

## Indicator parameters for PCDD/PCDF from electric arc furnaces

Tomas Öberg (tomas.oberg@hik.se)

Kalmar University, Department of Biology and Environmental Science, SE-391 82 Kalmar,  
Sweden

### Abstract

The unintentional formation and release of persistent organic pollutants (POP) from industrial sources is of environmental concern and efforts are now made to reduce these emissions [1].

The emissions of chlorinated trace organics from electric arc furnaces (EAF) have been monitored on a regular basis in Sweden since the 1980s. Most analyses have encompassed not only polychlorinated dibenzo-*p*-dioxins (PCDD) and dibenzofurans (PCDF), but also chlorinated benzenes and phenols. Emissions of 2,3,7,8-substituted PCDD/PCDF from municipal solid waste incinerators (MSWI) can be modelled and predicted from analyses of chlorinated benzenes and phenols, which are suspected to be precursors in the formation process. The purpose of this investigation was to extend and update previously reported models with new samples from EAF, to describe the main sources of variation, and to compare multivariate calibration with univariate regression. The measurement data consisted of 27 samples collected between 1987 and 2002 and analysed by two different laboratories. A general multivariate calibration model was able to describe 97% of the variation in the TEQ (toxic equivalent quantity) value over five orders of magnitude. Univariate regression models cannot account for changes in the congener pattern and thus gave a poorer performance. In plant-specific applications the univariate approach did, however perform equally well. It was

therefore concluded that both multivariate and univariate regression models can be used in process optimisation studies, but that multivariate models are better suited for emission monitoring and evaluation of removal efficiencies in the off-gas cleaning systems.

*Keywords: Partial least squares; PLS; Dioxins; Chlorobenzenes; Chlorophenols; Surrogates*

## **1. Introduction**

Polychlorinated dibenzo-*p*-dioxins (PCDD) and dibenzofurans (PCDF) are persistent organic pollutants that resist degradation and may accumulate in terrestrial and aquatic ecosystems. PCDD/PCDF are unintentionally formed and released from combustion and other thermal processes involving organic matter and chlorine [2]. Thermal processes in the metallurgical industry are one such industrial source, with the first report on PCDD/PCDF formation and release from a wire reclamation operation appearing more than 20 years ago [3].

Sampling and analyses of micropollutants in off-gases increase in complexity and cost as the detection limit is lowered: this applies particularly to ultratrace components such as PCDD/PCDF. The development of indirect measurement methods began in the 1980s. It was then shown that emissions of PCDD/PCDF could be modelled and predicted from analyses of chlorinated benzenes and phenols [4-6]. A substantial number of papers have been published since then, though most of the work has focused on emissions from municipal waste incinerators [7-13]. In Europe, however, emissions from waste combustion plants are of less importance today. Instead, much interest and attention is now directed towards other industrial sources, in particular the metallurgical industry. Scrap remelting is one example, and a recent environmental strategy paper by the European Commission suggested that

electric arc furnaces are the only industrial sources with constant or increasing emissions to air [14].

Emissions from electric arc furnaces (EAF) have been monitored on a regular basis in Sweden since the 1980s, and most analyses have encompassed not only PCDD/PCDF but also chlorinated benzenes and phenols [15-18]. These samples from the metallurgical industry have been included in both general and plant-specific multivariate calibration models [6, 19]. The purpose of the current investigation was to extend and update these previously reported models with eleven new samples from EAF (1993-2002), to describe the main sources of variation, and to compare multivariate calibration with univariate regression. Furthermore, from the outcome it should also be possible to give recommendations for the practical use of these models in process optimisation, control of off-gas cleaning systems and monitoring of emission levels.

## **2. Experimental**

### 2.1 Measurement data

The data set evaluated in this investigation consisted of 23 off-gas samples collected between 1987 and 2002 from electric arc furnaces at four Swedish steel mills. High alloy steels are manufactured in two of these steel mills and standard merchant steel in the two others. Four additional samples were included from a study of adsorbents in a pilot scale electric arc furnace at the Metallurgical Research Institute in Luleå, Sweden. The EAF process in these plants is batch-wise and each scrap heat is charged separately. The available samples were unevenly distributed among plants and type of samples (Table 1). Samples from both raw gas,

clean gas, and after adsorbent injection were included to span the whole domain of possible off-gas compositions.

*Table 1*

Recycled steel scrap is the primary raw material in these steel plants and was also used in the pilot scale trials. The metal composition of the steel scrap is carefully monitored, but the scrap also contains a varying degree of organic impurities such as oil, paint, and plastic coatings. It has been shown that the composition and purity of the raw material is one important factor influencing the formation and release of chlorinated aromatic compounds from scrap remelting processes [16].

