Real-life metabolomics data analysis: how to deal with complex data?
Rubingh, C.M.

Citation for published version (APA):
Rubingh, C. M. (2010). Real-life metabolomics data analysis: how to deal with complex data?
DISCOVERY OF SUBTLE EFFECTS IN A HUMAN INTERVENTION TRIAL
THROUGH MULTILEVEL MODELING

Carina M Rubingh, Marjan J van Erk, Suzan Wopereis, Trinette van Vliet¹,
Elwin R Verheij, Nicole HP Cnubben, Ben van Ommen, Jan van der Greef,
Henk FJ Hendriks, Age K Smilde²

TNO Quality of Life, Business Unit Analytical Sciences, P.O. Box 360, 3700 AJ Zeist, The Netherlands

¹Currently at: CCMO, the Hague, The Netherlands
²Biosystems Data Analysis, SILS, University of Amsterdam, Amsterdam, The Netherlands

Accepted for publication in Chemometrics and Intelligent Laboratory Systems
Abstract

Many benefits can be gained if multi-factorial diseases with a high incidence and prevalence are better understood. Sophisticated approaches like multilevel analyses are needed to discover subtle differences between healthy people and people at the onset of disease in these types of studies. Multilevel analysis generates different sub-models for each level of variation. For instance, within and between subject variation can be split and analyzed separately if the two factors are orthogonal (i.e., not confounded). In the present paper, the benefits of a multilevel approach in multi-way analysis (nPLS-DA) will be described for the analysis of metabolomics data of an double blinded, randomized, parallel intervention trial with twenty slightly overweight men, whom received a diclofenac or placebo treatment for nine days. Blood samples were taken on multiple time points on 5 treatment days.

The cross-validated error rate for classifying subjects in the correct treatment group for the multilevel nPLS-DA was compared with the error rate from the ordinary nPLS-DA. 42.1% of the subjects were misclassified using ordinary nPLS-DA, whereas only 5% were misclassified using the multilevel approach. Metabolites which contributed in different ways to treatment group differences could be determined and used for biological interpretation.

The multilevel multi-way technique turned out to be a much stronger tool for modeling differences between treatment groups than the ordinary method. The metabolites that contributed most to treatment differences were not only statistically, but also biologically relevant. The multilevel approach found the effects that were better interpretable, whereas the ordinary nPLS-DA failed to do so. The methodology that was described in this paper is not only limited to human intervention studies, but can be used also for studies with a similar data structure. The multilevel approach is able to investigate effects on all levels of variation of every well designed study, hence improving the interpretability of the results.

Abbreviations

CV  Cross Validation
GC-MS  Gas Chromatography – Mass Spectroscopy
IS  Internal Standard
LC-MS  Liquid Chromatography – Mass Spectroscopy
LV  Latent Variable
nPLS  Multi-way Partial Least Squares
nPLS-DA  Multi-way Partial Least Squares Discriminant Analysis
OGTT  Oral Glucose Tolerance Test
PLS  Partial Least Squares
PLS-DA  Partial Least Squares Discriminant Analysis
Introduction

Many benefits can be gained if multifactorial diseases with a high incidence and prevalence are better understood. For instance, metabolic syndrome, cardiovascular diseases, obesity and diabetes type 2 as well as underlying factors such as insulin resistance, cause serious health problems. Cure and prevention is still difficult because the underlying causes are not completely understood. Therefore, studies are performed to obtain insight into molecular mechanisms of diseases in order to cure and/or prevent disease and hence to improve health status. For instance, nowadays it is thought that low-grade inflammatory status, often seen in overweight subjects, plays an important role in the development of insulin resistance (Hu et al., 2004; Spranger et al., 2003).

Techniques such as liquid chromatography mass spectrometry (LC-MS) and gas chromatography mass spectrometry (GC-MS) (Koek et al., 2006; Coulier et al., 2006), among others, are used to obtain system level information (van der Greef et al., 2007). The techniques, often employed as metabolomics tools, can ideally be used to detect metabolic aspects that are related to specific phenotypes of a disease. A property of these techniques is the generation of large amounts of data consisting of many correlated variables. In this huge amount of data, it is difficult to identify subtle intervention differences. Consequently, a deliberate experimental design and subsequent data analysis is needed in studies where small treatment effects can be expected. Such a design is focused on ruling out, as much as possible, all sources of variation other than those caused by the intervention. To increase the power of the study, repeated measurements within subjects over time can be taken or the study can be set up using a cross over study design. Hence, the treatment effects are estimated using changes within subjects rather than between subjects, which often show more variability. It will also limit the number of subjects that needs to be studied.

