Revealing Subject-specific Temporal Patterns from Longitudinal Data - Supplementary Material

Christos Chatzis[1,2*], David Horner[3], Rasmus Bro[4], Ann-Marie Malby Schoos[3,5,6,7], Morten A. Rasmussen[3,4], 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, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark

[4]Department of Food Science, University of Copenhagen, Copenhagen, Denmark

[5]Department of Pediatrics, Copenhagen University Hospital - Næstved, Slagelse and Ringsted, Slagelse, Denmark

[6]Department of Pediatrics, Copenhagen University Hospital - Amager and Hvidovre, Hvidovre, Denmark

[7]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[*]Corresponding authors: christos@simula.no (C.C.) and evrim@simula.no (E.A.).
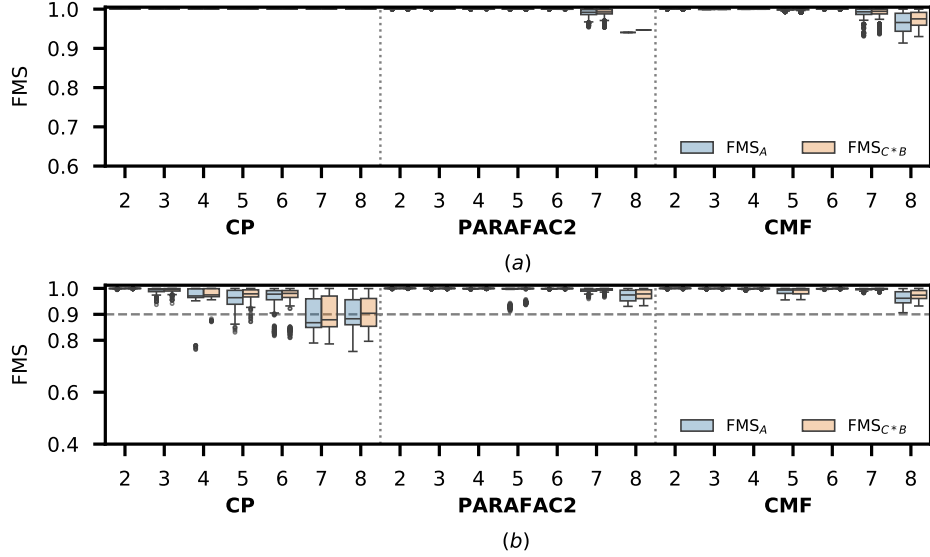
1   Metabolomics

