Supplementary material 'Scaling in ANOVA-simultaneous component analysis'

## 1. Separate PCA's on effect matrices equal a least-squares problem on the observed data matrix

The separate PCA's on the effect matrices can be viewed as a least-squares estimation problem, in terms of the observed data matrix **X** and sum constraints of zero on each of the component matrices. This is so because all effect matrices (i.e., $\mathbf{1}\,\mathbf{m}^{\mathrm{T}}$, $\mathbf{X}_{\alpha}$, $\mathbf{X}_{\beta}$ $\mathbf{X}_{\alpha\beta}$, **E** in Equation (2)) are mutually orthogonal, and because the PCA on the effect matrices results in component matrices that satisfy the sum constraints (Jansen et al., 2005; Timmerman, 2006). For the ASCA model involving the between and within effects, the associated ordinary least-squares (OLS) loss function boils down to

$$\underset{\mathbf{T}_{(\alpha+\alpha\beta)},\mathbf{P}_{(\alpha+\alpha\beta)},\mathbf{T}_{\mathrm{E}},\mathbf{P}_{\mathrm{E}}}{\operatorname{argmin}} \left\|\mathbf{X}-\mathbf{1}\,\mathbf{m}^{\mathrm{T}}-\ \mathbf{X}_{\beta}\ -\ \mathbf{T}_{(\alpha+\alpha\beta)}\mathbf{P}_{(\alpha+\alpha\beta)}^{\mathrm{T}}-\mathbf{T}_{\mathrm{E}}\mathbf{P}_{\mathrm{E}}^{\mathrm{T}}\ \right\|^{2}= \tag{1}$$

$$\underset{\mathbf{T}_{(\alpha+\alpha\beta)},\mathbf{P}_{(\alpha+\alpha\beta)},\mathbf{T}_{\mathrm{E}},\mathbf{P}_{\mathrm{E}}}{\operatorname{argmin}} \|\mathbf{X}-\mathbf{1}\,\mathbf{m}^{\mathrm{T}}\ \|^{2}+\left\|\mathbf{X}-\ \mathbf{X}_{\beta}\ \right\|^{2}+\left\|\mathbf{X}-\ \mathbf{T}_{(\alpha+\alpha\beta)}\mathbf{P}_{(\alpha+\alpha\beta)}^{\mathrm{T}}\ \right\|^{2}+\ \left\|\mathbf{X}-\ \mathbf{T}_{\mathrm{E}}\mathbf{P}_{\mathrm{E}}^{\mathrm{T}}\ \right\|^{2}=$$

$$\underset{\mathbf{T}_{(\alpha+\alpha\beta)},\mathbf{P}_{(\alpha+\alpha\beta)},\mathbf{T}_{\mathrm{E}},\mathbf{P}_{\mathrm{E}}}{\operatorname{argmin}} \left\|\mathbf{X}_{(\alpha+\alpha\beta)}\ -\ \mathbf{T}_{(\alpha+\alpha\beta)}\mathbf{P}_{(\alpha+\alpha\beta)}^{\mathrm{T}}\right\|^{2}+\left\|\ \mathbf{X}_{\mathrm{E}}-\mathbf{T}_{\mathrm{E}}\mathbf{P}_{\mathrm{E}}^{\mathrm{T}}\ \right\|^{2},$$

where $\mathbf{T}_{(.)}$ ($NK \times Q_{(.)}$) and $\mathbf{P}_{(.)}$ ($L \times Q_{(.)}$) denote the component score and loading matrices for effect (.), with ($\alpha + \alpha\beta$) the between effect and (E) the within effect; further there are sum constraints of zero on each effect matrix and each component score matrix (i.e., $\sum_{k=1}^{K}\mathbf{x}_{\beta,k}^{\mathrm{T}}=\mathbf{0}^{\mathrm{T}}$; $\sum_{j=1}^{J}\mathbf{x}_{(\alpha+\alpha\beta),jk}^{\mathrm{T}}=\mathbf{0}^{\mathrm{T}},\forall\, k;\ \sum_{k=1}^{K}\mathbf{x}_{(\alpha+\alpha\beta),jk}^{\mathrm{T}}=\mathbf{0}^{\mathrm{T}},\forall\, j;\ \sum_{i=1}^{I}\mathbf{x}_{\mathbf{E},jki_{j}}^{\mathrm{T}}=\mathbf{0}^{\mathrm{T}},\forall\, j,k$; $\sum_{j=1}^{J}\mathbf{t}_{(\alpha+\alpha\beta),jk}^{\mathrm{T}}=\mathbf{0}^{\mathrm{T}},\forall\, k;\ \sum_{k=1}^{K}\mathbf{t}_{(\alpha+\alpha\beta),jk}^{\mathrm{T}}=\mathbf{0}^{\mathrm{T}},\forall\, j$; $\sum_{i=1}^{I}\mathbf{t}_{\mathbf{E},jki_{j}}^{\mathrm{T}}=\mathbf{0}^{\mathrm{T}},\forall\, j,k$. The constraints on the component matrices are active in the first and second terms of Equation 5, but inactive in the third term, implying that the PCA's of the effect matrices result in component matrices that meet the required constraint.