Three different consultant companies collected the samples using an all-glass sampling train. PCDD/PCDF, chlorinated benzenes and phenols were analysed in 21 samples by ALcontrol AB (previously Miljökonserterna) in Nyköping, Sweden. Six samples were analysed for the same components by Umeå University in Sweden. The analytical methods used at the two laboratories have been described elsewhere [20, 21]. The sampling and analysis of PCDD/PCDF followed the recommended Nordic method before 1997 (18 samples) and the European standard EN 1948 thereafter [22, 23].

Censored data, with substantial portions of the data below the limit of detection (LOD), are common in environmental applications [24]. Kaune and Kettrup discussed this problem in correlation analysis of PCDD/PCDF and recommended not including data below the LOD [25]. In our previous reports on indicator parameters, values below the LOD have been treated either as missing or assigned a value of half the detection limit. It can be argued that both these approaches are unbiased, but in trace organic analysis the LOD is sometimes close to the reported measurement values. In this investigation the LOD values in one sample were often higher than the reported positive identification for the same variable in another sample,

and assigning a value of half the detection limit could then add noise, hide data structures and weaken correlations. Data below the LOD were therefore treated as missing values.

The chemical analytical variables and measurement results are summarised in Table 2. All data are reported as the amounts in the sample. The number of missing values (below LOD) is generally low. The toxic equivalent quantity (TEQ) was calculated using the WHO weighting scheme [26]. The sums of the homologous groups of PCDD/PCDF are not included in this investigation since only the 2,3,7,8-substituted congeners are required by the European standard. We have shown in a previous study that the 2,3,7,8-substituted congeners can be predicted with higher precision than the sums of the homologous groups, presumably due to better analytical accuracy and precision [19]. The gas volumes of the samples varied between 2-24 m<sup>3</sup> dry gas at 101.3 kPa and 0 °C.

#### *Table 2*

Skewed frequency distributions are often encountered in environmental pollution monitoring [24, 27], and this data set is no exception. However, a symmetric distribution is easily obtained by a logarithmic transformation as seen from the Box-Whisker plots, Fig. 1.

#### *Fig. 1*

### 2.3 Data analysis

The data analysis was carried out using the software Unscrambler 7.6 SR-1 (CAMO ASA, Oslo, Norway). Multivariate calibration models were developed using partial least squares regression for either a block of dependent Y-variables (PLSR2) or a single Y-variable (PLSR1). Univariate regression models are included for comparison. Applied regression

analysis and multivariate calibration using PLSR have been extensively described both in textbooks and in a recent paper in this journal [28-30].

Prediction testing using an external test set is often recommended as a severe test of model validity. It is then important that the test samples cover the same range as the intended future use of the model, but this is difficult to accomplish with small data sets. Results from two recent studies suggest that cross-validation is to be the preferred method when the sample size is small [31, 32]. The difficulty in selecting a relevant test set will of course become more pronounced if the available data are inhomogeneous, as here with an unevenly distributed population of samples. Cross-validation was therefore selected as the currently best alternative to evaluate the model performance.

All variables were log-transformed and auto-scaled to zero mean and unit variance. The significance of each model was evaluated with full cross-validation (leave one out), both to establish the rank of the calibration model and to estimate the prediction error.

### **3. Results and discussion**

The predictor variables (X) consisted of 22 congeners and groups of chlorinated benzenes and phenols (nos. 1-22 in Table 2). The dependent variables (Y) were the 17 toxic congeners of PCDD/PCDF and TEQ (nos. 23-40 in Table 2).

#### **3.1 A general PLSR2 model**

Initially, the overall correlation between X and Y was investigated using the PLSR2 method. A model with two significant components explains 68% of the variance in Y and 88% of the

variance in X. The correlation between the X- and Y-blocks is shown in a plot of the first score vectors, Fig. 2. The samples from plant A effectively span and dominated the regression model. This is no surprise since these samples dominate in number and vary substantially in composition. The two points that fall distinctly below the regression line are samples from tests with adsorbent injection in plant E. One could easily dismiss these as outliers, but removing these samples would only result in new “lonely samples” since the data set is small in number and heterogeneous in composition.

*Fig. 2*

The contribution of the original variables to the overall regression can be visualised in a plot of the loading weights for the X-block and loadings for the Y-block, Fig. 3. The X- and most of the Y-variables line up with increasing degree of chlorination, and hence also molecular weight and volatility. This correlation pattern has been reported previously, and it is a prerequisite for the success of a multivariate calibration approach to this task [5, 6].

