \times J \times |\mathcal{S}_1|}$ and $\mathcal{X}^{(2)} \in \mathbb{R}^{I \times J \times |\mathcal{S}_2|}$, formed by stacking the slices $\{\mathbf{X}_k \mid k \in \mathcal{S}_1\}$ and $\{\mathbf{X}_k \mid k \in \mathcal{S}_2\}$, respectively, where $\mathcal{S}_1 \subset \mathcal{S}$ and $\mathcal{S}_2 \subset \mathcal{S}$ are randomly selected, partially overlapping subsets of indices and $|\mathcal{S}|$ is the cardinality of the set $\mathcal{S}$. To quantify the similarity between factors extracted from different subsampled tensors (e.g. $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(2)}$), we use the factor match score (FMS). For two CP models, e.g., CP models of $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(2)}$, FMS over the taxa and time modes (FMS_{AB}) is computed as:

$$\text{FMS}_{\text{AB}} = \frac{1}{R} \sum_{r=1}^{R} \frac{|{\mathbf{a}_r^{(1)}}^\top \mathbf{a}_r^{(2)}|}{\|\mathbf{a}_r^{(1)}\| \|\mathbf{a}_r^{(2)}\|} \frac{|{\mathbf{b}_r^{(1)}}^\top \mathbf{b}_r^{(2)}|}{\|\mathbf{b}_r^{(1)}\| \|\mathbf{b}_r^{(2)}\|}, \tag{3}$$

where $\|.\|$ denotes the vector 2-norm, $\mathbf{a}_r^{(1)}$ and $\mathbf{b}_r^{(1)}$ are the r-th columns of factor matrices in the taxa and time modes from the CP model of $\mathcal{X}^{(1)}$ (first submodel), and $\mathbf{a}_r^{(2)}, \mathbf{b}_r^{(2)}$ are the corresponding columns from the CP model of $\mathcal{X}^{(2)}$ (second submodel).

In the case of PARAFAC2, we compute FMS_{A} using the relevant terms (i.e. taxa mode loadings $\mathbf{a}_r$) of Eq. 3. Since the time mode factors are specific to the set of subjects in the data tensor, they can only be compared using the loading vectors corresponding to the set of subjects common to any two submodels. To do this, we define maps such as $\pi_1 : \mathcal{S}_1 \to \{1, \ldots, |\mathcal{S}_1|\}$ and $\pi_2 : \mathcal{S}_2 \to \{1, \ldots, |\mathcal{S}_2|\}$ from a position in the complete tensor to its positional index in the submodel. Let $\mathbf{v}_r^{(1)}$ and $\mathbf{v}_r^{(2)}$ be the vectors of vertically concatenated scaled time loadings corresponding to the subjects shared between submodels one and two:

$$\mathbf{v}_r^{(1)} = \begin{bmatrix} \vdots \\ c^{(1)}_{\pi_1(k),r} [\mathbf{b}^{(1)}_{\pi_1(k)}]_r \\ \vdots \end{bmatrix}_{k \in \mathcal{S}_{1,2}}, \mathbf{v}_r^{(2)} = \begin{bmatrix} \vdots \\ c^{(2)}_{\pi_2(k),r} [\mathbf{b}^{(2)}_{\pi_2(k)}]_r \\ \vdots \end{bmatrix}_{k \in \mathcal{S}_{1,2}},$$

where $\mathbf{v}_r^{(1)}, \mathbf{v}_r^{(2)} \in \mathbb{R}^{J|\mathcal{S}_{1,2}| \times 1}$ and $\mathcal{S}_{1,2} = \mathcal{S}_1 \cap \mathcal{S}_2$. We then calculate $\text{FMS}_{\text{C*B}}$ as

$$\text{FMS}_{\text{C*B}} = \frac{1}{R} \sum_{r=1}^{R} \frac{|{\mathbf{v}_r^{(1)}}^\top \mathbf{v}_r^{(2)}|}{\|\mathbf{v}_r^{(1)}\| \|\mathbf{v}_r^{(2)}\|}. \tag{4}$$

The factor match score is calculated over the scaled time loadings to avoid overemphasizing poorly estimated time loadings, which can arise when the corresponding subject loading is close to zero (Roald et al., 2022).

Because of the inherent permutation ambiguity in both CP and PARAFAC2 decompositions, we first align the order of components to the optimal permutation before computing FMS. The FMS ranges from 0 to 1, with higher values indicating greater similarity. More specifically, we assess the replicability of an R-component CP and PARAFAC2 model as follows.

1. Randomly split the dataset along the subjects dimension into F approximately equal parts, preserving class proportions if class labels are relevant.
2. For each part $f = 1, ..., F$ create a subsampled tensor by excluding the f-th part.
3. Fit an R-component submodel to each of the F subsets.
4. Compute factor similarity for all $\binom{F}{2}$ pairs of the F submodels, using FMS_{AB} for CP, and both FMS_{A} and $\text{FMS}_{\text{C*B}}$ for PARAFAC2.
5. Repeat (1-4) ten times.

The number of subsets F depends on the characteristics of the dataset in use, such as the number of subjects K and the presence of relevant grouping labels. We consider an R-component model to be replicable if at least 90% of the pooled FMS_{AB} (or FMS_{A} and $\text{FMS}_{\text{C*B}}$ in the case of PARAFAC2) values exceed 0.9. In general, to maximize explanatory power, we select the largest number of components that still yield a replicable model. Beyond the replicability criteria, we adopt an exploratory approach by evaluating the interpretability of the components for biological relevance (Rošťáková and Rosipal, 2022).

Datasets

Simulation study

In order to illustrate the use of PARAFAC2 to capture subject-specific temporal trends in longitudinal microbiome data, we constructed a simulated data tensor including variation in the temporal patterns across subjects. The data was generated in a way that it does not fully conform to the structural assumptions of either CP or PARAFAC2, to better represent real-world scenarios where the true data-generating process is unknown. The simulated data tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ with $I = 36$ microbes, $J = 21$ time points, and $K = 12$ subjects was constructed according to the ground truth factors shown in Figure 2. Each of the three patterns, corresponding to the $R = 3$ components, describes the evolution of a microbial signature $\mathbf{a}_r$ over time according to the temporal profile $[\mathbf{b}_k]_r$ with strength $c_{k,r}$ in subject k. All three components are constructed to encode group-level differences between subsets of subjects through the subject loadings ($\mathbf{c}_r$). In addition, the second component includes subject-specific delays in the temporal profiles ($[\mathbf{b}_k]_2$), through variation introduced to the time loadings across a subset of subjects. In particular, the first pattern is the bell-shaped trend in time of taxa 1-12 and is unique to the group of subjects {1,2,6,7,11,12}. The third pattern is the dual-peaking temporal trend of microbes 19-30 and is unique to the group of subjects {7,8,9,10,11,12}. The second pattern is the subject-specific saturation of microbes 10-21 that is unique to each subject within the group {4,5,6,7,8,9}. Microbes 31-36 have low feature loadings in all subjects at all time points. Of particular interest are the between-group differences, with only one group of subjects exhibiting the taxa signatures in each pattern, as well as the between-subject differences that are specific to a group of subjects in component 2. Simulated tensors were constructed both in the absence and presence of noise. To introduce noise, we constructed a noisy tensor as

$$\mathcal{X}_{noisy} = \mathcal{X} + \eta \left\| \mathcal{X} \right\|_F \frac{\mathcal{E}}{\left\| \mathcal{E} \right\|_F},$$

where $\mathcal{X}$ denotes the noiseless tensor, $\mathcal{E} \sim \mathcal{N}(0, 1)$ is a tensor of i.i.d. Gaussian noise, $\|.\|_F$ denotes the Frobenius norm, and the noise level is set to $\eta = 0.25$.