## 2. Generation of simulated data

The experimental design pertains to 3 treatments ($j$=1,…,$J$), including a reference condition ($j$ = 1), with per condition $I$ = 20 individuals ($i$=1,…,$I$), who are measured at 10 non-equidistant time points ($k$ = 1,…, $K$), on 7 variables ($l$=1,…,$L$). The $k$ = 1,…, $K$ time points reflect $t_k$ = 0.00001, 1, 2, 4, 6, 9, 12, 16, 24, and 34 hours after intake, respectively.

The simulated scores were based on three basis functions, $C1_k$, $C2_k$ and $C3_k$, evaluated at the 10 time points $t_k$. Those functions express an early, middle and late peak, respectively, and they were generated as:

$$C1_k = k1 * (\exp(-k1 * t_{k-1}) - \exp(-k2 * t_{k-1})) / (k2 - k1),$$

$$C2_k = k1 * (\exp(-k1 * t_{k-5}) - \exp(-k2 * t_{k-5})) / (k2 - k1),$$

$$C3_k = k1 * (\exp(-k1 * t_{k-7}) - \exp(-k2 * t_{k-7})) / (k2 - k1),$$

with $t_{k-x} = 0$ for $x > k$, k1 = 50 and k2 = 100.

The simulated scores $Vl_{ijk}$ (on the $l^{th}$ variable, for individual $i$ in treatment condition $j$ at time point $k$) were generated as follows:

$$V1_{ijk} = \{abs(rn_l(0,1) + (j-1) * 4 * C1_k) .* (1+ rn_{ijk}(0,.05)\} + rn_{ijk}(0,.03333)$$

$$V2_{ijk} = \{abs(rn_l(0,1) + (j-1) * 2 * C1_k) .* (1+ rn_{ijk}(0,.1)\} + rn_{ijk}(0,.05)$$

$$V3_{ijk} = abs(rn_l(0,1) + (j-1) * 10 * C1_k + 0.05*k) * (1+ rn_{ijk}(0,.2)$$

$$V4_{ijk} = abs(rn_l(0,1) + (j-1) * 10 * C1_k + 0.15*k) + (1+ rn_{ijk}(0,.2))$$

$$V5_{ijk} = abs(rn_l(0,1)) + 0.15 * k + (j-1) * 3* C2_k + rn_{ijk}(0,.1), \text{ for } i = 1,\ldots,10$$

$$V5_{ijk} = abs(rn_l(0,1) + 0.15 * k + (j-1) * 3*C3_k + rn_{ijk}(0,.1), \text{ for } i = 11,\ldots,20$$

$$V6_{ijk} = abs(rn_l(0,1) + (j-1) * 3 * C2_k + rn_{ijk}(0,.05), \text{ for } i = 1,\ldots,10$$

$$V6_{ijk} = abs(rn_l(0,1) + (j-1) * 3*C3_k + rn_{ijk}(0,.05), \text{ for } i = 11,\ldots,20$$

$$V7_{ijk} = abs(rn_l(0,1) + rn_{ijk}(0,.1),$$

where rn.(0,1) is a randomly drawn number from $N$(0,1), and abs(.) indicates the absolute value.