Differences within a subject (intra-individual) caused by an intervention are often smaller than the differences between subjects (inter-individual). Therefore, it will be difficult to detect small differences within a subject if total variance, being the sum of within and between subject variation, is taken into account. Basic multivariate data analysis tools like Partial Least Squares (PLS) (Martens and Naes, 1989; Geladi and Kowalski, 1986) and Partial Least Squares Discriminant Analysis (PLS-DA) (Barker and Rayens, 2003) do not distinguish inter- from intra- subject variation and thus are not ideal to be used. More sophisticated approaches like multilevel analyses are needed to take the full experimental design into account.

The basic idea of multilevel analysis is that different sub-models for each level of variation are generated, similar to analysis of variance (ANOVA). For instance, within and between subject variation can be split and analyzed separately if the two factors are orthogonal (i.e., not confounded) Multilevel data analysis has proven its value already in the field of metabolomics (Jansen et al., 2005) and psychometrics (Timmermans, 2006). Recently, the
use of a multilevel multivariate discriminant analysis of a metabolic experiment with a crossover design showed major advantages compared to the traditional data analysis approach (Van Velzen et al., 2008). In the present paper, the benefits of a multilevel approach in multi-way analysis will be described for the analysis of metabolomics data from a human intervention trial.

Methods

Study design

A human intervention trial was performed to gain more insight into the association between inflammatory status and insulin sensitivity in slightly overweight men. The study was designed to identify genes, proteins and metabolites responding to a diclofenac treatment as compared to a placebo treatment. Diclofenac, a non-steroidal anti-inflammatory compound, was chosen as anti-inflammatory model compound. A challenge test, the Oral Glucose Tolerance Test (OGTT), was used to determine changes related to glucose metabolism as a consequence of diclofenac treatment. The question of interest was to determine differences between the two treatment groups in their response to the challenge test after nine days of treatment compared to their response at baseline.

Twenty slightly overweight men (BMI range: 26.1 – 30.9 kg/m²) participated in the double blinded, randomized, parallel intervention trial. Ten subjects received a placebo treatment and ten subjects received diclofenac. One subject in the diclofenac treatment group dropped out. Blood samples were taken at day 0 and after 2, 4, 7 and 9 days of treatment. An OGTT using 75g glucose was performed on day 0 and day 9 during which blood was sampled at eight different time points, namely 0, 15, 30, 45, 60, 90, 120 and 180 minutes after the glucose intake. Metabolites were measured for each day and each time point, whereas the genes and proteins were measured at a selection of these. More details about the study design and data collection can be found elsewhere (Wopereis et al., 2009).

Data set

Four metabolomics platforms were used, namely LC-MS lipids, LC-MS free fatty acids, LC-MS polar and GC-MS global. Since the emphasis of this paper is on the analysis strategy, only the results of one of these platforms, namely the LC-MS polar data set, are presented. However, the analysis approach was applied to all platforms and results of all platforms can be found in Wopereis et al. (2009).

LC-MS polar data were corrected for the recovery of the Internal Standard (IS) for injection. Batch to batch differences were removed by synchronizing medians of quality control (QC) samples per batch. Duplicate measurements were combined into a single measurement (Bijlsma et al., 2006). When both analytical duplicates had a zero value or a non-zero value, measurements were averaged, whereas the single value was taken in case only one of the duplicates was above zero. Data were additionally cleaned up by removing glucose-related
peaks and IS-isotopes, since these could disturb the data analysis and may lead to trivial solutions. Finally, 120 peaks were included in the LC polar data set. The data set was of size $I \times (J \times K \times M)$, in which $I = 19$ subjects, $J = 120$ metabolites, $K = 8$ time points, and $M = 2$ measurement days.

