Multivariate statistical analysis of human monitoring in Würzburg, Germany

Bert van Bavel [a] & Günter Löffler [b]

[a] Institute of Environmental Chemistry, Umeå University, S-90187 Umeå, Sweden
[b] Department of Geography, University of Würzburg, D-97074 Würzburg, Germany

Introduction

The hypothesis that in the vicinity of Würzburg extensive cultivation, industrial production, domestic fires and traffic [1] lead to accumulation of pollution in the outside air is investigated in this study. This accumulation is, in particular, amplified by the topographical situation of the city in combination with its micro-climatic conditions. By means of human monitoring in the vicinity of Würzburg it is examined whether the organo-chlorine levels in blood data can in part be explained by pollution from outside air. The hypothesis of the existence of invasion via the outside air cannot be ruled out. Rather, the origin is the spatial variation of the pollution, which can be directly traced back to micro-climatic processes, local relief conditions, and the functional set-up of the town.

Material and Methods

Research data

The basis for this study is a group of patients from a Würzburg geophysician. The samples consisted of persons for whom the first reading of the organo-chlorine level in whole blood were measured in the period January 1992 to June 1995 and who lived in the vicinity of Würzburg. All the patients showed unspecified symptoms. In October and November 1995 a written questionnaire on personal data was made. Finally nine organo-chlorine compounds, 12 variables raised in or derived from the questionnaire could be used in the following analysis (Table 1).

Multivariate statistical methods

Multivariate data analysis consists of several methods to extract information from data matrices. The samples (objects k), in our case the different persons from the Würzburg area, are seen as points in a multidimensional space with as many dimensions as variables (i) measured on the samples. In a principal component analysis (PCA) these points are projected down onto a line, a plane or a hyperplane to give the best low-dimensional representation of the data. The data is this way approximated by a systematic part (loadings and score vectors) and a random part (the residuals). Score plots show relations between objects, in our case the monitored group of people from Würzburg, while the loading plots reveal relations between the different variables, in this study the organo-chlorine levels, personal and geographical data. The dimension of a PCA is determined by means of cross validation (Q2). In partial least squares to latent variables (PLS) a similar projection is made with the intention of predicting one or more dependent variables. Before applying PCA or PLS the data was centred and scaled to unit variance, all calculations were performed using the SIMCA S60 programme. PCA and PLS have been discussed in detail by Wold et al. [2,3].

Table 1. 21 Variables used in the analysis code and description

Organo chlorine compounds	Personal variables, questionnaire	Geographical variables, derived from the place of residence
PCB101 = PCB#101 **PCB138** = PCB#138 **PCB153** = PCB#153 **PCB180** = PCB#180 **HCB** = HCB **DDTsum** = sum of DDT+DDE+DDD **b-HCH** = beta-HCH **g-HCH** = gamma-HCH **PCP** = PCP	**Sex** =0=male, 1=female **Age** = in years **Bodex** = body-index (100*weight[g]/height[cm]^3) **hh-day** = hours spent home per day **hh-tot** = hours spent at home in total **wh-day** = working hours per day **wh-tot** = working hours in total **occupa** = occupation (ordinal scaled after possible exposure)	**altmet**= altitude of the current place of residence in meter above sea level **altlev** = place of residence in the average altitude of the intercepting layer of inversion (0=no, 1=yes) **DisleE**= distance from the industrial estate in Würzburg-East (includes a waste incinerator) **DisleS**= distance from the industrial estate in Würzburg-South

Results and discussion

A PCA model of the total data set resulted in two significant PCs according to cross validation. The statistics of the analysis are given in Table 2. Figure 1 shows the score plot and the loading plot is given in Figure 2. From the PCA of the complete data set the following conclusions can be drawn. In Fig 1 several object are located outside the main cluster.