## 3. Additional figures

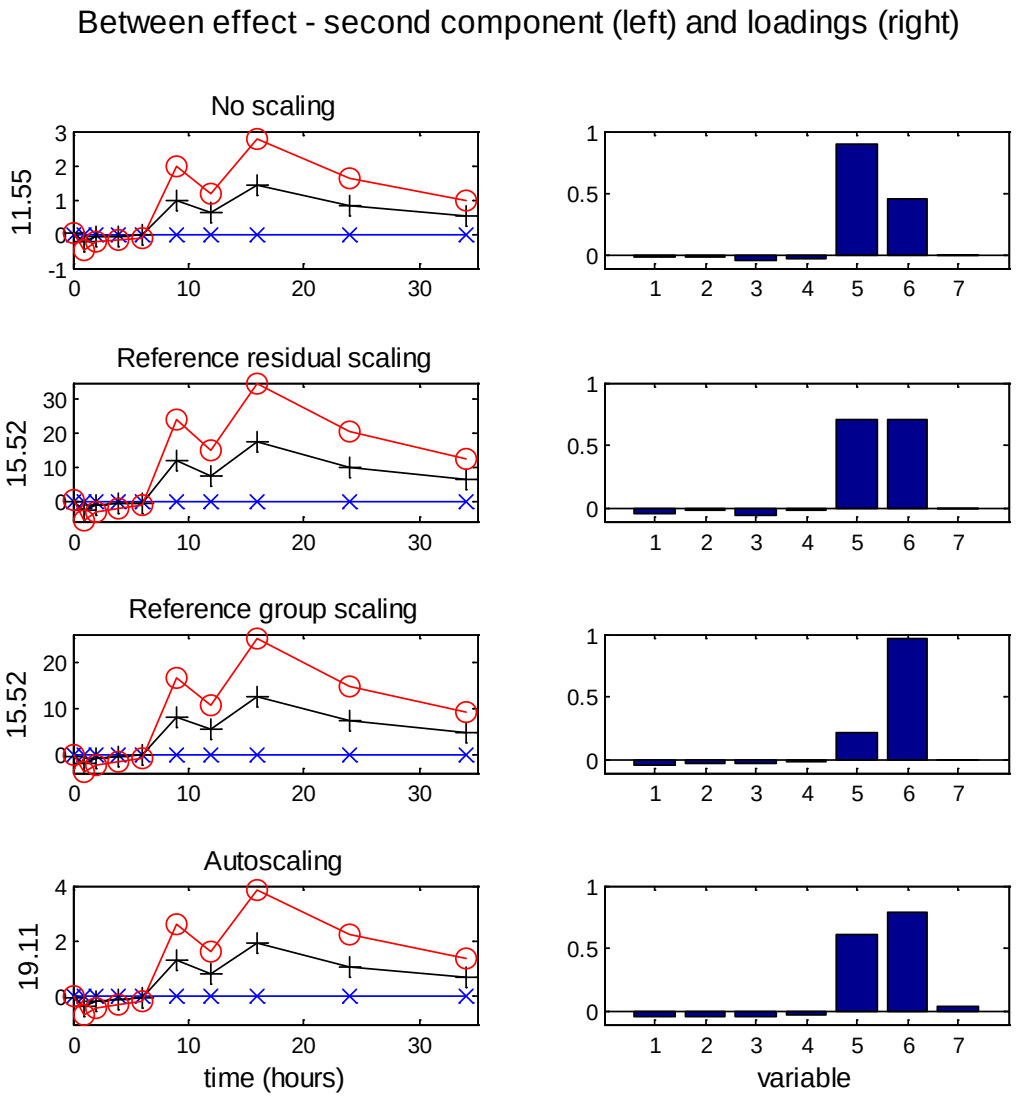


**Figure 1** Between effect after four types of scaling of simulated data. Left: scores on the second component plotted across time for each condition; right: the associated loadings for v1 to v7

