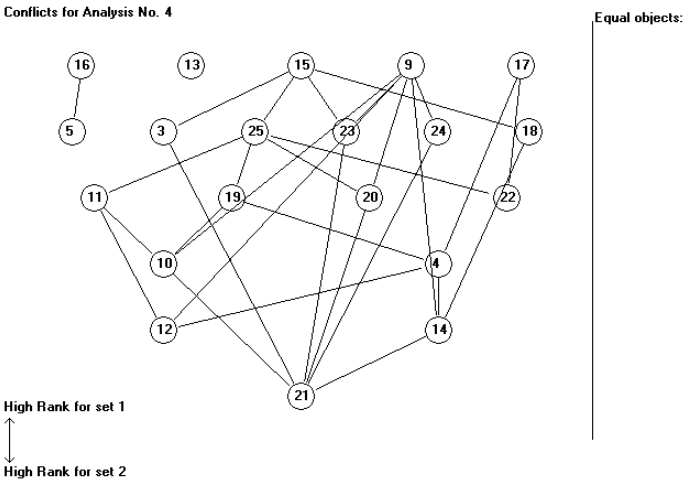

Fig. 12. Conflict diagram for the data set where the banned pesticides are excluded (Ids. 1, 2, 6, 7, 8) and where the only variable used in *Set 2* is *SpArea*. (Figure drawn using Po Correlation)

The numbered circles show the correlation results for a series of different variable combinations used for respectively *Set 1* and *Set 2*, which also refer to the columns in the Table to the right in Fig. 13. The small circles in this table identify the included variables. So *e.g.* for analysis no. 1, *Set 1* was made by the two variables *DetFreq* and *MedMax* and *Set 2* was made by only *Dose*. The variable denoted PRODUCT is the product between *Dose* and *SpArea* as presented by Table 4. The analysis 1, 2 and 3 is a test of every single variable including PRODUCT in relation to *Dose* and *SpArea* as *Set* 1. The product is best as single variable followed by respectively *Dose* and *SpArea* and this follows naturally the τ values in Table 4. However, the analysis no. 4 shows that the partial order of *Dose* and *SpArea* together performs better than the product if only the T (0,0) value is considered. The value drop on the y-axis from analysis no. 3 to analysis no. 4 is due to increased discordance in the ranking introduced when two variables are included in *Set* 2 instead of only a single one. The use of respectively *DetFreq* and *MedMax* as single variable in *Set* 1 is also tested as respectively no. 5 and no. 6. Neither analysis no. 5 nor 6 can make the same good correlation measured by T (0,0) as analysis no. 4.

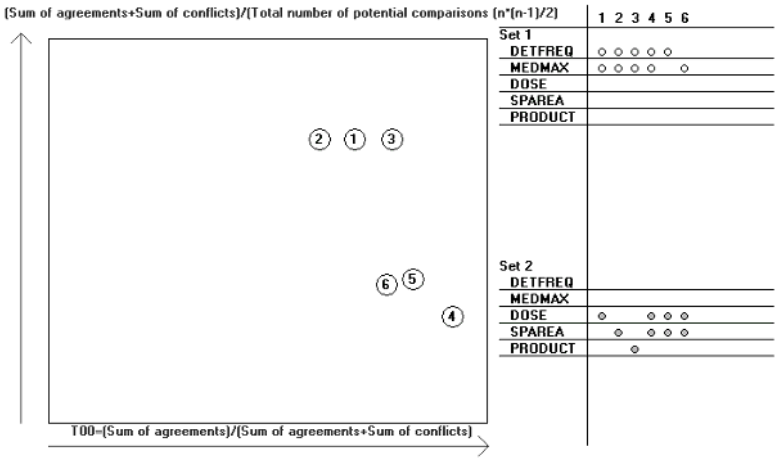

Fig. 13. A graphical display of the correlation result for a series of different variable combinations. The numbered circles refer to the numbering in the top row in the table to the right. Each number is a correlation analysis and the small circles in the tables indicate which variables that have been used in respectively *Set 1* and *Set 2*. The x and y axis is similar to the axis in the significance plot as explained for Fig. 8. (Figure drawn using Po Correlation)

References

Brüggemann R, Halfon E, Welzl G, Voigt K, Steinberg C (2001) Applying the Concept of Partially Ordered Sets on the Ranking of Near-Shore Sediments by a Battery of Tests. J Chem Inf Comp Sc 41:918-925

Brüggemann R, Bücherl C, Pudenz S, Steinberg CEW (1999) Application of the concept of partial order on comparative evaluation of environmental chemicals. Acta Hydrochim Hydrobiol Vol 23:170-178

Brüggemann R, Bartel HG (1999) A theoretical concept to rank environmentally significant chemicals, Journal of Chemical Information and Computer Sciences Vol 39(2):211-217

Conover WJ (1999) Practical nonparametric statistics, John Wiley & Sons, ISBN 0-471-16068-7

Gibbons JD (1993) Nonparametric measures of association, SAGE University papers. series: quantitative applications in the social sciences, ISBN 0-8039-4664-3

Goodmann LA and Kruskal WH (1963) Measures of association for cross-classifications III: Approximate sample theory. Journal of American Statistical Association Vol 58:310-364