COPSAC_{2010} cohort

The Copenhagen Prospective Studies on Asthma in Childhood 2010 (COPSAC_{2010}) is a population-based birth cohort of 700 children recruited in pregnancy and followed prospectively with deep clinical phenotyping (Bisgaard et al., 2013). The study was conducted in accordance with the guiding principles of the Declaration of Helsinki and was approved by the Local Ethics Committee (H-B-2008-093), and the Danish Data Protection Agency (2015-41-3696). Both parents gave written informed consent before enrollment. The data used here comprises fecal samples collected at 1 week, 1 month, 1 year, 4 years, and 6 years after birth and analyzed via 16S rRNA gene amplicon sequencing. Details of the laboratory workflow, sequencing and data processing have been described previously (Stokholm et al., 2018; Poulsen et al., 2025). Additional covariates include mode of delivery (vaginal delivery/cesarean section) and maternal intrapartum antibiotics (yes/no). Children with one or more missing samples (incomplete cases) were excluded from the analysis. Species not detected in $\geq 10\%$ of subjects in at least one time point were filtered out. To account for the compositional nature of the data, the centered log-ratio (CLR) transformation was applied, using a pseudo-count of 0.5 (Filzmoser et al., 2010). After pre-processing and data transformation, the data were arranged as a third-order tensor of size $364 \times 5 \times 267$ with modes taxa, time, and subjects, respectively. The data tensor contained no missing entries.

FARMM study

The Food and Resulting Microbial Metabolites (FARMM) study collected daily fecal samples over a 15-day period from 30 adult participants, evenly distributed across three diet groups: vegan, omnivore, and exclusive enteral nutrition (EEN) lacking dietary fiber (Tanes et al., 2021). The study comprised three distinct phases: the dietary phase (days 1–5), the gut microbiota purge phase (days 6–8), and the recovery phase (days 9–15). All participants received antibiotic and polyethylene glycol (Abx/PEG) treatment during study days 6 to 8. For a detailed description of the study design, sample collection, and sequencing procedures, we refer to the original publication (Tanes et al., 2021). Preprocessed shotgun metagenomics-based relative abundance data were obtained from (Ma and Li, 2023). Taxa detected in fewer than five samples with a relative abundance of at least 10^{-5} were filtered out. Additionally, time point zero was removed because no subjects in the vegan group had samples at that time point. Subsequently, the CLR transformation was applied after adding a pseudo-count of 0.5. The data were arranged as a third-order tensor of size $343 \times 15 \times 30$ with modes taxa, time, and subjects, respectively. The data tensor contained 11.8% missing entries.

Experimental setup

Implementation details

Before applying the tensor decompositions, data tensors were divided by their Frobenius norm. For fitting the CP models, we used an alternating least squares-based algorithm (Bro, 1998) implemented in Tensorly (v0.9.0) (Kossaifi et al., 2019). The PARAFAC2 models were fit using alternating optimization with the alternating direction method of multipliers (AO-ADMM) algorithm (Roald et al., 2022) implemented in MatCoupLy (v0.1.6) (Roald, 2023). In experiments with missing data, we used an expectation maximization (EM)-based approach with both CP and PARAFAC2 (Kossaifi et al., 2019; Tomasi and Bro, 2005; Chatzis et al., 2025). We incorporated the following constraints and regularization terms during the analysis of each dataset:

- **Simulation study:** Non-negativity in the subjects mode (i.e., on factor matrix $\mathbf{C}$) in PARAFAC2.
- **COPSAC$_{2010}$ data:** Ridge penalty in all modes in CP. Non-negativity in the subjects mode and ridge penalty in all modes in PARAFAC2.
- **FARMM data:** Ridge penalty in all modes in CP. Non-negativity in the subjects mode, ridge penalty in the taxa and subjects modes (i.e., $\mathbf{A}, \mathbf{C}$), and graph Laplacian penalty in the time mode (i.e., $\{\mathbf{B}_k\}_{k=1}^{K}$) in PARAFAC2.

The ridge penalty was added to aid convergence (Boyd and Vandenberghe, 2004; Huang et al., 2016), using a penalty parameter set to 10^{-3}. Non-negativity was imposed to resolve the sign ambiguity of PARAFAC2. Due to missing samples in the FARMM study, the dataset contains completely missing mode-1 fibers (i.e., all taxon relative abundances $\mathbf{X}_k(:, j)$ for subject k at time point j are missing). PARAFAC2 struggles with such scenarios, as $\mathbf{B}_k(j, :)$ can be chosen arbitrarily, provided it conforms to the imposed constraints. To improve the robustness of the recovered $\mathbf{B}_k$ factor matrices in PARAFAC2, we applied a smoothness constraint in the form of a graph Laplacian penalty (with penalty parameter 10^{-3}) on columns of the $\{\mathbf{B}_k\}_{k=1}^{K}$ factor matrices to penalize large differences in neighboring time points (Roald et al., 2022).

To avoid convergence to local minima, each model was fit using multiple random initializations. After confirming the uniqueness of the solution, the solution with the best fit was selected for further analysis. Finally, to address the scaling ambiguity of CP and PARAFAC2, we scaled the columns of factor matrices to unit norm.

Performance metrics

To evaluate how well CP or PARAFAC2 models the data, we calculated the model fit as follows:

$$\text{Fit}(\%) = (1 - \frac{\left\|\mathcal{X} - \hat{\mathcal{X}}\right\|_F^2}{\|\mathcal{X}\|_F^2}) \times 100, \tag{5}$$

where $\mathcal{X}$ is the data tensor, $\hat{\mathcal{X}}$ is its reconstruction based on the model. A model fit of 100% indicates that the model perfectly reconstructs the data, leaving no residual error. Any fit below 100% implies that some variation in the data remains unexplained.

For simulations where the ground truth factors were known, we used FMS to assess how accurately the ground truth factors were recovered. Here, we used the FMS across all modes, defined as:

$$\text{FMS} = \frac{1}{R}\sum_{r=1}^{R} \frac{|\mathbf{a}_r^\top \hat{\mathbf{a}}_r|}{\|\mathbf{a}_r\|\|\hat{\mathbf{a}}_r\|} \frac{|\mathbf{b}_r^\top \hat{\mathbf{b}}_r|}{\|\mathbf{b}_r\|\|\hat{\mathbf{b}}_r\|} \frac{|\mathbf{c}_r^\top \hat{\mathbf{c}}_r|}{\|\mathbf{c}_r\|\|\hat{\mathbf{c}}_r\|}, \tag{6}$$

where $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ are the r-th columns of the ground truth factor matrices, and $\hat{\mathbf{a}}_r$, $\hat{\mathbf{b}}_r$, and $\hat{\mathbf{c}}_r$ are the corresponding model estimates. In the case of PARAFAC2, $\mathbf{b}_r$ and $\hat{\mathbf{b}}_r$ refer to vectors produced by vertically concatenating $\{[\mathbf{b}_k]_r\}_{k=1}^K$ and $\{[\hat{\mathbf{b}}_k]_r\}_{k=1}^K$, respectively.