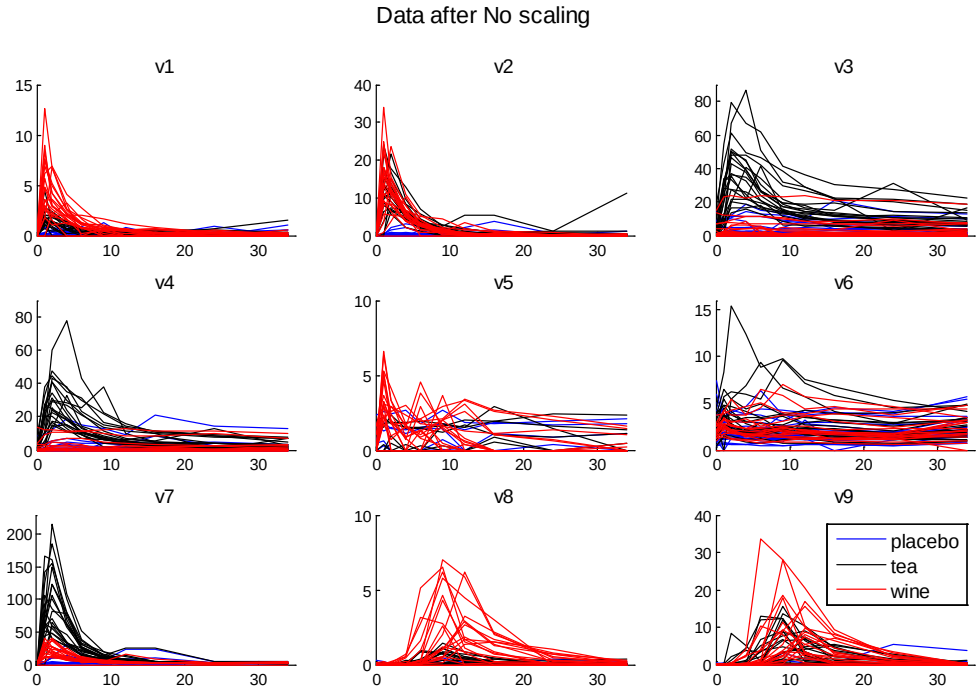


**Figure 2** Measured nutrikinetics data, of 20 individuals per treatment condition, at 8 time points, on 9 variables

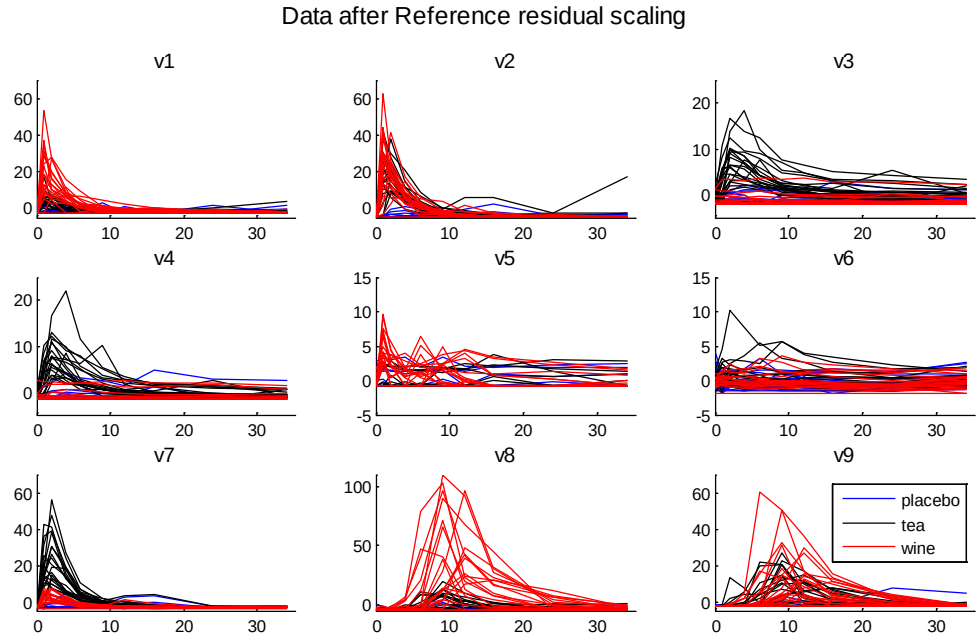


**Figure 3.** The nutrikinetics data after reference residual scaling

## 4. List of metabolites used in Empirical data example (nutrikinetics study)

The metabolites, using names according to the Pubchem ontology, and their abbreviations in the text are as follows.

| Metabolite | Abbreviation in text |
|---|---|
| (-)-Catechin | v1 |
| (-)-Epicatechin | v2 |
| (-)-Epicatechin gallate | v3 |
| (-)-Epigallocatechin gallate | v4 |
| Resveratrol | v5 |
| Isorhamnetin | v6 |
| 3/4-O-methylgallic acid | v7 |
| 5-(3′-Methoxy-4′-hydroxyphenyl)-gamma-valerolactone | v8 |
| 5-(3′,4′-Dihydroxyphenyl)-gamma-valerolactone | v9 |