*Fig. 3*

### 3.2 A specific PLSR1 model

A PLSR1 model often has better predictive ability than a PLSR2 model, and nonlinear structures can sometimes be modelled by including additional factors. Here it is also warranted to focus on one of the dependent variables, namely WHO-TEQ, because it is used for compliance reporting and expresses the environmental significance. The TEQ value is a weighted average, where the weights are assigned according to the toxic properties of each of the 2,3,7,8-substituted PCDD/PCDF congeners.

The optimal rank of the PLSR1 model was estimated to seven components, corresponding to the first maximum in the explained validation variance ( $Q^2_Y$ ), Fig. 4. A more conservative approach would be to select only 4-5 components. A model with five components explained 96% of the variance in the calibration data ( $R^2_Y = 0.957$ ) for the TEQ value, and 79% of the variance in cross-validation segments ( $Q^2_Y = 0.790$ ). The same model also describe a large portion of the variance in the predictor variables, i.e. chlorinated benzenes and phenols ( $R^2_X = 0.933$  and  $Q^2_X = 0.870$ ).

*Fig. 4*

This model provides an adequate description of variations in the WHO-TEQ value over almost five orders of magnitude. The prediction results are shown for the calibration data in order to facilitate comparison with the univariate approaches described elsewhere, Fig. 5.

*Fig. 5*

The loading weights for the X-block predictor variables in the first two components were almost exactly the same as in the PLSR2 model, Fig. 6. The two first components describe both variations in amounts and in congener pattern, i.e. degree of chlorination. That these main relationships are stable is thus an indication that both these aspects are important in modelling PCDD/PCDF from chlorinated benzenes and phenols.

*Fig. 6*

### 3.3 Comparison with other regression approaches

It has been suggested that PLS regression is an unnecessary complicating step in using chlorinated benzenes and phenols as indicator parameters for PCDD/PCDF [7, 33]. Many

studies have instead focused on finding univariate regression relationships to specific congeners, groups or the sum of chlorinated benzenes or phenols.

The present data set provides an opportunity to compare these different approaches. The coefficients of determination,  $R^2$  and  $Q^2$  (cross-validated  $R^2$ ) are listed in Table 3 for various homologous groups and the sums of chlorinated benzenes and phenols.

*Table 3*

Hexachlorobenzene was the first variable to be suggested as an indicator parameter for PCDD/PCDF [4]. It is one of the best univariate relationships for this data set, but is clearly outperformed by the PLSR1 model shown previously. The variation between plants and processes is due to variations in halogen input and removal efficiency in off-gas cleaning, and both of these factors affect congener patterns of all chlorinated aromatics in the off-gas. Univariate models are therefore less suitable under such circumstances.

We have found univariate approaches useful in process studies that are performed within a specific plant [16, 34, 35]. Ten of the clean-gas samples in this investigation came from the same steel mill (plant A), and here the correlation between PCDD/PCDF and the sum of chlorinated benzenes was much better, Fig. 7.

*Fig. 7*

A PLSR1 model with one significant component has the same performance, with  $R^2_Y = 0.954$  compared to 0.958 and  $Q^2_Y = 0.923$  compared to 0.930.

#### 4. Conclusions

The results of this investigation show that the amounts and distribution of various chlorinated aromatic compounds in the off-gases from electric arc furnaces vary substantially. The correlation patterns were relatively stable and it is possible to predict the amounts and composition of 2,3,7,8-substituted congeners of PCDD/PCDF from the amounts and congener composition of chlorinated benzenes and phenols. A multivariate calibration model provided an adequate description of variations in the WHO-TEQ value over almost five orders of magnitude. Comparison with univariate regression showed that models thus simplified do not give an adequate description of the whole data set.

Changes in the composition of contaminated metal scrap or in the removal efficiency of off-gas cleaning can change the congener pattern of the unintentionally formed chlorinated aromatics. Univariate regression cannot account for variations in the congener pattern and is thus likely to fail in situations where samples are from different processes or plants. In contrast, measurement data from a single plant showed that univariate regression within such a limited domain could perform equally well as multivariate calibration. This is also the likely background to why so much research effort has focused on this approach, and indeed most of the cited references in a recent review are of studies performed within single plants [13].

Chlorinated benzenes and phenols are easier to analyse with high precision at a low cost. The presented correlation patterns can form the basis for indirect measurements of PCDD/PCDF with an indicator parameter approach. Analyses of both chlorinated benzenes and phenols were needed in order to model PCDD/PCDF over the whole of the investigated domain. It would probably be sufficient to use only one of these compound groups for plants with similar operating conditions and off-gas cleaning. Use of indicator parameters for emission monitoring and evaluation of removal efficiencies in off-gas cleaning are best performed with

multivariate calibration models. In process optimisation studies not involving the off-gas cleaning system, univariate regression models usually perform equally well.