Figure S.1: (a) Reproducibility, and (b) Replicability of different models of the metabolomics data using different number of components (R) for the subset of the female subjects.
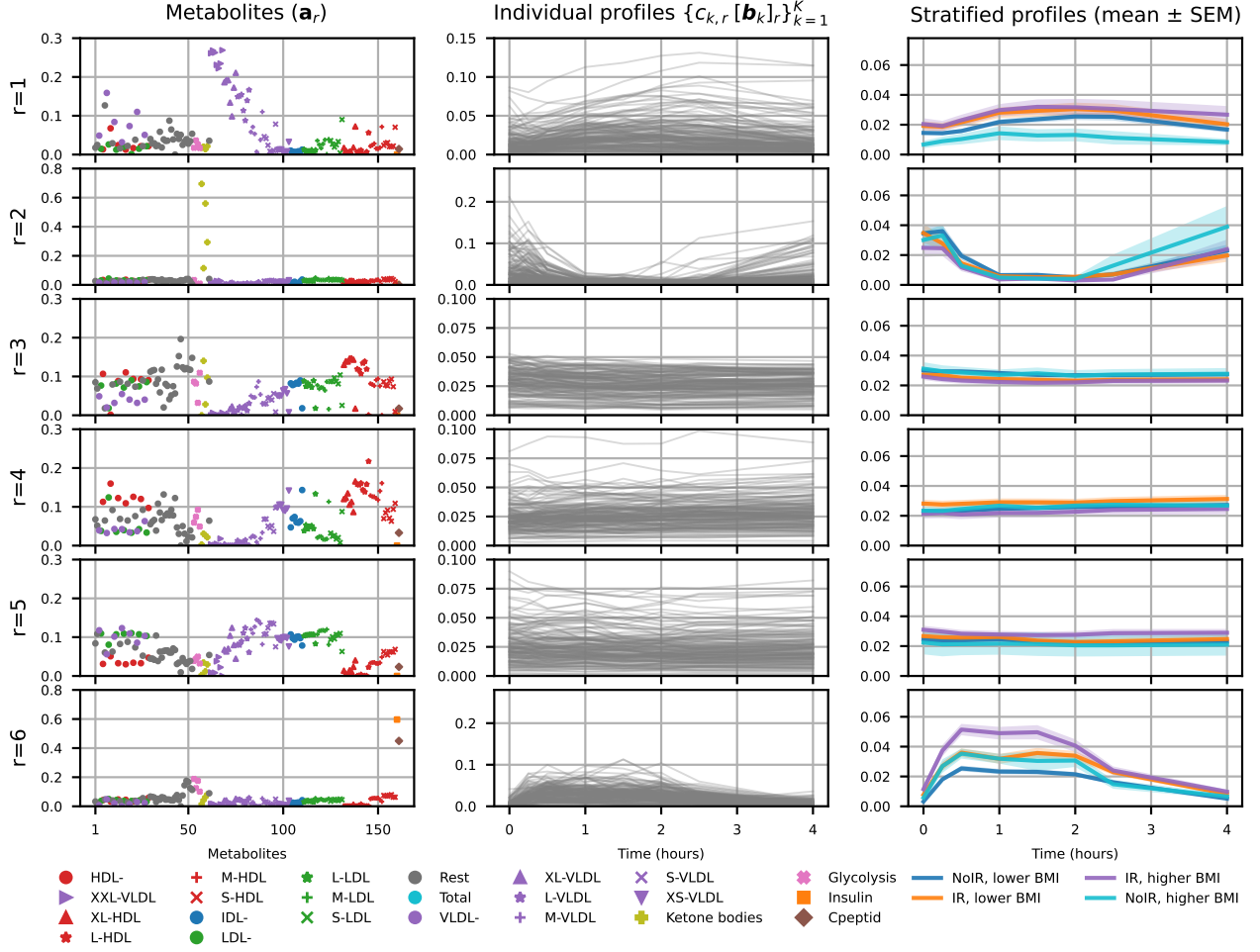
Figure S.2: Metabolomics (female subjects). Components of a 6-component PARAFAC2 model (with non-negativity constraints in all modes) of the metabolomics data from males. $\mathbf{a}_r$ denotes the pattern in the metabolites mode, where metabolites are colored by lipoprotein classes. Different shapes are used for lipoprotein subclasses. Subject-specific time profiles scaled by the corresponding subject scores, i.e., $c_{k,r}[\mathbf{b}_k]_r$ for each component are shown in the middle column. The last column shows scaled subject-specific time profiles colored according to four BMI/IR groups.