**Multilevel multi-way regression**

The challenge test was used as a 'systems read-out'-parameter: the hypothesis was that the resilience of a system will be demonstrated and possibly quantified especially after stressing or perturbing a homeostatic metabolic situation. To determine differences between the two treatment groups in their response to the challenge test on day 9 compared to the day 0 response, the question of interest was stated as a multi-way regression problem. For all $I$ subjects, $J$ metabolites were measured at $K$ different time points at $M$ days. A multi-way regression problem is concerned with finding a model which predicts the value of $y$ from the data block $X$. One way of doing this is multi-way version of PLS (Martens and Naes, 1989; Geladi and Kowalski, 1986), called nPLS (Smilde et al., 2004; Smilde, 1997; Bro, 1996). In the present study, the metabolic response $X$ is related to treatment groups, hence $y$ is not a continuous parameter as in regular regression, but a dichotomous vector containing the treatment group membership. Therefore, the model is a multi-way version of PLS-DA (Barker and Ryans, 2003), called nPLS-DA.

The following model is used:

$$T = XV$$

$$X = TG(W^M \otimes W^K \otimes W^J)' + E_x$$

$$y = TB + e_y$$

$$\max \text{ cov}(t_c, y^{(c-1)}); c=1, ..., C$$

$$w^{M}_{c}, w^{K}_{c}, w^{J}_{c}$$

where $V$ is a matrix of weighing coefficients which can be written in terms of $W$, $G$ is the core array, $B$ is the regression matrix for regressing $y$ on $T$, and $E_x$ and $e_y$ are the residuals of the model for $X$ and $y$, respectively (Smilde et al., 2004).

This model can be used to relate the metabolic response to the challenge test (size $I \times J \times K \times M$) to treatment class membership (size $I \times 1$). However, this means that both inter- and intra-individual variation is taken into account. A multilevel approach (Jansen et al., 2005; Timmermans, 2006) can be used to split the variance into a between subject (inter-individual) and a within subject (intra-individual) part, hence the metabolic changes can be investigated at different levels of variation. Since the interest is in intra-individual differences specifically, the inter-individual variation can be removed by subtracting the day 0 data from the day 9 data. This can be best illustrated using a one way ANOVA model. For
simplicity reasons, an example is given to test for treatment effects over a certain number of
days at a specific time point:

\[ x_{ijk} = \mu + \alpha_i + \tau_k + \delta_j + (\tau\delta)_{kj} + (\alpha\tau)_{ik} + \varepsilon_{ijk} \]  

(2)

where \( x_{ijk} \) = measurement for subject \( i \) at day \( j \) for treatment \( k \), \( \mu \) = the overall mean, \( \alpha_i \) = effect of subject \( i \), \( \tau_k \) = effect of treatment \( k \), \( \delta_j \) = effect of day \( j \), \( (\tau\delta)_{kj} \) = treatment x day interaction, \( (\alpha\tau)_{ik} \) = subject x treatment interaction, and \( \varepsilon_{ijk} \) = residual error. If there are, for instance, measurements taken at two different days \((j = 2)\), and \( x_{i2k} \) is subtracted from \( x_{i1k} \) to test for treatment effects over the two days, all terms that are independent of \( j \) are dropped out, including the effect of each individual subject \( \alpha_i \). The model that is left is:

\[ d_{ik} = \mu_k + \varepsilon_{ik} \]  

(3)

where \( d_{ik} \) = the change in response for subject \( i \) for treatment \( k \), \( \mu_k \) = mean change in response for treatment \( k \), and \( \varepsilon_{ik} \) = residual error. This residual error takes only the changes within a subject into account.

A multilevel multi-way model was created which regresses parameter \( y \) containing the treatment group membership to the changes in metabolic response between day 0 and day 9, \( X_9 - X_0 \) (size \( l \times J \times K \)). The model was adapted as follows:

\[ T = (X_9 - X_0) \ V \]

\[ (X_9 - X_0) = TG(W^M \otimes W^K \otimes W^J)' + E_{X0-X9} \]  

(4)

\[ y = TB + e_y \]

\[ \text{max cov}(t_c,y^{(c-1)}); c=1, ..., C \]

\[ w_{c}^M, w_{c}^K, w_{c}^J \]

Note that by using \( X_9 - X_0 \) instead of \( X \) the sets of parameters \( T, V, B, W, E \) and \( e \) in (4) are not the same as in (1). Especially \( W^M \) is different as \( M \) is 1 in (4) and 2 in (1) whereas the dimension of \( T, W^K \) and \( W^J \) is the same. The model given in (4) handles variation between two time points by subtraction. However, the method can be generalized for data with more time points then two. The creation of the \( X \)-block that was used for multilevel nPLS-DA modeling is illustrated in Figure 1. First of all, a 3-way matrix \( X_0 \) of size 19 x 120 x 8 was created out of a 19 x 960 matrix. This matrix contained the metabolic data of day 0, determined at eight different time points for each subject. A matrix \( X_9 \) of the same size was also created, containing similar information for the day 9 measurements. Finally, the \( X_0 \) matrix was subtracted from the \( X_9 \) matrix and this \( X \)-block was used for data analysis. In this way an additive treatment effect will be more clear. If the treatment effect is suspected to non additive, e.g. a multiplicative change, logarithmic transformation of data prior to subtraction can be considered to improve the results.
Centering and scaling

