Pattern recognition of polychlorinated dibenzo-*p*-dioxins and dibenzofurans: similarity values and analysis of variance

Andreas Kaune
Technische Universität München, Lehrstuhl für Ökologische Chemie und Umweltanalytik, D–85350 Freising, Germany, and GSF – Forschungszentrum für Umwelt und Gesundheit GmbH, Institut für Ökologische Chemie, Neuherberg, D–85764 Oberschleißheim, Germany; E-mail: kaune@gsf.de
Heidelore Fiedler
Universität Bayreuth, Lehrstuhl für Ökologische Chemie und Geochemie, D–95440 Bayreuth, Germany, E-mail: heidelore.fiedler@uni-bayreuth.de

1. Abstract

Two methods are applied for pattern recognition of polychlorinated dibenzo-*p*-dioxins (PCDD) and dibenzofurans (PCDF) and related compounds, i.e. a similarity value (*S*) and an analysis of variance (ANOVA) or its non-parametric analogue (*U* test). *S* was defined such that it becomes zero if two measurements are identical. It is a measure for the overall pattern similarity. To compare individual congeners and homologues in more detail, either an ANOVA or a *U* test was used. Both approaches were tested on PCDD/PCDF measurements of pentachlorophenol and perchloroethylene dry cleaning residues. *S* indicated a close agreement of both patterns while the *U* test identified differences in the furans to dioxins ratio and in the relative concentrations of some 2,3,7,8-chlorinated dibenzofurans.

2. Introduction

The results of the analysis of PCDD/PCDF and related compounds are often reported as toxicity equivalency concentrations (TEQ) using different national or the international toxicity equivalence factors. This does not give any information on the congener-specific results or the contribution of individual congeners to the I-TEQ. However, the distribution of homologues and congeners can provide very useful information because they may be characteristic for a given matrix. To obtain such information, usually multivariate statistical techniques are employed, e. g. principal component analysis and hierarchical cluster analysis. In this contribution, we propose two additional methods for pattern recognition of PCDD/PCDF: similarity values and analysis of variance of concentration ratios.

This approach is based on the idea that absolute concentrations of homologues and congeners in two different samples from the same matrix may vary considerably but that relative concentrations (or concentration ratios) may not and may therefore be very characteristic for a given matrix. Relative concentrations are, e.g., obtained by dividing a congener concentration by the sum of PCDD/PCDF, by the homologue concentration or by the concentration of other congeners. The proposed procedure enables the comparison of two data sets each containing PCDD/PCDF measurements of the same matrix. The procedure was implemented in a computer program and was used before to compare samples from a hazardous waste dump site with those from nearby soils [1,2].

3. Materials and Methods

Concentrations of congeners and homologues were first divided by the sum of PCDD/PCDF. From these relative concentrations (C_i), the similarity value (S) was generated according to

$$S = 1/n \sum_{i=1}^{n} (C_{1i} - C_{2i})^2 W_i \quad (1)$$

where the weighting factor (W_i) either equals 1 or was expressed as

$$W_i = \frac{\sum_{j=1}^{n} (s_{1j} + s_{2j})/n}{(s_{1i} + s_{2i})} \quad (2)$$

where s is standard deviation and n is the number of concentration values (n = 28 in congener-specific PCDD/PCDF analyses). S can be computed for two individual measurements. Then C_{1i} is the relative concentration of the first measurement, C_{2i} the relative concentration of the second measurement, and $W_i = 1$. This approach is similar to that of Morselli et al. [3] who applied the similarity value concept on homologue concentrations. However, contrary to ref. [3] the division by n (eq. 1) makes S independent from the number of congeners or homologues used. This procedure is useful when non-2,3,7,8-substituted PCDD/PCDF congeners, polychlorinated biphenyls or other analytes are considered, too.

Similarity values were computed for all possible combinations of measurements included in data sets 1 and 2. These similarity values can be distinguished in those where C_{1i} and C_{2i} are from the same data sets (intra similarity values) and those where C_{1i} and C_{2i} are from different data sets (inter similarity values). Arithmetic mean and standard deviation of inter S and intra S were computed. A large standard deviation of the intra S (rough estimate: standard deviation > mean) indicates a large inhomogeneity within this data set and should lead to an examination whether or not all measurements within this data set belong to the same parent population. Outliers may be identified by very large intra S values. A large difference between the mean intra S and the mean inter S is a first hint that measurements of data sets 1 and 2 are characterized by different patterns.