## 5. Acknowledgements

Support from the Swedish Steel Producers' Association (Jernkontoret) is gratefully acknowledged.

## 6. References

- [1] The Stockholm Convention on persistent organic pollutants; United Nations Environment Programme: Geneva, 2001.
- [2] Bumb, R. R.; Crummett, W. B.; Cutie, S. S.; Gledhill, J. R.; Hummel, R. H.; Kagel, R. O.; Lamparski, L. L.; Luoma, E. V.; Miller, D. L.; Nestruck, T. J.; Shadoff, L. A.; Stahl, R. H.; Woods, J. S. *Science* 1980, 210, 385-390.
- [3] Hryhorczuk, D. O.; Withrow, W. A.; Hesse, C. S.; Beasley, V. R. *Archives of Environmental Health* 1981, 36, 228-234.
- [4] Öberg, T.; Bergström, J. *Chemosphere* 1985, 14, 1081-1086.
- [5] Öberg, T.; Bergström, J. *Chemosphere* 1987, 16, 1221-1230.
- [6] Öberg, T.; Bergström, J. *Chemosphere* 1989, 19, 337-344.
- [7] Kaune, A.; Lenoir, D.; Nikolai, U.; Kettrup, A. *Chemosphere* 1994, 29, 2083-2096.

- [8] Kaune, A.; Lenoir, D.; Schramm, K. W.; Zimmermann, R.; Kettrup, A.; Jaeger, K.; Ruckel, H. G.; Frank, F. *Environmental Engineering Science* 1998, 15, 85-95.
- [9] Zimmermann, R.; Heger, H. J.; Blumenstock, M.; Dorfner, R.; Schramm, K. W.; Boesl, U.; Kettrup, A. *Rapid Communications in Mass Spectrometry* 1999, 13, 307-314.
- [10] Fiedler, H.; Lau, C.; Eduljee, G. *Waste Management & Research* 2000, 18, 283-292.
- [11] Blumenstock, M.; Zimmermann, R.; Schramm, K. W.; Kettrup, A. *Chemosphere* 2001, 42, 507-518.
- [12] Kato, M.; Urano, K. *Waste Management* 2001, 21, 63-68.
- [13] Öberg, T.; Neuer-Etscheidt, K.; Nordsieck, H.; Zimmermann, R. *Organohalogen Compounds* 2002, 59, 37-44.
- [14] Communication from the Commission to the Council, the European Parliament and the Economic and Social Committee: Community strategy for dioxins, furans and polychlorinated biphenyls; *Official Journal of the European Communities* 2001, C322, 2-18.
- [15] Öberg, T.; Bergström, J. *Organic micro-pollutants from steel mills (In Swedish); report D621, Jernkontoret: Stockholm, Sweden, 1988.*
- [16] Öberg, T.; Allhammar, G. *Chemosphere* 1989, 19, 711-716.
- [17] Öberg, T. *Formation of dioxin – Literature review and evaluation of measurement data (In Swedish); report D673, Jernkontoret: Stockholm, Sweden, 1992.*

- [18] Öberg, T. Emissions to air of environmental organic pollutants from electric arc furnaces: Occurrence and possible measures to minimise the environmental impact (In Swedish); report D793, Jernkontoret: Stockholm, Sweden, 2003.
- [19] Öberg, T.; Bergström, J. *Organohalogen Compounds* 1992, 8, 197-200.
- [20] Öberg, T.; Warman, K.; Bergström, J. *Chemosphere* 1987, 16, 2451-2465.
- [21] van Bavel, B.; Fängmark, I.; Marklund, S.; Söderström, G.; Ljung, K.; Rappe, C. *Organohalogen Compounds* 1992, 8, 225-228.
- [22] Jansson, B.; Bergvall, G. *Waste Management & Research* 1987, 5, 251-255.
- [23] Stationary source emissions - Determination of the mass concentration of PCDDs/PCDFs EN 1948:1-3, European Committee for Standardisation: Brussels, 1997.
- [24] Gilbert, R. O. *Statistical methods for environmental pollution monitoring*; Van Nostrand Reinhold Co.: New York, 1987.
- [25] Kaune, A.; Kettrup, A. *Chemosphere* 1994, 29, 1811-1818.
- [26] van den Berg, M.; Birnbaum, L.; Bosveld, A. T. C.; Brunström, B.; et al.. *Environmental Health Perspectives* 1998, 106, 775-792.
- [27] Dean, R. B. In *Chemistry in water reuse*; Cooper, W. J., Ed.; Ann Arbor Science: Ann Arbor, 1981; Vol. 1, pp 245-258.
- [28] Draper, N. R.; Smith, H. *Applied regression analysis*, 2d ed.; Wiley: New York, 1981.
- [29] Martens, H.; Næs, T. *Multivariate calibration*; Wiley: Chichester, 1989.