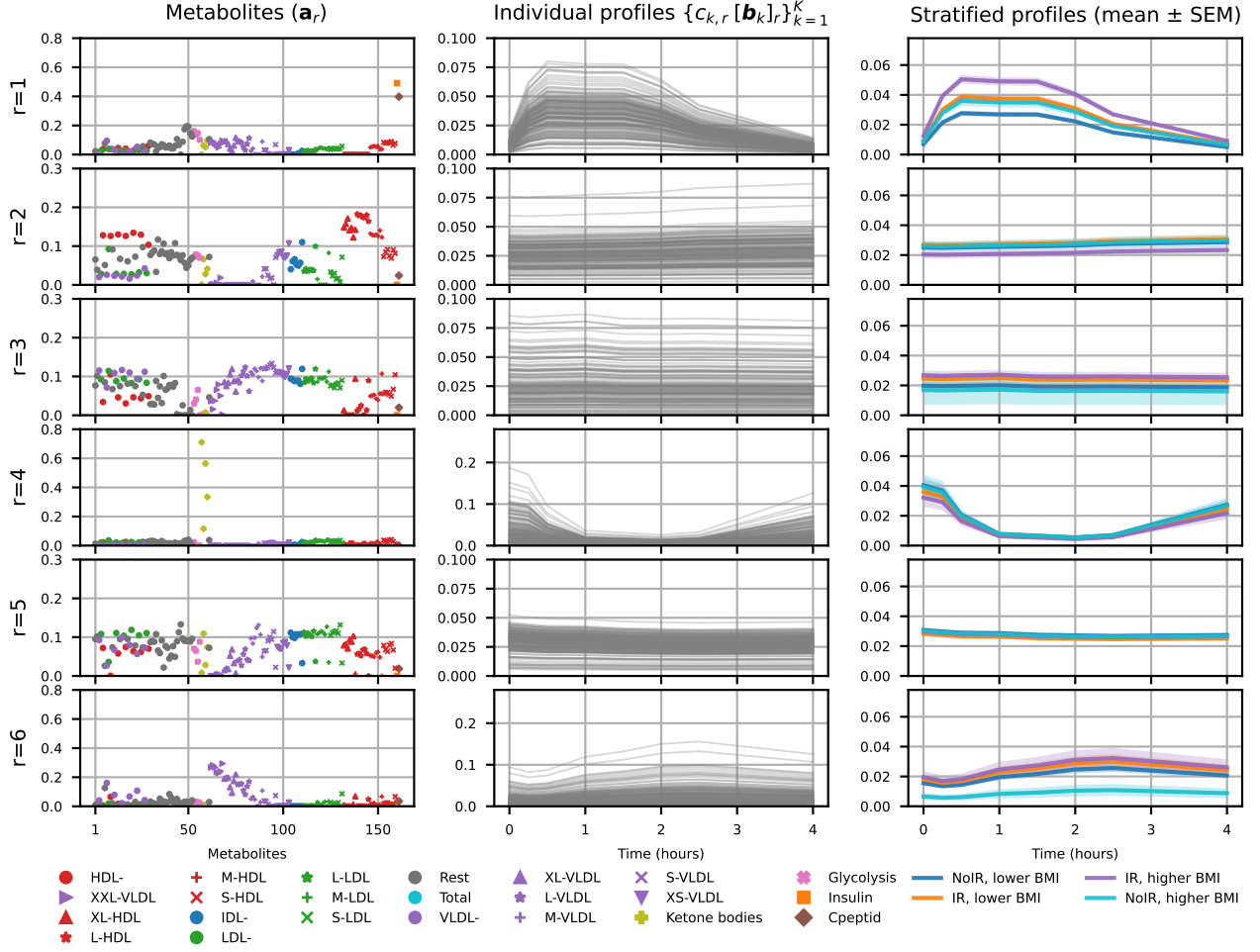
Figure S.3: Metabolomics (female subjects). Components of a 6-component CP model (with nonnegativity constraints in all modes) of the metabolomics data from males. $\mathbf{a}_r$ denotes the pattern in the metabolites mode, where metabolites are colored by lipoprotein classes. Different shapes are used for lipoprotein subclasses. Subject-specific time profiles scaled by the corresponding subject scores, i.e., $c_{k,r}[\mathbf{b}_k]_r$ for each component are shown in the middle column. The last column shows scaled subject-specific time profiles colored according to four BMI/IR groups.