When looking at the corresponding loadings plot (Fig. 2), it can be seen that the objects 3, 78, 218, 231, 244, 265 and 293 mainly differ in organo-chlorine compound levels (very high levels of HCB and DDTsum), while for the objects 37, 99 and 275 the cause of the difference originates in the variables, describing the occupation (wh-day is far above average). The objects 95, 118 and 276 differ mainly in hh-day.

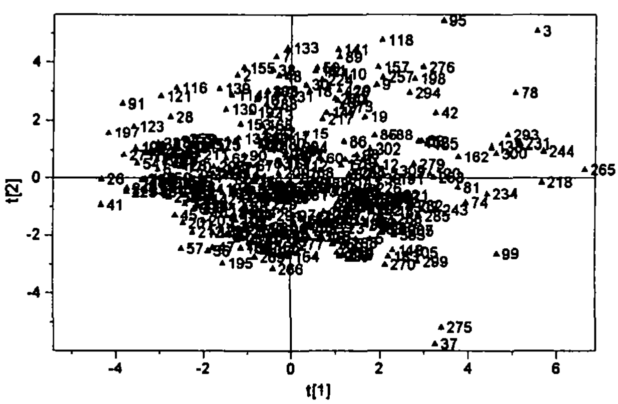

Figure 1. Score plot of the total data set PC1 versus PC2

Table 2. Model statistics PCA, including all 21 variables

	R2X	R2X(cum)	Q2	Q2 limit	Q2(cum)
Comp 1	0.21100	0.21100	0.14691	0.048544	0.14691
Comp 2	0.16135	0.37235	0.12781	0.050711	0.25594

The loading plot shows which variables are related. Two clusters of organo-halogen compounds (PCB #138, #153, #180 and DDT sum, HCB, b-HCH) are related with age. The age relation between these compounds has been described in literature [4]. PCB #101 falls outside this cluster, which is not so strange, because it is known to be metabolised at a much higher rate then the other three PCBs [5].

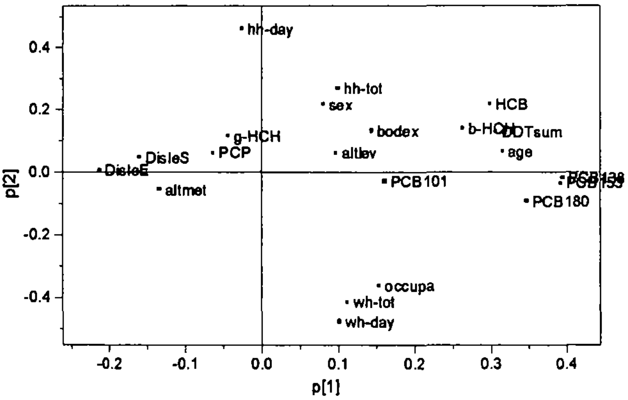

Figure 2. Loadings plot of the total data set

PCP and g-HCH are located somewhat in the middle of clusters of geographical variables (DisleE, DisleS, altmet) and personal variables (sex, bodex, hh-tot), while the variables, which described the occupation form another cluster. The significance of the different relations is further evaluated with Partial Least Squares to latent variables (PLS). PLS is used to evaluate the relation between the geographical and personal variables on the one hand and the different organo-chlorine levels in human blood for each compound on the other hand. From nine models only four models (for PCB #153, DDTsum, HCB, b-HCH) were found to be significant.

***Table 3. Statistics PLS** a) geographical/personal data and level of PCB153*
b) geographical/personal data and level of DDTsum

	R2X	R2X(cum)	R2Y	R2Y(cum)	Q2	Q2(cum)
a. PCB #153	0.20281	0.20281	0.13387	0.13387	0.0993	0.0993
b. DDTsum	0.18136	0.18136	0.19736	0.19736	0.17179	0.17179