Results

PARAFAC2 recovers subject-specific temporal patterns

To demonstrate the utility of PARAFAC2 in microbiome data analysis, we compared its performance with that of CP in recovering the underlying true factors from a simulated dataset specifically designed to reflect subject-specific variation in the temporal patterns. The ground truth as well as the recovered factors are shown in Figure 2.

Model fit was high for both methods, i.e., for CP, it was 98.8% and for PARAFAC2, the model fit was 99.9%. The factor match score between the ground truth and recovered factors was 0.994 for CP and 0.997 for PARAFAC2. Both models successfully recovered the taxa factors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ and identified the group differences associated with the encoded microbial signatures (dashed colored/continuous gray lines). Furthermore, the scaled time patterns in components 1 and 3 ($c_{k,1}[\mathbf{b}_k]_1$, $c_{k,3}[\mathbf{b}_k]_3$, for $k = 1, ..., K$), encoding the common temporal dynamics across subjects, were also recovered by both models. However, CP did not recover the subject-specific time loadings of component 2. Since CP assumes a shared time loading between subjects per component, the temporal profile of the recovered time loading is identical across subjects; it is only scaled by the subject loadings $c_{k,2}$. In contrast, the added flexibility of PARAFAC2 allowed for the accurate recovery of both the between-group and between-subject variability in the scaled time loadings of component 2. The temporal patterns of subjects 4-9 were faithfully reproduced (Figure 2), with a mean absolute error in $\hat{c}_{k,2}[\hat{\mathbf{b}}_k]_2$ across time points and subjects of 0.0068 for the CP model versus 0.0013 for PARAFAC2, demonstrating the better performance of PARAFAC2 in terms of capturing subject-specific dynamics.

Similar results were observed, with PARAFAC2 outperforming CP in the case where the ground truth tensor contained 25% noise (Figure S1).

Subject-specific maturation patterns of the infant gut microbiome in the COPSAC_{2010} cohort

To characterize infant gut microbiome maturation while accounting for subject-specific differences in the temporal evolution of dominant microbial signatures, we applied the PARAFAC2 decomposition to 16S rRNA longitudinal gut microbiome data on species level from 267 children of the COPSAC_{2010} cohort. To highlight the distinctions in the extracted patterns, we also compared the results with those obtained using the CP decomposition. We focus on the model components capturing microbial variation in early life and related to mode of delivery and maternal intrapartum antibiotic exposure to illustrate the approaches. For downstream analysis and visualization, subjects were grouped into exposure groups according to mode of delivery and maternal intrapartum antibiotics: (i) vaginal delivery without intrapartum antibiotic prophylaxis (VD-no IAP), (ii) vaginal delivery with intrapartum antibiotic prophylaxis (VD-IAP), and (iii) Cesarean section (with intrapartum antibiotics; CS).

We selected a 7-component CP model and a 5-component PARAFAC2 model based on our model selection approach focusing on replicability and interpretation of the components (Figure S2). The replicability analysis was conducted using $F = 10$ subsets, generated through stratified sampling based on delivery mode to account for its known impact on gut microbiome development (Dominguez-Bello et al., 2010; Jakobsson et al., 2014; Shao et al., 2019; Stokholm et al., 2020). For the PARAFAC2 model, the dimensions of the dataset limited the maximum number of extractable components to five, in accordance with the model's uniqueness conditions (Kiers et al., 1999). The selected CP and PARAFAC2 models explained 49.8 and 50.0% of the data, respectively. Both CP and PARAFAC2 identified components that capture early-life patterns—specifically, variation in samples collected at 1

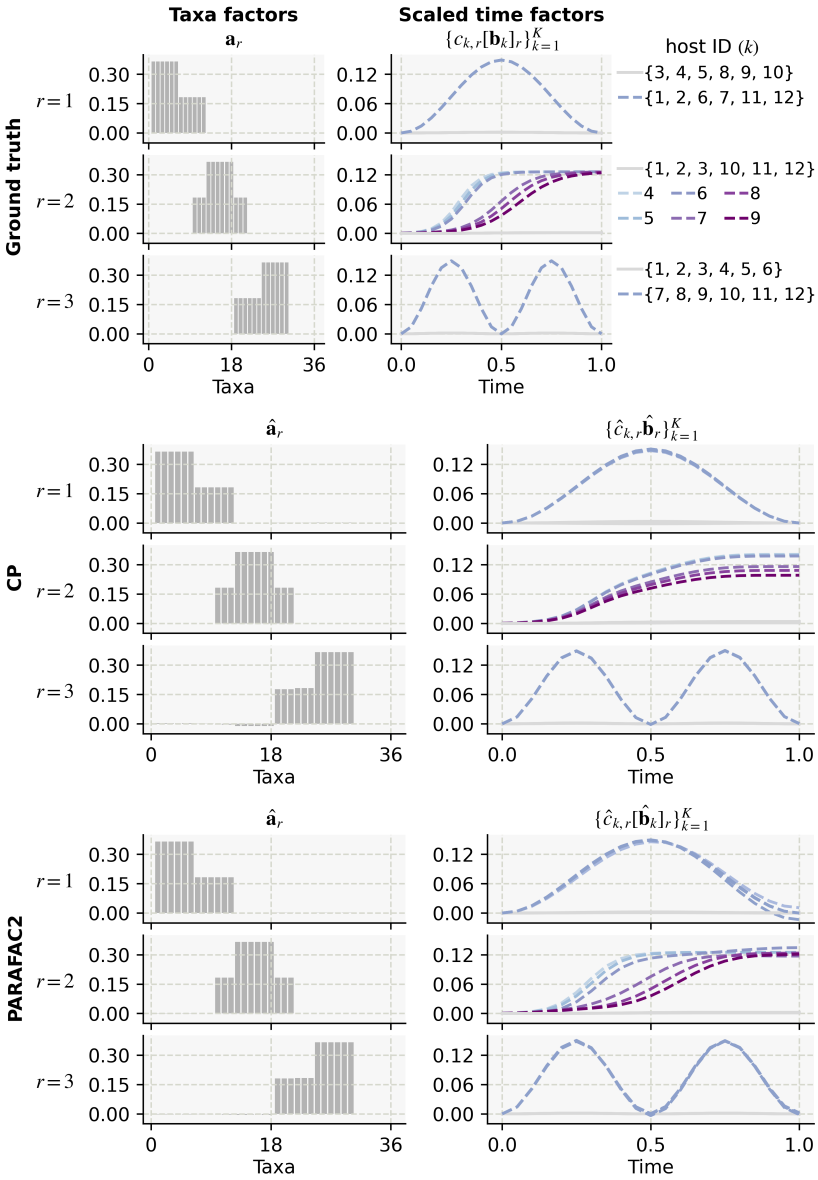

Figure 2: Ground truth factors underlying the simulated dataset and their recovery by CP and PARAFAC2. Top panel: ground truth factors including the taxa ($\mathbf{a}_r$), time ($[\mathbf{b}_k]_r$), and subjects ($\mathbf{c}_r$) loadings visualized after scaling the time loadings by the corresponding subject loading ($c_{k,r}[\mathbf{b}_k]_r$). The microbial signature $\mathbf{a}_r$ is present in subject k over time according to their scaled time loadings $c_{k,r}[\mathbf{b}_k]_r$. Group differences characterized by a presence/absence of the microbial signature in the subject is indicated by dashed colored/continuous gray lines. The patterns in component 2 contain subject-specific time loadings $c_{k,2}[\mathbf{b}_k]_2$ for $k \in \{4, 5, 6, 7, 8, 9\}$ as highlighted by the respective line colors. Middle and bottom panels: factors recovered by CP and PARAFAC2, respectively.

week, 1 month, and 1 year of age (PARAFAC2: components 3 & 5, Figure 3; CP: components 2 & 7, Figure S5). The complete models, including all components, are visualized in Figures S3 and S4.