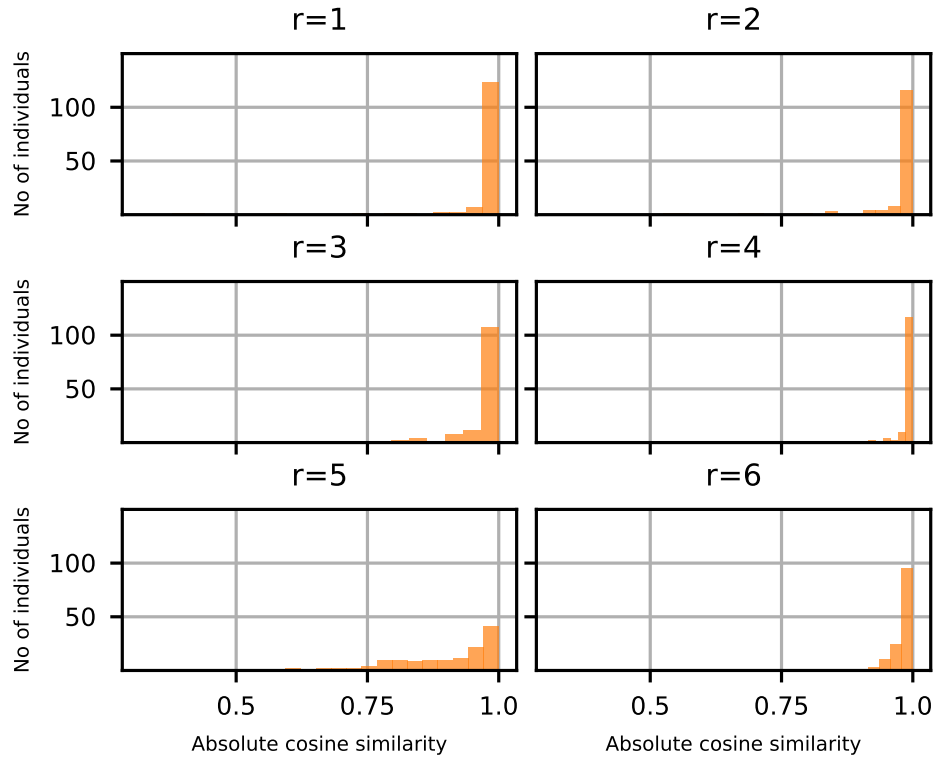
Figure S.4: Comparison of subject-specific time profiles $(c_{k,r}[\mathbf{b}_k]_r)$ captured by CMF and PARAFAC2 for each component for metabolomics, where each histogram contains 140 data points, one for each (male) subject.
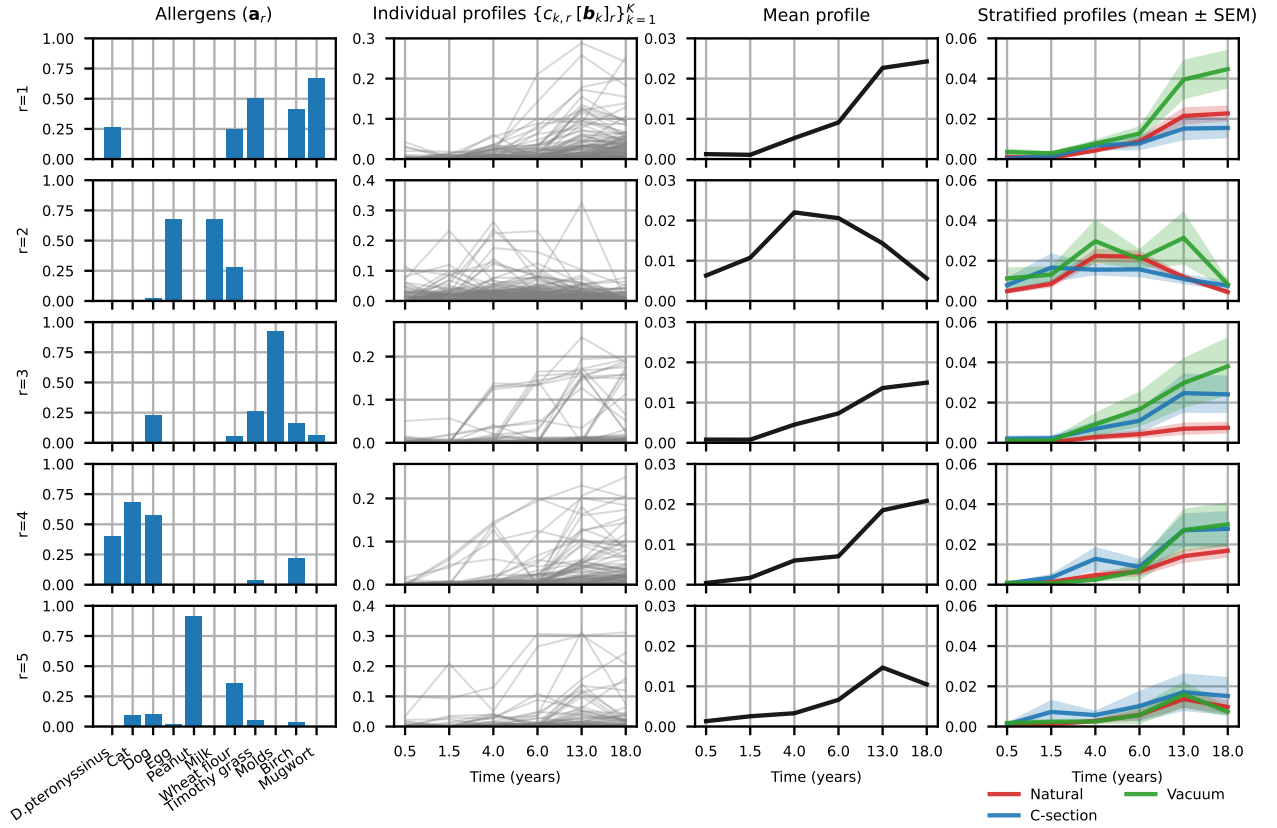
2   Sensitization

Figure S.5: Sensitization. Components of a 5-component PARAFAC2 model (with nonnegativity constraints in all modes) of the sensitization data. $\mathbf{a}_r$ denotes the pattern in the allergens mode. Subject-specific time profiles scaled by the corresponding