Furthermore, a mean (unweighted) and a mean weighted similarity value was calculated. In both cases, C_{1i} and C_{2i} are the mean relative concentrations of all measurements within data set 1 and 2, respectively. For the mean unweighted similarity value, W_i equals 1. For the mean weighted similarity value, W_i was computed according to equation 2. The weighting factor has the effect that congeners or homologues with a high standard deviation become less important in the calculation of S. On the other hand, small discrepancies between the non-weighted and the weighted S indicate that the standard deviations are relatively similar for all the congeners and homologues considered.

The smaller S, the more similar are the two patterns. S is a measure for the similarity of two patterns as a whole. However, large differences in only one or very few relative concentrations may not be detected by using S. Therefore, an analysis of variance (ANOVA) was computed separately for all pairs of analytes. Prerequisites of an ANOVA are that the data are from a normally distributed parent population of equal variance (ref. [4], p. 238). Normal distribution was tested employing the normal probability plot correlation method. Equality of variances was tested according to ref. [4] (p. 224). For normality and equality testing, a significance level $p \geq 0.10$ was used. If the requirements of normal distribution and equal variances were not met, the non-parametric U test according to Wilcoxon, Mann and Whitney (ref. [5], p. 230) was used. Values below the detection limits were set equal to zero.

To test the similarity and ANOVA/U-test approach, PCDD/PCDF concentrations in pentachlorophenol and sodium pentachlorophenate (PCP) [6] were compared with those found in perchloroethylene (PER) residues from the dry cleaning process. We expected very similar patterns in these two

matrices because PCDD/PCDF concentrations in dry cleaning PER residues (DCR) result from the textiles cleaned [7] and many textiles are treated with PCP and may thus be contaminated with the PCDD/PCDF impurities of the commercial PCP. The homologue patterns of the PCP and DCR samples used are shown in Figure 1.

4. Results and Discussion

Individual similarity values between relative PCDD/PCDF concentrations of PCP and DCR ranged from 1.4×10^{-2} to 4.5×10^{-5}. Measurements with high S values were checked but there was no solid reason to assume any error in the sampling and analytical procedure so that the measurements were not excluded from the statistical analysis. The mean unweighted S was 1.3×10^{-3} and the mean weighted S 4.9×10^{-4}. According to the discrepancy between both values, analytes whose relative concentrations differ largely between PCP and DCR samples also exhibit large standard deviations so that these analytes have little impact on the mean weighted S. To summarize these results, the similarity values were very small and indicated a close agreement in the PCDD/PCDF patterns of PCP and DCR.

Individual congeners and homologues were investigated in more detail using an ANOVA or U test. Relative concentrations of congeners and homologues of DCR samples were usually not normally distributed, and analytes of the PCP and DCR data sets had unequal variances so that the parametric ANOVA could not be used and the non-parametric U test was employed instead. The U test identified discrepancies in the furans to dioxins ratio. The ratio of the sum of PCDD to the sum of PCDD/PCDF was 76 ± 15 % for PCP and 93 ± 4 % for DCR. From the standard deviations it is also obvious that the PCDD fraction of PCP was much more variable than that of DCR. On the other hand, the PCP samples contained more PCDF (relative to the sum of PCDD/PCDF). This was also true for four 2,3,-7,8-substituted congeners, i.e. 2,3,7,8-Cl_4DF, 1,2,3,7,8-Cl_5DF, 2,3,4,7,8-Cl_5DF and Cl_8DF (Table 1).

Table 1. PCDD/PCDF congeners and homologues with significantly ($p < 0.01$) different relative concentrations between PCP and DCR samples (according to the non-parametric U test).

Analyte	relative concentrations (mean ± s) in samples of			
	PCP		dry cleaning PER residues	
PCDD	0.76	± 0.15	0.93	± 0.04
2,3,7,8-Cl_4DF (4F1)	8.3×10^{-6}	$\pm 9.7 \times 10^{-6}$	3.2×10^{-4}	$\pm 2.7 \times 10^{-4}$
1,2,3,7,8-Cl_5DF (5F1)	2.8×10^{-5}	$\pm 3.4 \times 10^{-5}$	5.0×10^{-4}	$\pm 6.1 \times 10^{-4}$
2,3,4,7,8-Cl_5DF (5F2)	2.1×10^{-5}	$\pm 2.6 \times 10^{-5}$	8.4×10^{-4}	$\pm 1.4 \times 10^{-3}$
Cl_8DF	0.16	± 0.11	0.020	± 0.014
PCDF	0.24	± 0.15	0.068	± 0.042