The early-life representations of the CP and PARAFAC2 models were similar, both in terms of the uncovered microbial signatures and the scaled time loadings (Figure S5, panels i-iv; Figure 3, panels i-iv). The top 15 taxa with the largest absolute weights in each of the early-life-related components (CP component 2 vs. PARAFAC2 component 3, and CP component 7 vs. PARAFAC2 component 5) showed an 87% and 80% overlap between CP and PARAFAC2 at the ASV level. When considering the union of these top 15 ASVs across the components, the overlap between models is 89%, highlighting the consistency in the captured microbial signatures. The corresponding scaled time loadings indicated a similar decreasing trend over time in both models, with some exposure group differences between the models.

Both CP and PARAFAC2 recovered a component representative of a general early-life microbial signature (CP component 2 and PARAFAC2 component 3), characterized by typical early colonizers such as *Escherichia* species and Enterobacteriaceae (ASV annotated only at family level). (Figure S5, panel i; Figure 3, panel i) (Bäckhed et al., 2015; Hill et al., 2017). In both models, the scaled time loadings show that the prevalence of this signature decreases gradually over time (Figure S5, panel ii; Figure 3, panel ii). However, the models differed in how this signature varied across exposure groups. In CP, the subject loadings indicate a difference between VD-no IAP and VD-IAP (Figure S5, panel v), whereas the corresponding PARAFAC2 subject loadings distinguish VD (both without and with IAP) from CS. In addition, the PARAFAC2 scaled time loadings $c_{k,3}[\mathbf{b}_k]_3$ captured more detailed differential dynamics compared to CP (Figure 3, panel ii), demonstrating that the microbial signature was more dominant in CS compared to VD particularly at 1 week.

The second early-life component (CP component 7 and PARAFAC2 component 5) captured a microbial signature with higher abundances of *Bacteroides* species in both models, accompanied in CP by simultaneously decreased abundance of taxa such as *Clostridium perfringens* (Figure S5, panel ii; Figure 3, panel ii). The subject scores for these microbial signatures differed significantly by mode of delivery in both models (Figure S5, panel vi; Figure 3, panel vi) and were consistent with previously reported effects of delivery via Cesarean section (Fallani et al., 2010; Jakobsson et al., 2014; Azad et al., 2016; Shao et al., 2019; Stokholm et al., 2020). The scaled time loadings suggest that the early-life microbial signature diminishes over time in all groups—despite perturbations such as Cesarean section or antibiotic exposure—indicating the replacement of early colonizers and the maturation of the gut microbiomes (Bokulich et al., 2016; Milani et al., 2017). In particular, the scaled time loadings from the PARAFAC2 model provided a more detailed view of the temporal dynamics than those from the CP model, suggesting that the microbial signature remains relatively stable in Cesarean-section–born infants, characterized by a consistently lower initial abundance of *Bacteroides* species, whereas in vaginally delivered infants the initially higher abundance of *Bacteroides* gradually decreases over time (Figure 3, panel iv). Furthermore, when examining subject-specific profiles rather than group mean trajectories, PARAFAC2 revealed substantial heterogeneity in gut microbiome maturation, effectively capturing subject-specific temporal shifts and delays (Figure 3, panel vii; and Figure S5, panel vii).

The microbial signatures of the early-life components captured by both models were relatively sparse (i.e., taxa loadings were characterized by few taxa), whereas the later-life components reflected more mature and taxonomically diverse signatures (Figures S3 and S4) characterized by denser taxa loadings, in line with what is generally observed in childhood gut microbiome development (Bäckhed et al., 2015; Nunez et al., 2025).

Heterogeneous gut microbiome adaptation to dietary intervention in the FARMM study

We applied CP and PARAFAC2 to shotgun metagenomics-based longitudinal gut microbiome data from adults in the FARMM study to explore the effects of dietary and antibiotic/polyethylene glycol intervention (Abx/PEG) (Tanes et al., 2021). The replicability-based model selection approach resulted in a 3-component CP model and a 3-component PARAFAC2 model (Figure S6). Due to the smaller sample size ($N = 10$ in each of 3 study groups), the replicability analysis was conducted with $F = 5$ subsets, with stratified sampling for study group. The selected CP and PARAFAC2 models explained 48.3 and 51.0% of the data, respectively. Both models identified two components related to the dietary phase of the study (days 1-5) and one component related to the recovery phase (days 9-15) following the Abx/PEG perturbation (Figures S7 and S8).

The overall patterns of microbial features suggested a general concordance between the models