Data \( (X_9 - X_0) \) were centered across subjects and followed by auto-scaling within metabolites. The centering step was performed to remove constants, whereas the scaling to unit variance within the metabolite mode resulted in metabolite concentrations that were relative to the variation of that metabolite. By performing the scaling step after the centering step, the prior centering remained unaffected (Smilde et al., 2004; Kiers and van Mechelen, 2001; Harshman and Lundy, 1984).

Figure 1. The creation of the X-block that was used for nPLS-DA modeling.

Model validation

To determine the optimal number of latent variables (LVs) and to validate the multilevel nPLS-DA model, a “leave-one-subject-out” cross-validation (CV) was used (Martens and Naes, 1989). In the first CV-step, data of one subject (size 1 x J x K) was left out, a multilevel nPLS-DA model was built, and the class membership of the subject who was left out was predicted. This was repeated until all 19 subjects were left out once. The error rate of the model was determined by the difference between the original class membership and the predicted one by CV. The optimal number of LVs was determined based on the minimum value of this error rate. The final fit of the model was made using the optimal number of LVs.
The nPLS-DA models were optimized by performing variable selection based on a jack-knife approach. An nPLS-DA model was made for each CV-step using data without the subject who was left out in that CV-step and using the same number of LVs that was used for the final model. This resulted in 19 sets of regression matrices of size $J \times K$, of which the standard deviation was used to determine the relative standard deviations (RSD’s) of each regression coefficient. Only those variables which had RSD of less than 100% for all time points were included in a new data set, which was used to build a second nPLS-DA model. Components that contributed to treatment differences were identified based on absolute regression coefficients of this second model (Martens and Martens, 2001).

A permutation test was performed to test whether the treatment differences were indeed true differences. One thousand dichotomous $y$ vectors were randomly created using the same proportion of zeros and ones as the vector that was used for modeling. For each random vector, a multilevel nPLS-DA model was made using the same "leave-one-subject-out" cross-validation approach and the cross-validated error rate was calculated. The same variables were used in the permutation test as were used in the corresponding nPLS-DA model. So, the permutation test for the original model contained all variables and the permutation test for the optimized model contained only those variables which had an RSD of less than 100% for all time points. A permutated null distribution was made of all thousand error rates and compared to the error rate for the original model in order to calculate significance of treatment differences.

Performance

To assess the performance of the multilevel multi-way model, also an 'ordinary' multi-way analysis was done. A 4-way nPLS-DA model (referred to as ordinary nPLS-DA in the sequel), as described in (1), was defined as the 'ordinary alternative'. The four dimensional data set (size $I \times J \times K \times M$) was used as $X$-block and the treatment class membership was used as $y$-vector. The error rate based on cross-validation for both models was compared. The error rate of the ordinary multi-way model was obtained using a same cross-validation procedure as was used for the multilevel approach. Also this 'ordinary model' was optimized based on jack-knifing the regression coefficients and a permutation test was performed.

Software

All analyses were performed using Matlab Version 7.3 2000b (The Mathworks, Inc.) and the n-way toolbox version 2.11 (Andersson and Bro, 2000).

Results and Discussion

Multilevel Multi-way analysis

A minimal cross-validated error rate of 31.5% was found for the multilevel nPLS-DA model relating the treatment group membership to the changes in metabolic response between
day 0 and day 9. Five LVs were needed for this model. The relatively high number of LVs compared to the total number of subjects illustrates the complexity of the data. As could be expected, it was not possible to describe all metabolic changes in only two or three dimensions.