In Table 3 the statistics of the multivariate relation of the geographical and personal data and the level of PCB153 as well as the level of the DDTsum in blood data are given. The PLS relation is significant according to cross validation and uses 18.1% (DDTsum) and 20.3% (PCB153) of the variance in the X-block data (12 geographical or personal variables) to explain 19.7% of the Variance in the Y-block data (DDTsum) and 13.4% (PCB #153). In Figure 3 so-called VIP plots are displayed. These plots reveal which variables are most influential in the models. A value over 1 indicates that the variable has a significant contribution to the model. In the model for PCB153 in addition to age (share of explanation: 30%) these are the geographical variables DisleE (13%) and altmet (10%). The impact of this is in general: The level is higher for the older the patients, the lower the altitude and the closer the proximity of the place of residence to the industrial estates Würzburg East.

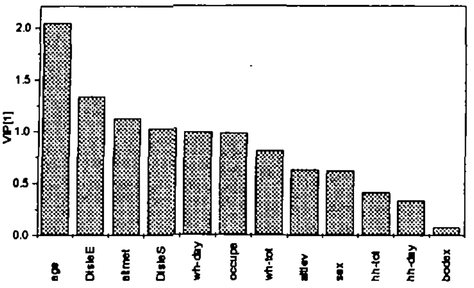

Figure 3a. VIP plot for PCB #153, indicating the most important variables in the PLS model

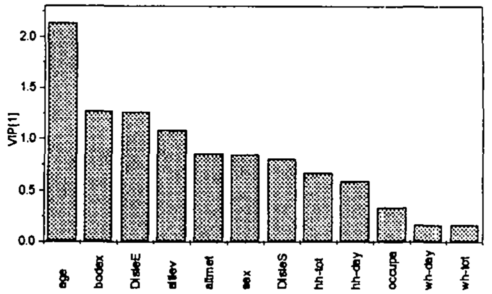

Figure 3b. VIP plot for sum of DDTs, indicating the most important variables in the PLS model

In the model for the DDTsum in addition to age (share of explanation: 40%) and bodex (13%), which is well known in literature [6], the geographical variables DisIeE (13%) and altlev (10%) are important variables in the PLS model. The impact of this is in general: The level is higher the older the patients, the higher the amount of fatty tissue, the closer the proximity of the place of residence to the industrial estates Würzburg and when living in the average altitude of the intercepting layer of inversion weather situations.

While the impact of the inversion weather situation is obvious, the relation with the proximity to the industrial estate Würzburg-East cannot be explained in detail here. Neither PCB #153 nor DDT, DDE or DDD respectively are expected to be emitted from the waste incinerator. Other local sources are expected to exist. Beside some industrial enterprises this may be uncontrolled incineration of the hospital waste, on-going until the early 90's, or leakage from equipment from military bases in this area.

The multivariate data evaluation showed that part of the variance (10-20%) in the level of some organo-chlorines (DDTsum, PCB #153, b-HCH) in blood data can be explained by pollution from outside air [7]. The whole amount of pollution in the vicinity of Würzburg caused by various emitters is not altogether necessarily higher than other cities. The special topographical conditions and the micro-climatic situations caused by the frequent inversion weather pattern pollution in the Würzburg area would lead to a distinct accumulation of the pollutants in or below the intercepting layer.

References

1. NRW-Meßprogramm 'Chloraromaten- Herkunft und Transfer'. Min. für U. R.L., Düsseldorf (1991).
2. S. Wold et al. In Martens and Russbaum (Eds). Food Research and Data Analysis. Applied Science.1983 pp147-188.
3. S. Wold et al. Environ. Health Pers. 61 (1995) 291-338.
4. L. Rylander et al. Sci. Total. Environ. (1997) 207 55-61.
5. B. van Bavel et al. Environ. Toxicol. and Chem., 15, 6: pp.947-954 (1996).
6. H. Marquardt and S. Schäfer (eds.), Lehrbuch der Toxokologie (BI-Wiss.-Verl., Mannheim 1994).
7. G. Löffler, Submitted to Toxicological and Environmental Chemistry (february 1998).