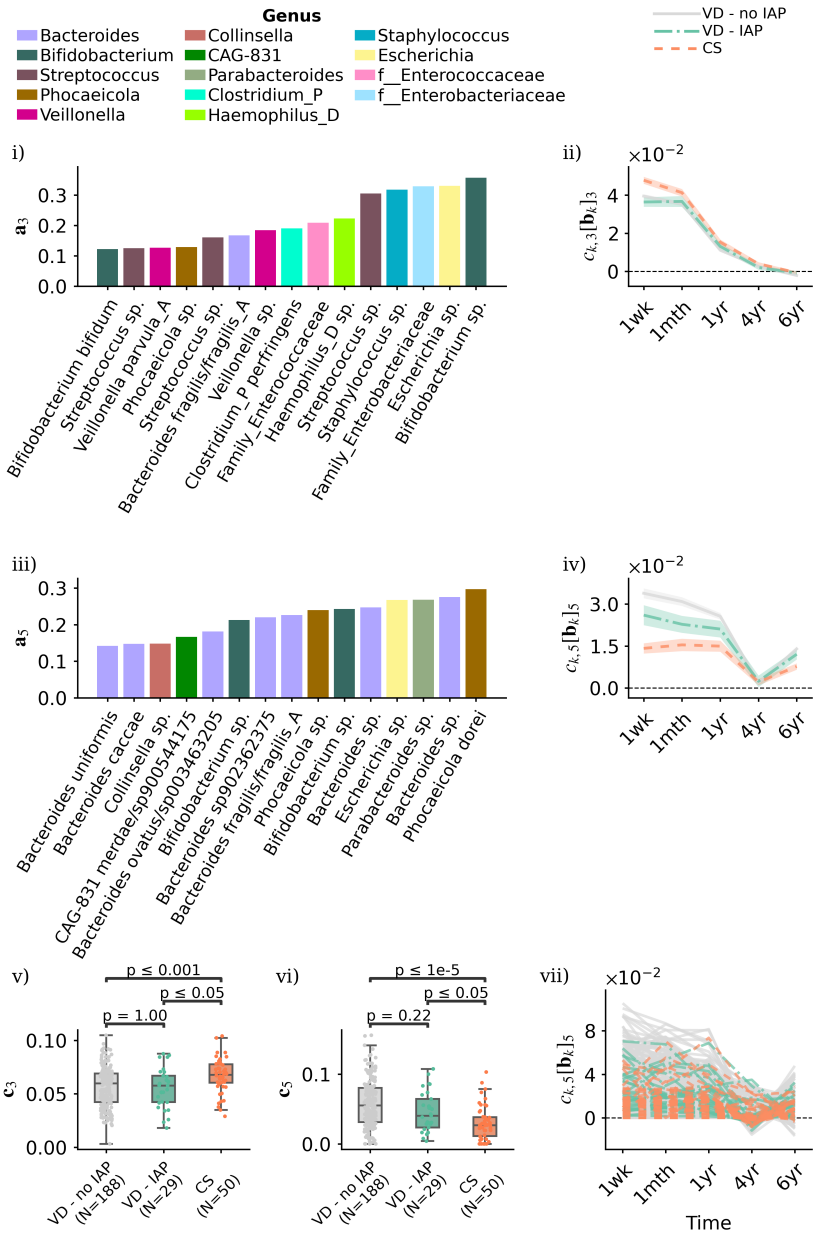

Figure 3: PARAFAC2 uncovers subject-specific temporal patterns of gut microbial signatures in early life, associated with delivery mode and maternal intrapartum antibiotics in the COPSAC_{2010} cohort. Components 3 (i–ii) and 5 (iii–iv) capture patterns in gut microbiome composition during the first year of life. Panels (i, iii) show taxa loadings with the 15 ASVs of largest absolute weight in the component, colored by genus. Scaled time loadings (ii, iv) at time points were averaged per exposure group. Shaded area indicate standard errors of group means. Panels (v, vi) present subject loadings in components 3 and 5 compared across exposure groups using the Mann–Whitney U test with Bonferroni correction. Panel (vii) illustrates subject-specific scaled time loadings of component 5, highlighting heterogeneity in the trajectories of the corresponding microbial signature.

across phases, as reflected in the substantial overlap of top-ranked species across components. In the first dietary phase–related component (component 1, both models; Figure 4), 87% of the top 15 species with the largest absolute weights overlapped between models. Overlap was complete (100%) for the recovery phase–related components (Figure S9 panels i, iv), with species typically associated with exposure to antibiotics such as *Veillonella* and *Klebsiella* being identified. The second dietary phase–related component showed greater divergence, with 60% overlap between the two models (Figure S10). Here, the microbial signature in both models indicated an increased relative abundance of *Clostridium* species, while CP attributed larger absolute weights to the simultaneously decreased relative abundance of species, such as *Eubacterium eligens* and *Bifidobacterium adolescentis* (Figure S10, panels i, iv). These patterns reflected the difference in microbial composition between the EEN group compared to the other diet groups (Figure S10, panels iii, vi).

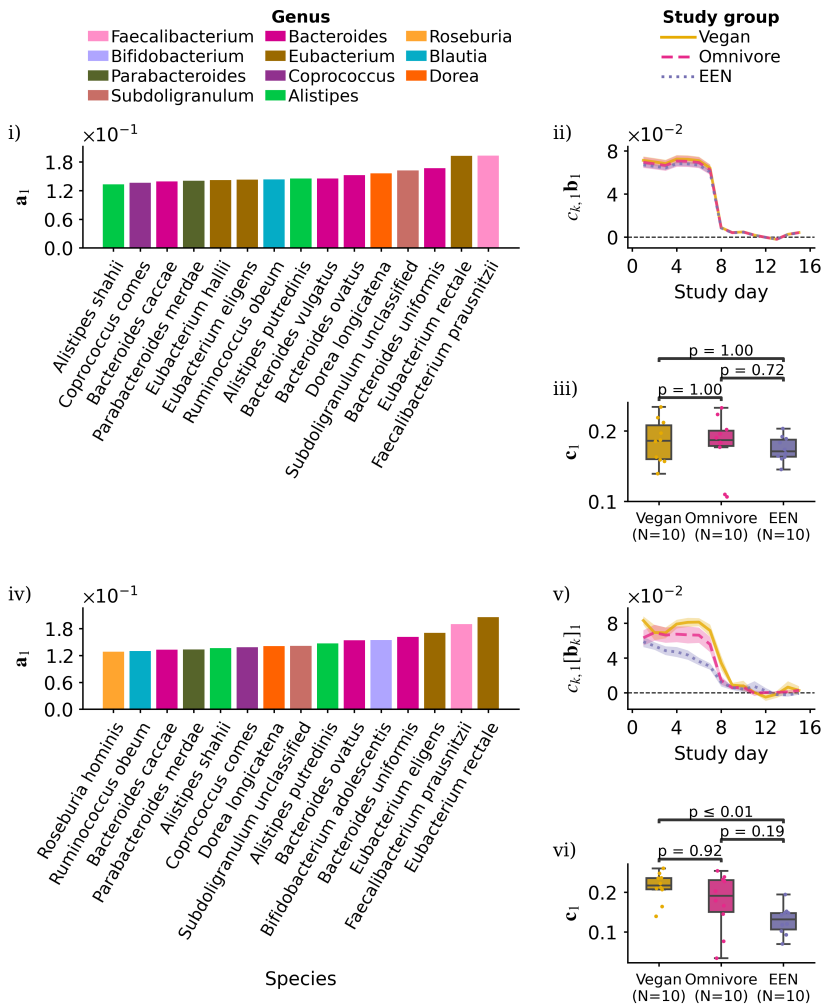

Figure 4: PARAFAC2 reveals distinct temporal patterns of gut microbiome response to dietary and antibiotic interventions in the FARMM study. Shown are component 1 of the selected CP (panels i–iii) and PARAFAC2 (panels iv–vi) models, both associated with the dietary phase (days 1–5) and the antibiotics phase (days 6–8) of the trial. Panels (i, iv) display the 15 species with the largest absolute weights in each component, representing the microbial signatures extracted by the models. Scaled time loadings (ii, v) at time points were averaged per study group. Shaded area indicate standard errors of group means. Panels (iii, vi) present subject loadings compared across exposure groups using the Mann–Whitney U test with Bonferroni correction.

Both models identified broadly similar temporal trajectories of the microbial signatures, as reflected in the scaled time loadings; however, PARAFAC2 captured additional nuance in these trajectories, revealing subtle variations over time not apparent in the CP results. PARAFAC2 uncovered distinct temporal dynamics in the microbial response to the exclusive enteral nutrition (EEN) diet compared to the vegan and omnivore diets in component 1, which was not captured by the CP model (Figure 4,

panel v). In particular, the EEN group exhibited a decrease in the abundance of the microbial signature over time during the dietary phase, unlike the other diet groups, which remained stable, as indicated by the scaled time loadings. The difference in the extracted temporal patterns is also reflected in the corresponding subject loadings of PARAFAC2, which identified a distinction between the EEN and vegan groups (Figure 4, panel vi). These findings align with those reported by Tanes et al. (2021), and highlight the suitability of PARAFAC2 for detecting nuanced, subject-specific temporal patterns. Additionally, the scaled time loadings in the recovery phase-related component indicated a more prominent difference between the EEN group compared to the other groups over time. In contrast, the CP model suggested a constant difference across the entire study duration (Figure S9, panels ii, v).