The model was optimized by using a jack-knife approach. If a subject is left out and the regression coefficient changes a lot, this will result in a relatively high RSD for that particular variable. A variable with a high RSD was considered to be unstable, hence unreliable to use in explaining the differences in response between the placebo and the diclofenac group. After variable selection, a new model was made based on a subset of 31 variables. This model had a cross-validated error rate of 5% and was using 5 LVs. It appeared that variables that where most contributing to the model based on the original 120 variables were maintained after variable selection. So, essentially the same information could be described using fewer variables, illustrating the fact that many variables were unimportant for the model. The error rate of 5% meant that the treatment group membership was correctly predicted for 18 out of 19 subjects using these 31 variables. The optimized model will be used for the interpretation of the results from the multilevel multi-way models.

In Figure 2, the results of the permutation test are visualized. The vertical line represents the cross-validated error rate of the nPLS-DA model that was made, whereas the histogram represents the distribution of error rates based on permuted classes. In Figure 2a the results of the overall multilevel nPLS-DA model is given, and in Figure 2b, the results of the optimized multilevel nPLS-DA is given. The results for the overall model is very moderate (p=0.47), but the treatment differences become more clear after optimization of the model (p=0.006).

![Figure 2. Permutation test results for the original multilevel nPLS-DA model (a), the optimized multilevel nPLS-DA model (b), the original ordinary nPLS-DA model (c) and the optimized ordinary nPLS-DA model (d).](image-url)
The multi-way regression model resulted in a regression matrix of size $J^* \times K$, in which $J^*$ is the number of variables after variable selection. To determine the variables which contributed most to treatment differences, the regression coefficients were sorted by their absolute value in descending order per time point $K$. For each time point, the first ten variables were selected and used as a starting point for biological interpretation. The selected variables are presented in Table 1. The contribution of each variable to the treatment effect can be followed over time by investigating its appearance in the list of parameters that contribute most to the differences between treatments. Some metabolites were important over the whole range of time, whereas others were contributing only for a period of time. The variables which appeared in the top 10 for only one time point were initially considered to be coincidently related to the treatment. Variable 'Isoleucine + Leucine (unresolved)' (V01 in next paragraphs) and 'Glycine' (V02 in next paragraphs) will be used to illustrate further interpretation.

**Table 1.** Top10-ranking of metabolites which contributed most to treatment differences based on their absolute regression coefficient at time point $K$ (light grey shade: a metabolite that contributes to the response differences between treatments at each time point; dark grey shade: a metabolite that contributes to the response differences between treatments only for a period of time).

<table>
<thead>
<tr>
<th>Time Point (minutes)</th>
<th>ranking</th>
<th>0</th>
<th>15</th>
<th>30</th>
<th>45</th>
<th>60</th>
<th>90</th>
<th>120</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Leucine &amp; Isoleucine (not resolved)</td>
<td>Leucine &amp; Isoleucine (not resolved)</td>
<td>Leucine &amp; Isoleucine (not resolved)</td>
<td>Leucine &amp; Isoleucine (not resolved)</td>
<td>unknown 79</td>
<td>unknown 74</td>
<td>Leucine &amp; Isoleucine (not resolved)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>unknown 60</td>
<td>unknown 76</td>
<td>unknown 60</td>
<td>unknown 60</td>
<td>unknown 60</td>
<td>4-Hydroxyproline</td>
<td>2-Amino-2-methyl butanoic acid</td>
<td>4-Hydroxyproline</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>unknown 76</td>
<td>Glutamic acid</td>
<td>unknown 76</td>
<td>4-Hydroxyproline</td>
<td>unknown 61</td>
<td>unknown 100</td>
<td>4-Hydroxyproline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>unknown 61</td>
<td>4-Hydroxyproline</td>
<td>unknown 61</td>
<td>unknown 76</td>
<td>unknown 76</td>
<td>unknown 110</td>
<td>Glycine</td>
<td>unknown 76</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>unknown 74</td>
<td>unknown 60</td>
<td>4-Hydroxyproline</td>
<td>unknown 61</td>
<td>4-Hydroxyproline</td>
<td>unknown 99</td>
<td>unknown 61</td>
<td>unknown 61</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Leucine &amp; Isoleucine (not resolved)</td>
<td>unknown 74</td>
<td>5-Oxoproline</td>
<td>1-Aminocyclopentanecarboxylic acid</td>
<td>Citrate + NH4</td>
<td>4-Hydroxyproline</td>
<td>1-Aminocyclopentanecarboxylic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>unknown 78</td>
<td>5-Oxoproline</td>
<td>unknown 70</td>
<td>2-Amino-2-methyl butanoic acid</td>
<td>2-Amino-2-methyl butanoic acid</td>
<td>unknown 78</td>
<td>Hippuric acid</td>
<td>2-Amino-2-methyl butanoic acid</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4-Hydroxyproline</td>
<td>unknown 79</td>
<td>unknown 78</td>
<td>unknown 70</td>
<td>unknown 100</td>
<td>unknown 108</td>
<td>unknown 60</td>
<td>unknown 74</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>unknown 64</td>
<td>Citrate + NH4</td>
<td>unknown 108</td>
<td>unknown 74</td>
<td>unknown 74</td>
<td>1-Aminocyclopentanecarboxylic acid</td>
<td>1-Aminocyclopentanecarboxylic acid</td>
<td>Glycine</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>unknown 108</td>
<td>Aspartic acid</td>
<td>2-Amino-2-methyl butanoic acid</td>
<td>glutamic acid</td>
<td>Glycine</td>
<td>Leucine &amp; Isoleucine (not resolved)</td>
<td>unknown 64</td>
<td>5-Oxoproline</td>
<td></td>
</tr>
</tbody>
</table>
V01 is an example of a metabolite that contributes to the response differences between treatments at each measurement point, as is illustrated by the light-grey shade in Table 1. This means that the response of this metabolite between day 0 and day 9 differed during the whole time course in subjects treated with diclofenac compared to the placebo group. This effect is illustrated in Figure 3, in which the mean difference between day 9 and day 0 response for V01 is plotted per treatment group. The placebo group had at fasting state (t0) a mean change of about zero between day 9 and day 0, whereas at the same time point the diclofenac group had a mean decrease of 2.5 units. The difference between treatment groups fluctuates between 1 and 2.5 units, depending of the time point, but it remains quite stable over time. In Figure 4, the regression coefficient of this variable is plotted against the time. The same conclusion towards this metabolite can be drawn from this figure.