- [30] Wold, S.; Sjöström, M.; Eriksson, L. *Chemometrics and Intelligent Laboratory Systems* 2001, 58, 109-130.
- [31] Martens, H.; Dardenne, P. *Chemometrics and Intelligent Laboratory Systems* 1998, 44, 99-121.
- [32] Hawkins, D. M.; Basak, S. C.; Mills, D. *Journal of Chemical Information and Computer Sciences* 2003, 43, 579-586.
- [33] Kaune, A.; Lenoir, D.; Nikolai, U. *Staub Reinhaltung der Luft* 1994, 54, 91-94.
- [34] Öberg, T. *Journal of Chemometrics* 2003, 17, 1-5.
- [35] Öberg, T.; Öhrström, T. *Environmental Science & Technology* 2003, 37, 3995-4000.

Table 1

Type of plant and number of off-gas samples.

Plant	Number of samples	
	Raw gas	Clean gas*
A: Merchant steel	5	10
B: Merchant steel	0	3
C: High alloy steel	0	3
D: High alloy steel	0	2
E: Pilot plant	2	2

\* After the fabric filter.

Table 2

Numbering and identification of variables, together with the number of measurement values, medians and ranges of the chemical variables. Chlorinated benzenes, phenols and PCDD/PCDF are denoted as B, P, D and F respectively. Values below the detection limit are treated as missing.

No	ID	n	Median ( $\mu\text{g}$ )	Min ( $\mu\text{g}$ )	Max ( $\mu\text{g}$ )	No	ID	n	Median (ng)	Min (ng)	Max (ng)
1	13B	27	7.2	0.57	180	23	2378D	15	0.43	0.077	22
2	14B	27	4.1	0.35	70	24	2378F	27	6.7	0.044	470
3	12B	27	14	1.3	950	25	12378D	22	0.90	0.038	66
4	135B	26	1.5	0.14	20	26	12378F	27	3.5	0.022	570
5	123B	26	9.8	0.9	220	27	23478F	27	5.4	0.024	1000
6	124B	27	7.9	1.2	470	28	123478D	24	0.40	0.012	17
7	1235/1245B	27	4.1	0.24	65	29	123678D	26	0.58	0.0091	17
8	1234B	27	3.4	0.23	66	30	123789D	25	1.2	0.023	130
9	P5CB	27	2.8	0.27	48	31	123478F	27	3.2	0.012	1200
10	H6CB	27	0.65	0.062	13	32	123678F	27	3.6	0.043	1100
11	24/25P	26	3.3	0.11	40	33	234678F	27	1.2	0.021	280
12	23P	22	0.40	0.052	12	34	123789F	26	0.95	0.0099	300
13	26P	26	0.73	0.06	17	35	1234678D	22	1.1	0.030	610
14	35P	17	0.15	0.018	4.9	36	1234678F	23	1.9	0.051	1800
15	34P	24	0.39	0.029	6.4	37	1234789F	22	0.22	0.0061	300
16	235P	23	0.14	0.015	3.8	38	OCDD	26	1.4	0.15	1000
17	246P	26	2.0	0.079	22	39	OCDF	26	0.78	0.023	690
18	245P	26	0.42	0.024	6.0	40	TEQ*	27	5.0	0.034	1000
19	234P	24	0.42	0.015	5.1						
20	236P	26	0.15	0.007	6.1						
21	T4P (sum tetra)	23	1.6	0.026	17						
22	P5CP	24	0.24	0.019	11						

Table 3

Univariate linear relationships between WHO-TEQ and some suggested indicator parameters (log-transformed data).

Parameter	n	R <sup>2</sup>	Q <sup>2</sup>	Parameter	n	R <sup>2</sup>	Q <sup>2</sup>
DCB	27	0.203	0.057	DCP	26	0.423	0.341
T3CB	27	0.343	0.260	T3CP	27	0.447	0.360
T4CB	27	0.398	0.317	T4CP	23	0.507	0.404
P5CB	27	0.456	0.366	P5CP	24	0.547	0.462
H6CB	27	0.537	0.456	Sum CP	27	0.415	0.308
Sum CB	27	0.264	0.142				