Replicability of subject-specific factors

Beyond the global replicability ($\text{FMS}_{\text{C*B}}$) used in model selection, we also assessed the uncertainty of individual trajectories to evaluate the stability of subject-specific temporal patterns in both the COPSAC_{2010} and FARMM datasets. To do this, we first identified, for each subject, the specific submodels from the replicability analysis in which they were included (89/100 and 39/50 submodels per subject in the case of COPSAC_{2010} and FARMM, respectively). We then calculated the standard deviation of their scaled time loadings at each time point across the submodels and averaged the standard deviations to use as an indicator of the variance in their inferred profiles. This analysis revealed that the average standard deviations were generally an order of magnitude smaller than the magnitude of the scaled time loadings, indicating good agreement across submodels and a robust recovery of their individual temporal patterns (Figure 5, panels i, iii).

To illustrate this spectrum of uncertainty, we visualized the scaled time loadings from all relevant submodels for a selection of representative subjects from both cohorts (Figure 5). Overall, the profiles showed good agreement across submodels, confirming their replicability at the subject-specific level, even within the same exposure group. For example, panel (ii) shows the mean (± 2 standard deviation) for three subjects (A, B, C) in the VD - no IAP exposure group from the COPSAC_{2010} dataset, all of which show high consistency (subject-specific average $\text{FMS}_{\text{C*B}} \geq 0.99$). Similarly, panel (iv) shows robust profiles for subjects 9003 and 9024 of the EEN study group from the FARMM dataset, with subject-specific average $\text{FMS}_{\text{C*B}} \geq 0.99$. The comparatively larger uncertainty observed for subject 9032 (subject-specific average $\text{FMS}_{\text{C*B}} = 0.97$) reflects the underlying missingness pattern of the data, as this subject was missing samples on study days 1-2 and 7-9 (see Discussion). Results displaying the complete set of model components are shown in Figures S11 and S12, for the COPSAC_{2010} and FARMM studies, respectively.

Discussion

In this study, we introduced the PARAFAC2 model as a powerful tool for the unsupervised analysis of longitudinal microbiome data, specifically to capture subject-specific temporal trajectories. Furthermore, to ensure the robustness of the uncovered patterns, we also proposed a replicability-based framework for determining the number of model components, ensuring the biological relevance and stability of the extracted signatures.

The results from the simulated data highlight a key limitation of CP-based tensor decomposition approaches in longitudinal microbiome data analysis: the inability to capture subject-specific microbial dynamics (such as accelerations or delays), phenomena frequently observed in real populations (Jakobsson et al., 2014; Mitchell et al., 2020; Nguyen et al., 2020). In the simulated data, the microbial signature represented by component 2 reached peak abundance at approximately $t \approx 0.4$ for subject 4 and $t \approx 0.9$ for subject 9, where t represents time (Figure 2). PARAFAC2 accurately recovered this difference in peak timing, whereas CP produced identical temporal profiles across subjects, with a peak near $t \approx 0.75$. Moreover, the CP-derived loadings suggested variability in the peak magnitude of the signature across subjects that was not supported by the data. In real data, subject-specific patterns, such as differences in the saturation timing of a microbial signature, may reflect distinct environmental exposures, host health conditions, or phenotypic states, underscoring the importance of accurately capturing these individual trajectories.

Although both CP and PARAFAC2 explained the data well and achieved near-perfect factor match scores in our simulation study, these standard model diagnostics were not sensitive enough to reveal differences in their recovery of subject-specific temporal patterns. Despite the large variation in the peak timing between subjects, the models appeared nearly equivalent by conventional performance

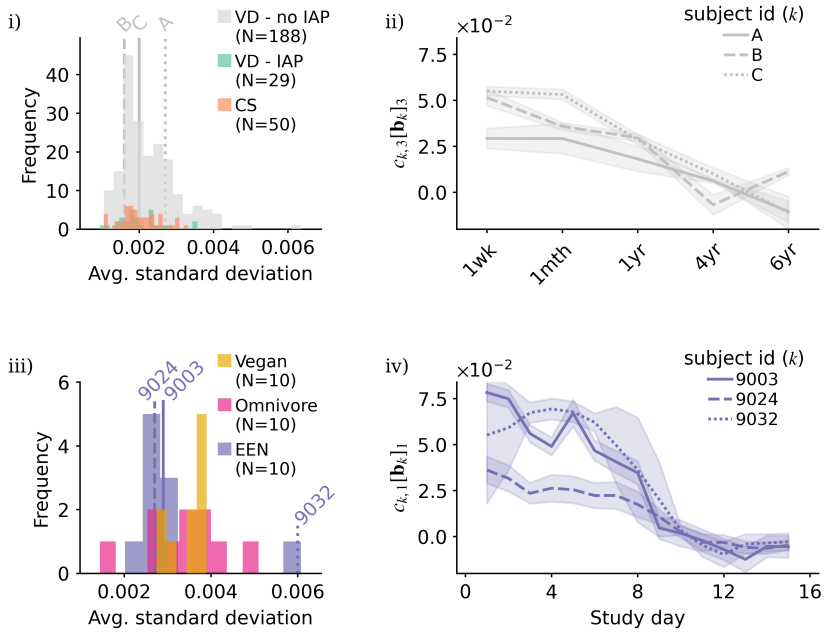

Figure 5: Replicability of PARAFAC2-based subject-specific scaled time loadings in the COPSAC2010 and FARMM datasets. Panels (i, iii) show histograms of the average standard deviation, calculated for each subject by first finding the standard deviation of the scaled time loadings at each time point across submodels, and then taking the average. Panels (ii, iv) show the mean (lines) ± 2 standard deviations (shaded areas) of scaled time loadings calculated at each time point across submodels for three representative subjects from the VD - no IAP group in the COPSAC_{2010} cohort (panel ii) and the EEN study group in the FARMM (panel iv) study.

metrics; only a closer visual examination of the estimated trajectories uncovered that PARAFAC2 more accurately preserved individual-level variation, whereas CP tended to misrepresent these subject-specific dynamics. It is important to note that neither model achieved perfect recovery of the ground truth factors, which is expected, as the simulated data did not fully conform to the structural assumptions of either model. Nevertheless, since the true data-generating process is often unknown, the added flexibility of PARAFAC2 may make it the more suitable modeling choice when individual variation in temporal profiles is expected or of biological interest.