Table 1 lists analytes for which the null hypothesis (means are identical) was rejected on the 0.01 level of significance ($p < 0.01$). We also examined relative concentrations with high p values for which the null hypothesis could not be rejected. There are two explanations for high p values. 1. The standard deviations are extremely large so that differences in the arithmetic mean (if present at all) are not considered statistically significant. 2. Standard deviations are small and the means are nearly identical. This was true for the relative concentrations of Cl_8DD, Cl_6DF and 1,2,3,4,6,7,8-Cl_7DF, which are therefore very similar in both PCP and DCR samples.

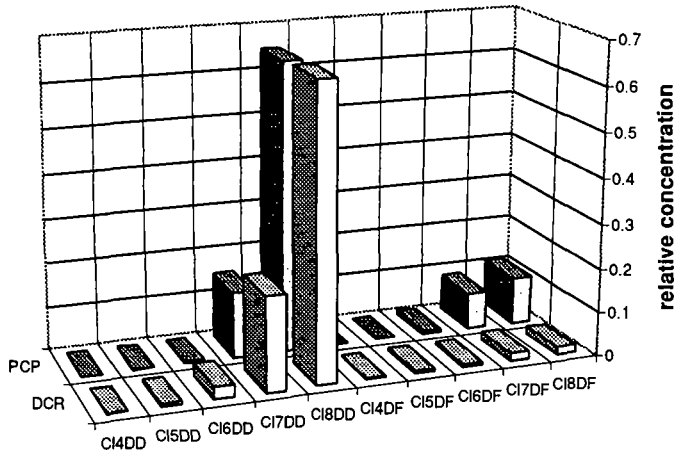

Figure 1. Homologue pattern of PCP and dry cleaning PER residues (DCR) (median values of 4 and 12 measurements, respectively). Concentrations were normalized to the sum of PCDD/PCDF = 1.

6. Conclusions

Similarity values and ANOVA or its non-parametric analogue (*U* test) can be applied for pattern recognition and source identification of PCDD/PCDF and related compounds. Individual similarity values can be calculated for two measurements. Mean similarity values and ANOVA/*U*-test allow the comparison of two matrices (two types of samples) and require a few measurements for each of the two matrices considered. Both approaches explicitly account for the variability of results typically encountered even if measurements are performed on the same matrix.

Further investigations should include different treatments of non-detected analytes and the choice of different significance levels so that the most appropriate one can be selected (we used $p < 0.01$).

6. References

1) Fiedler H., A. Lamprou, C. Lau, C. Maxin and O. Hutzinger (1993): Bewertung: Dioxin-Untersuchungen im Umfeld der Sonderabfalldeponie Hoheneggelsen. Gutachten für die Niedersächsische Sonderabfalldeponie Hoheneggelsen GmbH, Hannover, Germany.

2) Fiedler H., C. Lau and O. Hutzinger (1994): Dioxinmobilität im Umfeld der Sonderabfalldeponie Münchehagen. *In* G. Dörhöfer, J. Thein and H. Wiggering (eds): Altlast Sonderabfalldeponie Münchehagen. Umweltgeologie heute 4. Verlag Ernst & Sohn, Berlin, Germany, p. 95–102.

3) Morselli L., S. Zappoli, A. Liberti, M. Rotatori and E. Brancaleoni (1989): Evaluation and comparison of organic and inorganic compounds between emission and immision samples from municipal solid waste incinerator. Chemosphere 18, 2263–2273.

4) Kreyszig, E. (1982): Statistische Methoden und ihre Anwendungen. Vandenhoeck und Rupprecht, Göttingen, Germany, 7th Ed.

5) Sachs L. (1978): Angewandte Statistik. Springer-Verlag, Berlin, Germany.

6) Hagenmaier H. (1987): Abschlußbericht zum Forschungs- und Untersuchungsvorhaben: Belastung der Umwelt mit Dioxinen. Institut für Organische Chemie der Universität Tübingen, Germany.

7) Umlauf G., M. Horstmann, M. Klein and P. Kurz (1993): Mass balance of PCDD/F in a dry cleaning machine. Organohalogen Compd. 11, 413–416.