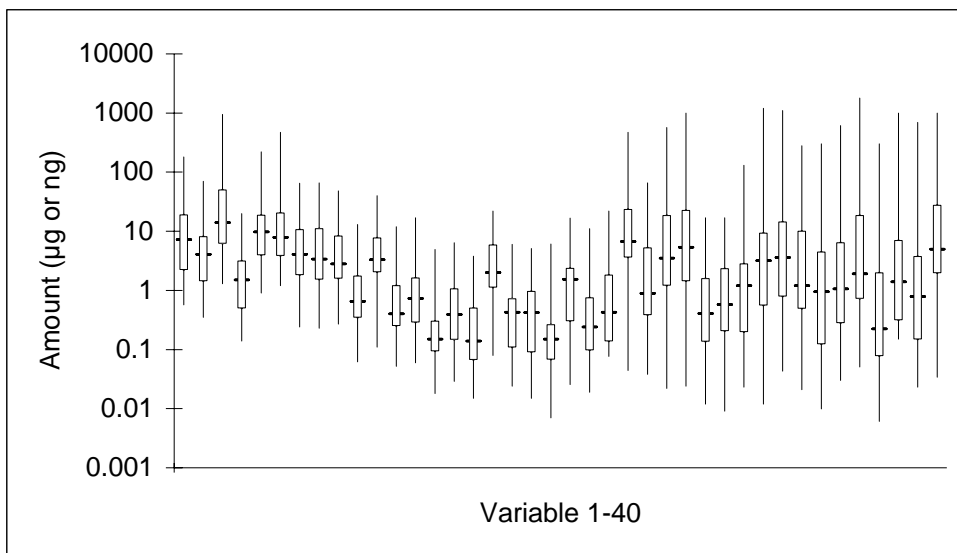


Fig. 1. Box-Whisker plots of chlorinated benzenes and phenols ( $\mu\text{g}$ ) and PCDD/PCDF (ng). Minimum, 1<sup>st</sup> quartile, median, 3<sup>rd</sup> quartile and maximum values in order of appearance, logarithmic scale.

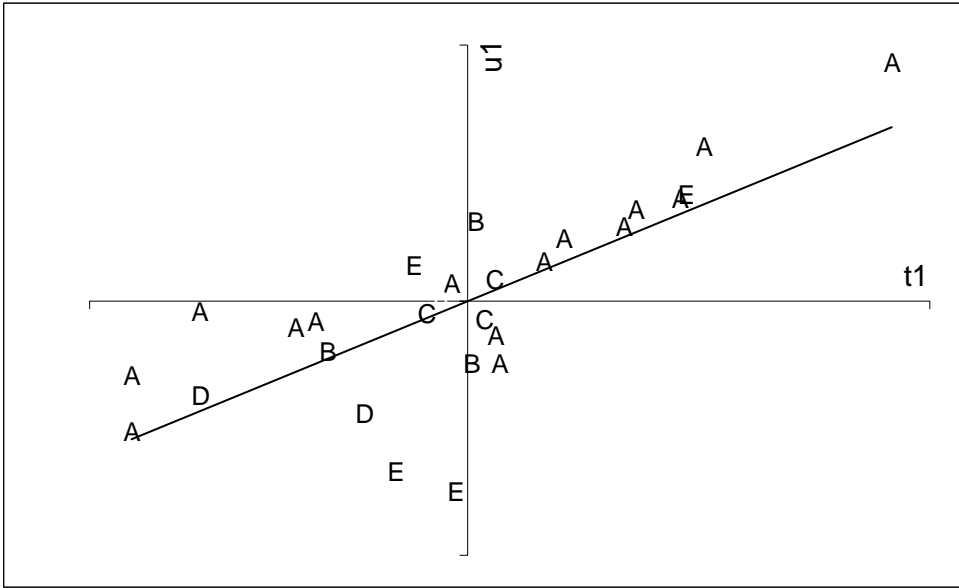


Fig. 2. The first score vector for Y ( $u_1$ ) vs the first score vector for X ( $t_1$ ), PLSR2 model. Samples denoted with letters from Table 1.