subject scores, i.e., $c_{k,r}[\mathbf{b}_k]_r$ are shown in the middle column. Mean of scaled subject-specific profiles are plotted in the third column while the last column shows mean (and standard error of mean) patterns of scaled subject-specific time profiles colored according to delivery/birth mode groups.
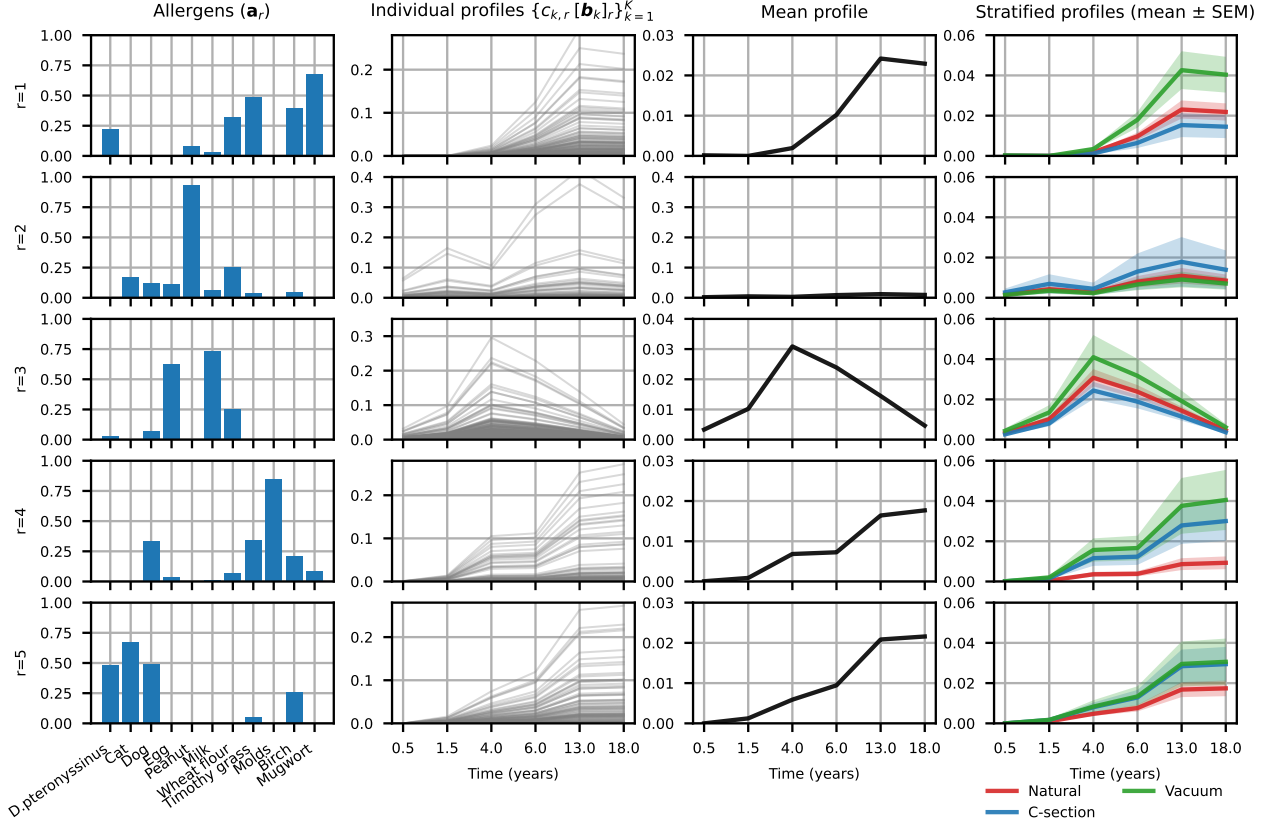
Figure S.6: Sensitization. Components of a 5-component CP model (with nonnegativity constraints in all modes) of the sensitization data. $\mathbf{a}_r$ denotes the pattern in the allergens mode. Subject-specific time profiles scaled by