8.1 Introduction

It is difficult, or maybe not even possible, to \textit{a priori} predict which regression method is best suited for the analysis of a specific data set. A researcher simply has to rely on experience and proceed by means of a trial and error protocol. Multiway analysis methods have been used for quite some time in the field of analytical chemistry\textsuperscript{1} and psychology. Not since 1988, when Cramer \textit{et al.} introduced the CoMFA\textsuperscript{2} method in 3D QSAR, no alternative for PLS has been reported. In this thesis, however, PLS\textsuperscript{3,4} has successfully been replaced by multilinear PLS,\textsuperscript{5} for the analysis of two 3D QSAR data sets\textsuperscript{6,7} (see Chapters 5 and 6).

In the field of pharmacology and medicinal chemistry, data from \textit{in vivo} (e.g., microdialysis) and \textit{in vitro} (e.g., receptor binding) experiments are generated. The utilization of multivariate methods, e.g., PLS and PCA for the analyses of the data has been very sparse\textsuperscript{8} and are, as yet, not really accepted. A further extension of the statistical boundaries in medicinal chemistry, is the introduction of multiway analysis.\textsuperscript{9,10} In the following, two real examples (Sections 8.2 and 8.4) taken from neuropharmacology and one hypothetical example (Section 8.3) from combinatorial chemistry will be presented. The results are presented as they originally were reported, together with suggestions of how two-way and/or multiway methods could be used as an alternative.

The objective of this chapter is to demonstrate the abundance of data, in medicinal chemistry, that can be arranged in multiway matrices. Consequently, no calculations are carried out here.

8.2 Example One: Neuropharmacology with Microdialysis

In the following example, two series of experiments were performed in rats,\textsuperscript{11} with the objective to find out whether the citalopram (a selective serotonin reuptake inhibitor) induced increase in 5-HT levels, had an effect on the release of acetylcholine in the \textit{ventral Hippocampus} area. A number of drugs administrated at different dosages were injected \textit{sub cutaneously}, and for the duration of several hours, starting 60 minutes before the injection, samples were collected each 15 minutes, by means of microdialysis.\textsuperscript{12} In the first and second series of experiments, the levels of serotonin (5-HT) and acetylcholine (Ach) were monitored, respectively. It was concluded from these experiments that no significant change in acetylcholine levels were observed, as the result of the increased serotonin levels in the \textit{ventral Hippocampus} area.
Data of this type, usually are presented in two dimensional plots where the concentration levels of each drug are plotted as a function of time (Figure 8.1). For clarity, in Figure 8.1 the 15 acetylcholine samples are appended after the 15 serotonin samples, the drug concentrations (percentage of basal levels) are plotted as logarithmic values and the zero level corresponds to the basal concentration level.

![Figure 8.1](image)

**Figure 8.1** The levels of serotonin (5-HT; 1–15) and acetylcholine (Ach; 16–30) measured in the ventral Hippocampus area. The values were originally reported as percentage of basal levels, but here the logarithm of the same levels are presented. After centering of the data, the zero level corresponds to the basal level. The ID-numbers of the drugs with the most significant increase in 5-HT concentration are reported.

Figure 8.1, is a graphical representation of the two-way matrix in Figure 8.2(b), which may be decomposed by means of a Principal Component decomposition, as in Figure 8.3(b), into a score vector, \( \mathbf{t} (I \times 1) \), and a loading vector, \( \mathbf{p} (JK \times 1) \).

As an alternative, this data set could also be assembled in a three-way array, \( \mathbf{X} (I \times J \times K) \), like in Figure 8.2(a) and accordingly decomposed by means of a three-way PARAFAC decomposition (Figure 8.3(a)) into a score vector, \( \mathbf{t} (I \times 1) \), and two loading vectors, \( \mathbf{w}_J (J \times 1) \) and \( \mathbf{w}_K (K \times 1) \), corresponding to the drug, the time and the response modes, respectively.

Independent of the decomposition method used, it is to be expected that most of the insignificant variation, especially in the acetylcholine mode, will be filtered off in the first few components and, consequently, not detrimentally affect the interpretation of the models. Furthermore, when the number of drugs is large, and with several responses, it is likely that PCA or PARAFAC models become more advantageous than traditional methods, e.g., Figure 8.1, since, among other things, all experiments can be analyzed simultaneously.
Figure 8.2 (a) Graphical representation of the three-way data set, $X (I \times J \times K)$, where nine different drugs or doses $(I = 9)$ were injected and monitored for 5-HT and Ach, in the ventral Hippocampus area, for the duration of 210 minutes, with samples taken every 15 minutes $(J = 15)$. (b) The unfolded three-way matrix, i.e., the $K$ slices $(I \times J)$ of $X$ concatenated to form $X (I \times JK)$.

Figure 8.3 In (a), the one component PARAFAC decomposition of $X (I \times J \times K)$ is presented, where $w_j (J \times 1)$ and $w_K (K \times 1)$ are the loading vectors, corresponding to the time and the response modes in Figure 8.2(a), respectively. In (b), the one component Principal Component decomposition of $X$ is presented, where $w (JK \times 1)$ is the loading vector. The score vector $t (I \times 1)$, corresponds in both (a) and (b) to the mode representing the drugs.