In real data, CP and PARAFAC2 revealed broadly consistent microbial signatures and temporal trajectories. Despite the similarity in microbial signatures, the scaled time loadings of PARAFAC2 provided more granular insights into the temporal evolution of microbial signatures at the subject level compared to those from CP. For instance, in our analysis of infant gut microbiome maturation, PARAFAC2 revealed variation among Cesarean-section-born infants, with some showing an earlier change toward a composition resembling that of vaginally born infants. Identifying these subject-specific maturation trajectories is important, as they may reflect differences in postnatal exposures, including environmental conditions or feeding practices that help to compensate for the early-life perturbations (Beller et al., 2021; Mercer et al., 2024; Nunez et al., 2025). In the FARMM study, PARAFAC2 detected a compositional change in the EEN group, characterized by decreasing relative abundances of species such as *Faecalibacterium prausnitzii*, *Parabacteroides merdae*, and *Bacteroides ovatus* during the dietary phase—patterns consistent with previously reported results (Tanes et al., 2021) but not captured by the CP model. The CP model assumes that all subjects follow the same temporal trajectory, allowing only for differences in the magnitude of how strongly each individual expresses the microbial signature according to the trajectory. Because of this assumption, inter-individual differences in the timing or progression of microbial changes could not be captured using CP, limiting the model's ability to represent subject-specific dynamics and overlooking biologically meaningful patterns.

A potential limitation of PARAFAC2 lies in the constant cross-product constraint it imposes on the $\mathbf{B}_k$ factor matrices. Although the constraint is essential for obtaining a unique solution, which is required for interpretability, the assumption that loading vectors of $\mathbf{B}_k$ should have constant angles for all k may impose biologically unwarranted structure (Anzardi et al., 2019), potentially limiting the

ability of the model to fully capture the diversity of subject-specific temporal dynamics. Nevertheless, in the analysis of functional magnetic resonance imaging (fMRI), PARAFAC2 has emerged as a robust compromise between models of differing flexibility, effectively capturing connectivity and spatial components while accounting for subject variability (Madsen et al., 2017). In addition, the model has been shown to perform well even when the PARAFAC2 constraint is only approximately satisfied, as demonstrated in our results on simulated data and in previous studies (Bro et al., 1999; Roald et al., 2022).

Another limitation of PARAFAC2 is that the maximum number of components it can extract is constrained by its uniqueness conditions. In particular, for datasets with only a few time points, a CP-based approach may therefore be a better choice, as it can potentially recover more microbial signatures.

An open challenge when fitting PARAFAC2 is the presence of completely missing measurements at a time point for a subject (i.e., a missing mode-1 fiber, such as in the case of the FARMM dataset), which makes the corresponding entries of the time loadings arbitrary as long as the imposed constraints are satisfied. To overcome this indeterminacy and improve the robustness of our results, we employed a smoothness constraint on the temporal mode. Nevertheless, in cases where such missingness is extensive (e.g., 9032 in panel iv, Figure 5), this approach may be insufficient, highlighting the need to investigate alternative strategies for handling missing mode-1 fibers.

Validating the subject-specific patterns inferred by PARAFAC2 is more challenging compared to CP, owing to its added flexibility. In the absence of biological replicates, it is not possible to directly assess whether the subject-specific trajectories reflect true biological variation or are partly driven by model assumptions and data idiosyncrasies. To account for this, we focused on group-level trends, aggregating subject-specific loadings within predefined exposure groups (e.g., by delivery mode and diet group). This approach highlights consistent between-group differences but may obscure individual-level nuances. Therefore, to more directly assess replicability at the subject level, we inspected the heterogeneity of subject-specific scaled time loadings between the submodels fitted during the replicability analysis. By leveraging the partial overlap of subjects between submodels, we could compare subject-specific trajectories across submodels, which confirmed their replicability and supported our model selection approach. Nevertheless, we note that specific parameters of the resampling, such as the number of subsamples, depend on dataset characteristics. Finally, replicate samples, when available, can aid in both determining the number of model components and assessing the biological validity of the discovered patterns. This approach has been successfully demonstrated in metatranscriptomic analysis of microbial communities using CP (Blaskowski et al., 2025), and the study of functional neuroimaging data via PARAFAC2 (Mørup et al., 2025).

In summary, PARAFAC2 provides a flexible and interpretable approach for the analysis of longitudinal microbiome data by explicitly modeling subject-specific temporal trajectories. By accommodating shifts, delays, and other forms of individual variation, PARAFAC2 enables a more accurate characterization of temporal changes. Furthermore, the integration of replicability as a criterion for model selection enhances the robustness of the extracted patterns. Together, these features advance the methodological toolkit for studying temporal microbiome dynamics and open avenues for future applications across diverse longitudinal omics datasets.

Competing interests

No competing interest is declared.

Acknowledgments

We express our deepest gratitude to the children and families of the COPSAC_{2010} cohort. We thank Frederiek Wesel and Roel van der Ploeg for the insightful discussions.

Funding

We acknowledge all funding received by COPSAC, listed on `www.copsac.com`. Specific support for this study was provided by the Lundbeck Foundation (grant no. R269-2017-5, COPSYCH). M.A.R. is funded by the Novo Nordisk Foundation (grant no. NNF21OC0068517).

Data availability

Data and code underlying the simulations are available at `https://github.com/blzserdos/parafac2_microbiome` and archived at `https://doi.org/10.5281/zenodo.17674126`. Shotgun metagenomic sequence data from the FARMM dataset (Tanes et al., 2021) were deposited under BioProject with accession code PRJNA675301. Processed data were obtained from Ma and Li (2023). Individual-level clinical data from the COPSAC_{2010} cohort are not publicly available to protect participant privacy, in accordance with the Danish Data Protection Act and European Regulation 2016/679 of the European Parliament and of the Council (GDPR) that prohibit distribution even in pseudo-anonymized form. Data can be made available under a joint research collaboration by contacting COPSAC's data protection officer (administration@dbac.dk).

References

T. Adali et al. Reproducibility in matrix and tensor decompositions: Focus on model match, interpretability, and uniqueness. *IEEE Signal Processing Magazine*, 39(4):8–24, 2022.

J. M. Amigo et al. Solving GC-MS problems with PARAFAC2. *TrAC Trends in Analytical Chemistry*, 27(8):714–725, 2008.

M. B. Anzardi, J. A. Arancibia, and A. C. Olivieri. Interpretation of matrix chromatographic-spectral data modeling with parallel factor analysis 2 and multivariate curve resolution. *Journal of Chromatography. A*, 1604:460502, 2019.

M. Azad et al. Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study. *BJOG: An International Journal of Obstetrics & Gynaecology*, 123(6):983–993, 2016.

F. Bäckhed et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe*, 17(6):852, 2015.

F. Becker, A. K. Smilde, and E. Acar. Unsupervised EHR-based phenotyping via matrix and tensor decompositions. *WIREs Data Mining and Knowledge Discovery*, 13(4):e1494, 2023.

L. Beller et al. Successional Stages in Infant Gut Microbiota Maturation. *mBio*, 12(6):e01857–21, 2021.

H. Bisgaard et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. *Clinical & Experimental Allergy*, 43(12):1384–1394, 2013.

S. Blaskowski et al. Simultaneous acclimation to nitrogen and iron scarcity in open ocean cyanobacteria revealed by sparse tensor decomposition of metatranscriptomes. *Science Advances*, 11(14):eadr4310, 2025.

N. A. Bokulich et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Science translational medicine*, 8(343), 2016.

S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

R. Bro. *Multi-way analysis in the food industry: models, algorithms, and applications*. Universiteit van Amsterdam, Amsterdam, 1998.

R. Bro and H. A. Kiers. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(5): 274–286, 2003.

R. Bro, C. A. Andersson, and H. A. Kiers. PARAFAC2—Part II. Modeling chromatographic data with retention time shifts. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 13 (3-4):295–309, 1999.