the corresponding subject scores, i.e., $c_{k,r}[\mathbf{b}_k]_r$ are shown in the middle column. Mean of scaled subject-specific profiles are plotted in the third column while the last column shows mean (and standard error of mean) patterns of scaled subject-specific time profiles colored according to delivery/birth mode groups.
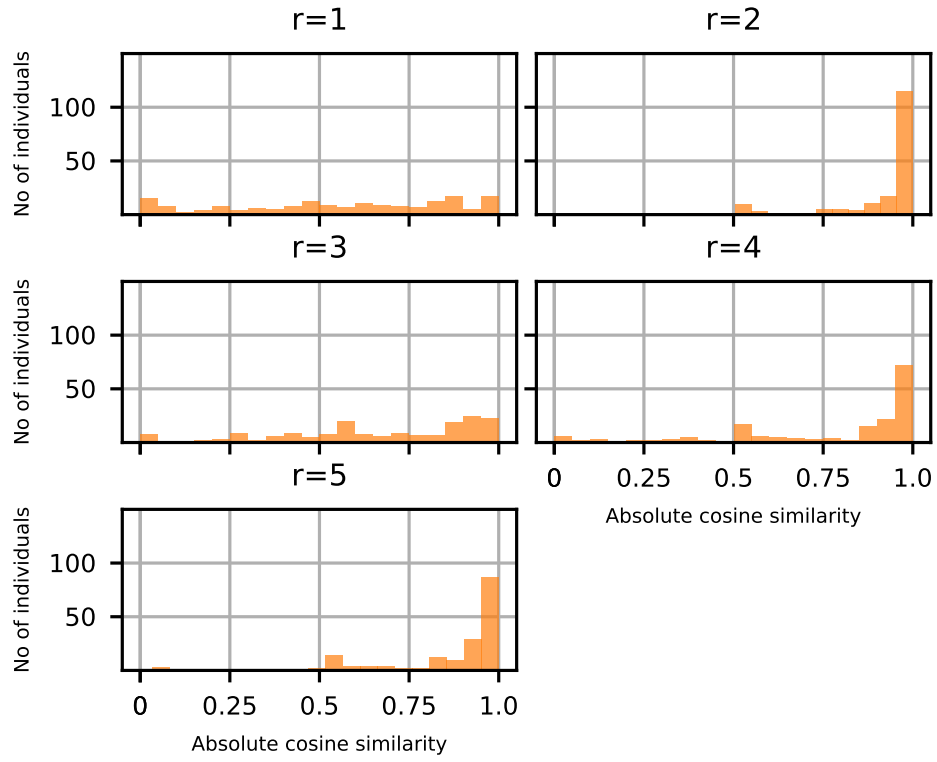
Figure S.7: Comparison of subject-specific time profiles $(c_{k,r}[\mathbf{b}_k]_r)$ captured by CMF and PARAFAC2 for each component for metabolomics, where each histogram contains 176 data points, one for each subject.