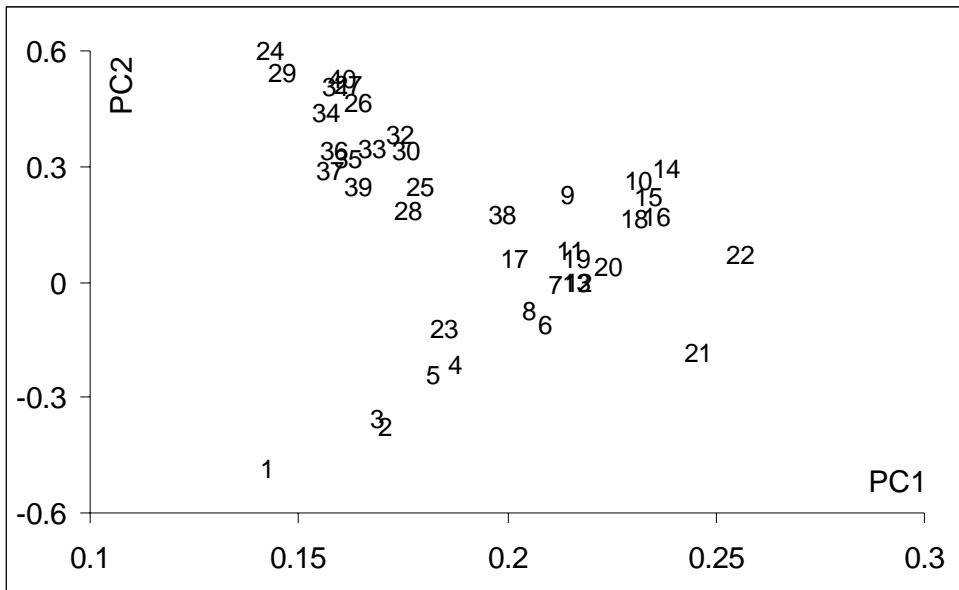


Fig. 3. Loading weights and Y-loadings for the first two components, PLSR2 model. Variables denoted with numbers from Table 2.

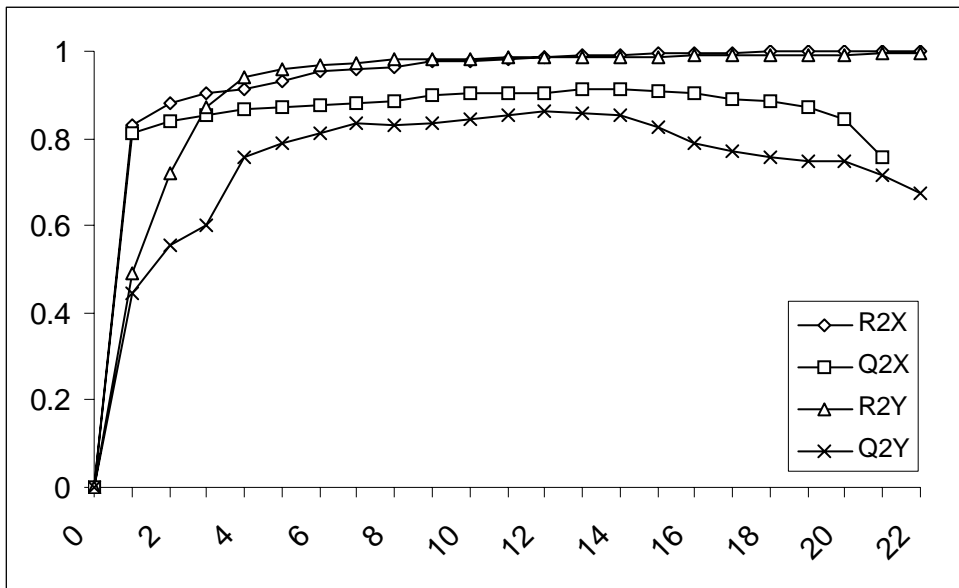


Fig. 4. Explained variances for calibration ( $R^2$ ) and cross-validation ( $Q^2_{\text{LOO}}$ ) vs number of components, PLSR1 model.

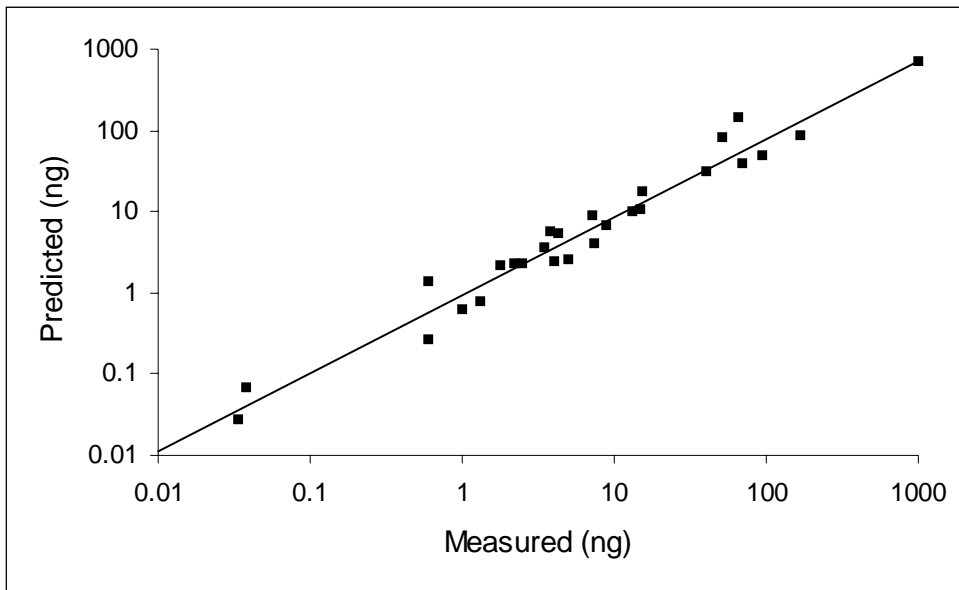


Fig. 5. Predicted vs measured amount of PCDD/PCDF in sample (log ng WHO-TEQ). PLSR1-model with five components.