J. D. Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3):283–319, 1970.

C. Chatzis, C. Schenker, M. Pfeffer, and E. Acar. tPARAFAC2: tracking evolving patterns in (incomplete) temporal data. *Data Mining and Knowledge Discovery*, 39(6):83, 2025.

L. Dethlefsen and D. A. Relman. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences*, 108:4554–4561, 2011.

M. G. Dominguez-Bello et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences*, 107(26):11971–11975, 2010.

M. Fallani et al. Intestinal Microbiota of 6-week-old Infants Across Europe: Geographic Influence Beyond Delivery Mode, Breast-feeding, and Antibiotics. *Journal of Pediatric Gastroenterology and Nutrition*, 51(1):77–84, 2010.

P. Filzmoser, K. Hron, and C. Reimann. The bivariate statistical analysis of environmental (compositional) data. *Science of The Total Environment*, 408(19):4230–4238, 2010.

H. Fog Froriep Halberg, M. Bevilacqua, and Å. Rinnan. Resampling as a Robust Measure of Model Complexity in PARAFAC Models. *Journal of Chemometrics*, 38(12):e3601, 2024.

J. K. Goodrich et al. Conducting a Microbiome Study. *Cell*, 158(2):250–262, 2014.

R. Harshman. Foundations of the PARAFAC procedure: Model and conditions for an explanatory factor analysis. *Technical Report UCLA Working Papers in Phonetics 16*, 1970.

R. A. Harshman. PARAFAC2: Mathematical and technical notes. *UCLA working papers in phonetics*, 22(3044):122215, 1972.

R. A. Harshman and W. S. De Sarbo. An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. *Research Methods for Multimode Data Analysis*, pages 602–642, 1984.

C. J. Hill et al. Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome*, 5(1):4, 2017.

F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927.

C. Hoskinson et al. Delayed gut microbiota maturation in the first year of life is a hallmark of pediatric allergic disease. *Nature Communications*, 14(1):4785, 2023.

K. Huang, N. D. Sidiropoulos, and A. P. Liavas. A flexible and efficient algorithmic framework for constrained matrix and tensor factorization. *IEEE Transactions on Signal Processing*, 64(19): 5052–5065, 2016.

B. Hunyadi, P. Dupont, W. Van Paesschen, and S. Van Huffel. Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. *WIREs Data Mining and Knowledge Discovery*, 7(1):e1197, 2017.

H. E. Jakobsson et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by Caesarean section. *Gut*, 63(4):559–566, 2014.

H. A. Kiers, J. M. Ten Berge, and R. Bro. PARAFAC2—Part I. a direct fitting algorithm for the PARAFAC2 model. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 13(3-4): 275–294, 1999.

S. Kodikara, S. Ellul, and K.-A. Lê Cao. Statistical challenges in longitudinal microbiome data analysis. *Briefings in Bioinformatics*, 23(4):bbac273, 2022.

T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

J. Kossaifi, Y. Panagakis, A. Anandkumar, and M. Pantic. TensorLy: Tensor Learning in Python. *Journal of Machine Learning Research*, 20(26):1–6, 2019.

J. B. Kruskal. 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, 1977.

S. Ma and H. Li. A tensor decomposition model for longitudinal microbiome studies. *The Annals of Applied Statistics*, 17(2):1105–1126, 2023.

K. H. Madsen, N. W. Churchill, and M. Mørup. Quantifying functional connectivity in multi-subject fMRI data using component models. *Human Brain Mapping*, 38(2):882–899, 2017.

C. Martino et al. Context-aware dimensionality reduction deconvolutes gut microbial community dynamics. *Nature Biotechnology*, 39:165–168, 2021.

E. M. Mercer et al. Divergent maturational patterns of the infant bacterial and fungal gut microbiome in the first year of life are associated with inter-kingdom community dynamics and infant nutrition. *Microbiome*, 12(1):22, 2024.

C. Milani et al. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiology and Molecular Biology Reviews : MMBR*, 81(4):e00036–17, 2017.

C. M. Mitchell et al. Delivery Mode Affects Stability of Early Infant Gut Microbiota. *Cell Reports Medicine*, 1(9):100156, 2020.

M. Mørup, E. Acar, and T. Adali. Tensor and coupled decompositions for interpretable pattern discovery in multimodal functional neuroimaging data. *IEEE Signal Processing Magazine*, 2025.

J. Nguyen, J. Lara-Gutiérrez, and R. Stocker. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiology Reviews*, 45(4), 2020.

H. Nunez et al. Early life gut microbiome and its impact on childhood health and chronic conditions. *Gut Microbes*, 17(1):2463567, 2025.

I. Perros et al. Temporal phenotyping of medically complex children via PARAFAC2 tensor factorization. *Journal of Biomedical Informatics*, 93:103125, 2019.

C. E. Poulsen et al. No association between the early-life gut microbiota and childhood body mass index and body composition. *Med*, 6(4), 2025.

A. Ramirez et al. Integrative, high-resolution analysis of single-cell gene expression across experimental conditions with PARAFAC2-RISE. *Cell Systems*, page 101294, 2025.

M. Roald. MatCoupLy: Learning coupled matrix factorizations with Python. *SoftwareX*, 21:101292, 2023.

M. Roald et al. An AO-ADMM Approach to Constraining PARAFAC2 on All Modes. *SIAM Journal on Mathematics of Data Science*, 4(3):1191–1222, 2022.

Z. Rošťáková and R. Rosipal. Determination of the number of components in the PARAFAC model with a nonnegative tensor structure: a simulated EEG data study. *Neural Computing and Applications*, 34(17):14793–14805, 2022.

Y. Shao et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature*, 574(7776):117–121, 2019.

P. Shi et al. TEMPTED: time-informed dimensionality reduction for longitudinal microbiome studies. *Genome Biology*, 25(1):317, 2024.

L. Spyrou, M. Parra, and J. Escudero. Complex Tensor Factorization With PARAFAC2 for the Estimation of Brain Connectivity From the EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(1):1–12, 2019.

J. Stokholm et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nature Communications*, 9(1):141, 2018.

J. Stokholm et al. Delivery mode and gut microbial changes correlate with an increased risk of childhood asthma. *Science Translational Medicine*, 12(569):eaax9929, 2020.

C. Tanes et al. Role of dietary fiber in the recovery of the human gut microbiome and its metabolome. *Cell Host & Microbe*, 29(3):394–407.e5, 2021.

G. Tomasi and R. Bro. PARAFAC and missing values. *Chemometrics and Intelligent Laboratory Systems*, 75(2):163–180, 2005.

G. R. van der Ploeg, J. A. Westerhuis, A. Heintz-Buschart, and A. K. Smilde. parafac4microbiome: exploratory analysis of longitudinal microbiome data using parallel factor analysis. *mSystems*, 10(6): e00472–25, 2025.

T. Vatanen et al. Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. *Nature Microbiology*, 4(3):470–479, 2019.

S. Yan et al. Characterizing human postprandial metabolic response using multiway data analysis. *Metabolomics*, 20(50):1–14, 2024.

X. Zhou et al. Longitudinal profiling of the microbiome at four body sites reveals core stability and individualized dynamics during health and disease. *Cell Host & Microbe*, 32(4):506–526.e9, 2024.