**Figure 3.** Mean change in metabolic response to the challenge test between day nine and day zero for subjects on placebo and diclofenac treatment, for V01 'Isoleucine + Leucine (unresolved)', a variable that contributes to treatment differences over the whole time course (error bars are based on standard errors).

**Figure 4.** Regression coefficients over time of a multilevel nPLS-DA model for V01 'Isoleucine + Leucine (unresolved)', a variable that contributes to treatment differences at each time point of the time course.
Variable V02 was only seen in the top 10 of contributing variables at 90 minutes and later of the OGTT, as illustrated by the dark-grey shade in Table 1. This metabolite was ranked 13 at t90 and therefore not included in Table 1 for this time point. The ranking at t0, t15, t30 and t45 was 21, 18, 29 and 26, respectively. So, after one hour this metabolite differed in response to the challenge test between day 9 and day 0 in subjects treated with diclofenac compared to the placebo group. This effect is illustrated in Figure 5: the differences in response are more or less the same up to 45 minutes and around zero, whereas they deviate from t60 and later. In Figure 6, the regression coefficient of this variable is plotted against the time. There is no significant contribution to treatment differences over the first 60 minutes of the curve. Only after an hour, this variable becomes more important.

![Figure 5](image1.png)

**Figure 5.** Mean change in metabolic response to the challenge test between day nine and day zero for subjects on placebo and diclofenac treatment, for V02 'Glycine', a variable that contributes to treatment differences in the second part of the time course (error bars are based on standard errors).

![Figure 6](image2.png)

**Figure 6.** Regression coefficients over time of a multilevel nPLS-DA model for V02 'Glycine', a variable that contributes only to treatment differences in the second part of the time course.
For the interpretation of the results of this type of modeling, it must be kept in mind that the regression coefficients, which were used to rank the metabolites, are based on a model in which other metabolites were also included. So, each coefficient reflects the relation between the treatment group and that particular metabolite, given the presence of the other metabolites that were used in that particular model. In Figure 4 and 6 the other metabolites are not taken into account, hence these are univariate illustrations of multivariate results.