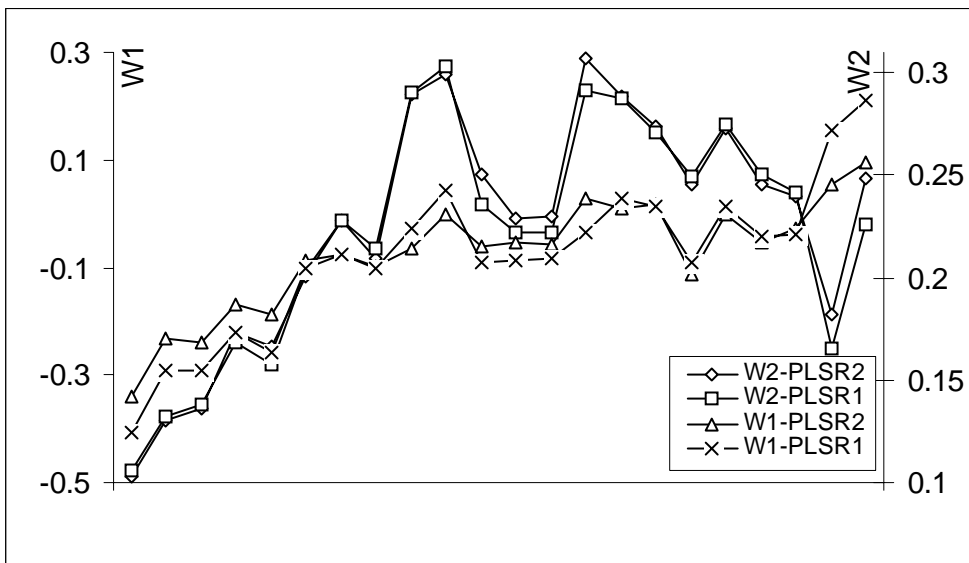


Fig. 6. Loading weights for the first two components, PLSR2 and PLSR1 model.

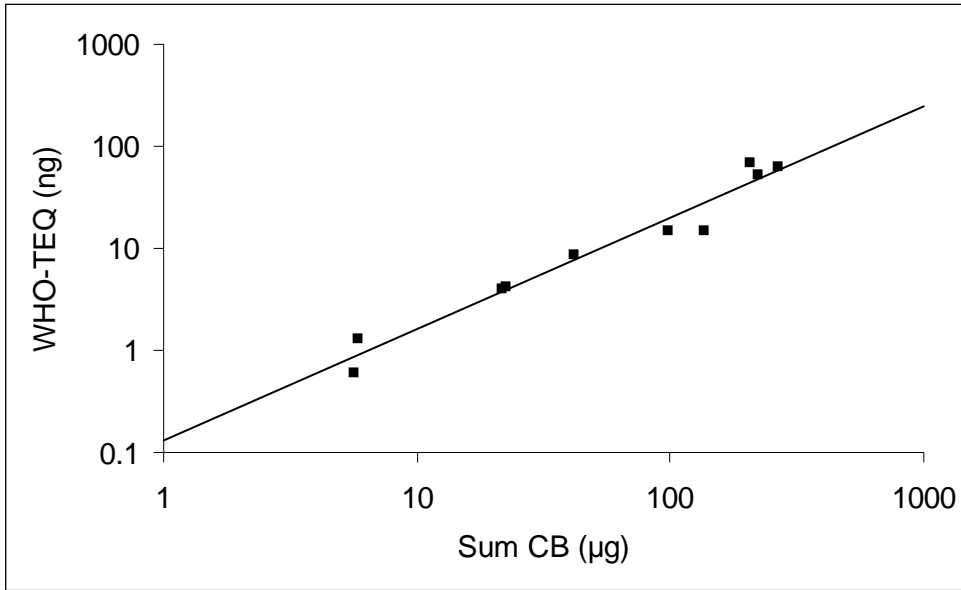


Fig. 7. PCDD/PCDF (log ng WHO-TEQ) vs. sum of chlorinated benzenes (log µg). Ten clean-gas samples from one plant.