9.3 Example two: Combinatorial Chemistry

In the field of combinatorial chemistry the following situation may occur. Imagine a number of compounds are to be synthesized by permuting all the possible combinations of three different types of building blocks (A, B and C). Typically, A could be $I$ different aromatic skeletons, B $J$ different substituents on position $R_1$ and C $K$ different substituents on position $R_2$. Since A, B and C consist of $I$, $J$ and $K$ different building blocks, respectively, $IJK$ number of compounds must be synthesized. The complete design can be comprised in the form of a cube, as in Figure 8.4. The compounds are synthesized with the help from robots and, subsequently, subjected to screening in a number of different receptor binding assays, e.g., the dopamine $D_2$, $D_3$ and $D_4$ receptors. In order to evaluate the results, the receptor affinities from the synthesized compounds are collected in a two-way matrix with $IJK$ rows and $L$ columns (assuming $L$ different receptors were considered).
Accordingly, the data may be analyzed by means of a PCA or with any other method able to handle two-way matrices. Alternatively, the cube structure of the designed compounds, in Figure 8.4, can be maintained. That is, for each receptor binding assay that the compounds are tested in, a new cube with binding results is obtained. In Figure 8.5, the data is collected in a four-way matrix $X (I \times J \times K \times L)$, which can be decomposed by means of a four-way PARAFAC\textsuperscript{15,16} or Tucker\textsuperscript{4,10,17,18} model. It is likely that compounds from the same region in the cube, are structurally associated, having affinity for about the same receptors. If the number of screened receptors is large, multivariate analysis methods (e.g., PCA) is recommended, and whether multiway methods, e.g., PARAFAC or Tucker, will improve the interpretation of the data is still to be found out.

**Figure 8.4** A possible combinatorial design $X (I \times J \times K)$ in three different modes, i.e., A, B and C.

**Figure 8.5** The matrix $X (I \times J \times K \times L)$, where the modes A, B, and C consist of I, J and K different building blocks and the response mode represents the L different receptor types.

### 8.4 Example Three: Neuropharmacology

In 1983, White and Wang\textsuperscript{19} reported on the effect of prolonged treatment with classical and atypical antipsychotic drugs on the number of spontaneously active dopamine neurons, in both the substantia nigra (A9) and the ventral tegmental area (A10). It was found that atypical antipsychotic drugs selectively decreased the number of A10 cells, while drugs with typical antipsychotic efficacy, failed to decrease the dopaminergic activity. The results from the investigation were comprised in two figures (Figures 8.6(a) and 8.6(b)). For each drug, the effect of one single injection (white bars) was compared with the effect from repeated treatments (black bars). After each experiment with repeated treatments, the effect of a single injection of apomorphine (0.063 mg/kg) was investigated (gray bars).
Figure 8.6 The effects of short and long-term treatment with various typical and atypical antipsychotics, on the number of spontaneously active A9 and A10 DA cells. Abbreviations: HAL (haloperidol), CPZ (chlorpromazine), CLZ (clozapine), TRZ (thioridazine), MOL (molindone), SUL (sulpiride) and MET (metoclapramide).

Also in this example a multiway approach can be employed by assembling the data in Figure 8.6, in a three-way array, like in Figure 8.7(a). The decomposition of $X$ is performed as in Figure 8.3(a) but here the score vector ($t$) corresponds to the drugs and the weight vectors $w_J$ and $w_K$ correspond to the type of injection and brain area, respectively. Since two modes in $X$ are very small ($J = 3$ and $K = 2$), a PCA of the unfolded three-way array, in Figure 8.7(b), is also likely to be successful.
Figure 8.7 (a) The data in Figure 8.6 assembled in a cube $\mathbf{X} (I \times J \times K)$, i.e., $J$ different types of treatments were performed with $I$ different drugs and DA neuron firing were measured in $K$ different brain areas. (b) The cube in (a), $\mathbf{X}$, is unfolded to form a two-way matrix, $\mathbf{X} (I \times JK)$.

8.5 Conclusions

In this chapter, three different examples where selected on rather arbitrary grounds, in order to demonstrate the abundance of data sets that can be arranged in multiway arrays. The objective was not to favor multiway methods before any of the present analysis methods, but to introduce the multiway methods as potential and effective analysis methods. Furthermore, there exist no clear solutions or suggestions of how to handle, for example, multiway data obtained from microdialysis where each experiment has been repeated several times. What kind of data scaling method is most effective? Is block scaling, as was used in Chapter 4 for 3D QSAR data, a suitable option for the example in Section 8.2? Clearly, before multivariate and multiway analysis methods, e.g., PCA, PLS, PARAFAC and Tucker, can be introduced and fully accepted in pharmacology several important questions need to be answered.

8.6 References

3. SYBYL- Molecular Modeling Software, 6.3, Tripos Incorporated, 1699 S. Hanley Rd, St. Louis, Missouri 63144-2913, USA,
11. Cremers T. [Unpublished data]; Department of Medicinal Chemistry; University of Groningen, The Netherlands **1997**