**Multilevel approach versus ordinary nPLS-DA**

The error rate based on cross-validation for the multilevel nPLS-DA was compared with the error rate from the ordinary nPLS-DA, before and after variable selection. In total, 47.5% of the subjects were misclassified using ordinary nPLS-DA: 6 out of 10 subjects receiving placebo treatment were classified in the diclofenac group and 3 out of 9 subjects on diclofenac treatment were classified as receiving placebo treatment. The percentage of misclassified subjects is higher compared to the multilevel nPLS-DA, which had an error rate of 31.5% before variable selection. Similar results were found after variable selection. The error rate of ordinary nPLS-DA after variable selection was 42.1%, whereas this error rate was 5% for multilevel nPLS-DA. Also the results of the permutation test are worse compared to the multilevel model, which is illustrated in Figure 2. In Figure 2c the results of the overall ordinary nPLS-DA model are given, and Figure 2d shows the results of the optimized ordinary nPLS-DA. Differences between the original and the optimized model are less clear compared to the multilevel variant. Having a $p$-value of 0.72 and 0.95, for the overall ordinary nPLS-DA and the optimized ordinary nPLS-DA respectively, it is clear that no difference between treatments could be identified.

Between subject variation is often much larger than within subject variation and in the ordinary nPLS-DA both inter- and intra-individual variation are entangled. The between subject variation is too large to detect the subtle differences within a subject, resulting in a much higher error rate. The multilevel approach splits the variation into an inter- and intra-individual part and, in this particular case, focusing on the intra-individual differences only, much better results were obtained.

Also for the 4-way analysis, the regression vector provides information on the contribution of each metabolite to the discrimination between treatment groups. V02, which was of any importance only after 1 hour based on the multilevel approach, appeared also in the top of the 4-way analysis. V02 was ranked around place 5 for each time point and for both days. However, V01 did not appear in the top of important metabolites at all. For some time points, the regression coefficient for V01 was even equal to zero, meaning that it had no contribution at all to the treatment difference.
Biological validation

Diclofenac is known to inhibit and activate several enzymes and transporters among which the inhibition of the enzyme aminopeptidase N (CD13) (Boeßlerli, 2003; Ware et al., 1998). CD13 is a broad specificity aminopeptidase that cleaves specifically the N-terminal bound neutral amino acids from oligopeptides. Especially essential neutral amino acids, like L-isoleucine, L-leucine, L-methionine, L-threonine, L-phenylalanine, L-valine and L-tryptophan are expected to show lower plasma concentration in diclofenac treated subjects, whereas most of the basic, acidic and non-essential neutral plasma amino acids, among which L-glycine, are expected not to show this concentration difference. Multiple metabolic intermediates of glutathione metabolism showed time-dependent suppression in response to the oral glucose tolerance test, among which glycine, but also 5-oxoproline and glutamic acid. The glutathione synthesis pathway is insulin sensitive and the difference in response suggests that diclofenac treatment may alter insulin signaling in overweight men (for more details see Wopereis et al., 2009).

Variable V01 and V02 were identified as Isoleucine + Leucine (unresolved) and Glycine, respectively. In the multilevel approach, Isoleucine + Leucine (unresolved) was found to be of high importance for explaining differences between the two treatment groups. Glycine appeared in the top 10 only after 1 hour. In ordinary nPLS-DA, Glycine was of importance at each time point, but Isoleucine + Leucine (unresolved) was of no importance at all. Given the effect of diclofenac on CD13 and its effects on amino acids, it can be concluded that the multilevel approach found the effects that were expected, whereas the ordinary nPLS-DA failed to do so.

Multilevel nPLS-DA revealed various metabolites from the same pathway that where contributing to treatment differences, which also endores to the strength of the methodology. Findings that were found for the LC global platform were also confirmed by the GC-MS platform. Since an in-depth exploration of the biological aspects of the study are beyond scope of the present paper, these results are not presented in more detail. In Wopereis et al. (2009), the biological interpretation is discussed in full detail.

Conclusions

In many (nutritional related) -omics studies, effects on subjects are subtle and hidden in the data. For some study designs it is possible to discover these small differences by using multilevel modeling.

The multilevel multi-way technique turned out to be a much stronger tool for modeling differences between treatment groups than the ordinary method. Taking into account the multilevel structure of the data, the modeling results can be improved. By splitting the variation into an inter- and intra-individual part, it is possible to focus on different variation sources in the data. In the present study, the between subject variation was left out, so that metabolites that contributed to the subtle differences between treatments in response to
the challenge test could be identified. The multilevel approach found the effects that were better interpretable, whereas the ordinary nPLS-DA failed to do so.

The methodology that was described in this paper is not limited to human intervention studies only, but can also be used for studies with similar data structures. The multilevel approach improves the interpretability of the results by taking into account the various levels of variation in a given design.

References