Kendall MG (1938) A new measure of rank correlation. Biometrika Vol 30:81-93
Pavan M (2003) Total and partial ranking methods in chemical sciences. PhD thesis in Chemical Sciences, Cycle XVI, University of Milano – Bicocca
Pudenz S (2004) Stefan Pudenz, Criterion - Evaluation and Information Management. Mariannenstr. 33, D-10999 Berlin, Germany (web: www.criteri-on.de)
Pudenz S, Brüggemann R, Komossa D, Kreimes K (1998) An algebraic/ graphical tool to compare ecosystems with respect to their pollution by Pb, Cd III: comparative regional analysis by applying a similarity index. Chemosphere Vol 36:441-450
Sørensen PB, Brüggemann R, Thomsen M, Lerche DB (2005) Application of multidimensional rank-correlation, MATCH Communications in Mathematical and in Computer Chemistry 54:643-670
Sørensen PB, Brüggemann R, Carlsen L, Mogensen BB, Kreuger J and Pudenz S (2003) Analysis of monitoring data of pesticide residues in surface waters using partial order ranking theory - Data interpretation and model development. Environmental Toxicology and Chemistry Vol 22(3):661-670

Evaluation of Biomonitoring Data

Dieter Helm

Robert Koch-Institute, Seestraße 10, D-13353 Berlin, Germany

e-mail: helmd@rki.de

Abstract

The construction of posets or Hasse diagrams is a profitable means for the evaluation of biomonitoring data. In contrast to other statistical approaches the Hasse Diagram Technique enables the consideration of multiple attributes at the same time and will result in at least partially ordered data sets. Moreover, the calculation of averaged ranks allows the construction of a total order for a given data set. For the evaluation of biomonitoring data, as obtained for the German Environmental Specimen Bank, the Hasse diagram technique was applied to achieve partially or totally ordered data. The following scheme was applied: i) Careful rounding of the original data to increase the number of comparabilities; ii) splitting of the data in smaller sub-sets; iii) construction of the posets for each sub-set; iv) construction of the total order for each sub-set (by means of averaged ranks) and, v) synopsis of the sub-sets.

Introduction

At about the same time when Hasse published his famous book about algebraic topics and made Hasse diagrams popular (Hasse 1967), the environmental pollution reached its first depressing climax in Middle Europe. DDT, for example, which was invented during World War II, being the most powerful insecticide the world had ever known, was used with gay abundance in such large quantities, that this abuse caused the death of countless song birds and, as it accumulated in the food chain, prevented breeding success of the birds of pray, thus motivating Rachel Carson to write her famous book The Silent Spring (Carson 1962) – actual data on

DDT are summarised in Sørensen et al. 2004. DDT was so popular in these times that even the plant enthusiast treated ornamental pot plants in the home with this pesticide, often in too large a quantity and in a negligent fashion (Buxbaum 1958). In Germany the banks of the River Rhine were regularly covered with froth and the corpses of innumerable fish during the late 1960ies and the 1970ies. Acid rain killed trees in coniferous woodlands. Traffic-borne lead (from leaded gasoline) reached about 1970 an average air concentration of 125 ng/m^3 in the Hamburg area; human body burden was estimated retrospectively to be about 150 µg per litre blood at the same time (von Storch et al. 2003). (Actual values: approx. 20 ng lead per cubic metre air und 19 µg lead per litre blood for the student participants of the Environmental Specimen Bank (ESB). In due time, these alarming figures prompted the governments to action. In 1963 the Convention on the International Commission for the Protection of the Rhine against Pollution (Bern Convention) was signed by Switzerland, France, Luxembourg, Germany and the Netherlands. The European Economic Community joint in as contracting party in 1976. The DDT Act from 1972 prohibited production and use of DDT in Germany. The Petrol Lead Act of 1971 accomplished reduction of airborne pollution from lead in gasoline.

The German Environmental Specimen Bank

To monitor the effects of these and other activities of the legislative, the ESB was brought into being by the Federal Environment Ministry (FEM) of Germany in the 1970ies as a component of ecological environment surveillance (FEM 2000) and started sampling in 1981. After a series of development and trial phases, expansion of the ESB to full-scale operation (13 ecosystems and 4 human sampling sites) started on 1 January 1994. According to the concept (FEM 2000), the ESB, collects ecologically representative environmental and human specimens, which are analysed for environmentally relevant substances, and stored. Long-term storage is performed under conditions, which exclude any change in composition or chemical properties over a period of several decades. This archive retains specimens for retrospective analytical characterisation concerning unpredictable questions, which may arise in future. Although the specimens are analysed for a number of environmental substances prior to storage (monitoring), the genuine value of the ESB is the storage of samples (archive) to serve as records for the conservation of eco-toxicological and toxicological evidence. In order to attain a high level of quality assurance, all stages from sampling itself, to the transportation of specimens, the preparation

and analysis of specimens, through to long-term storage, are set out in mandatory standard operating procedures (SOPs) for all types of environmental and human specimens. The environmental specimens are obtained from representative areas (ecosystems); representative ecosystems from the terrestrial, limnetic (riverine) and marine environment have been selected as examples. With regard to its entire composition each ecosystem is depicted by the biological, physical and chemical characterization of specimens taken from this system: producer, consumer, and destructor (FEM 2000). Table 1 gives an overview of the sampling programme of the ESB.

Table 1. Sampling programme of the ESB

Sampling area	Sampling sites	Ecosystem type	Specimens sampled
BR/NP Wadden Sea	3	Marine	1 - 4
Bodden NP of Western Pomerania	3	Marine	1 - 4
River Elbe	5	Riverine	5, 6, 15
River Rhine	4	Riverine	5, 6, 15
Bornhoeved Lake District	1	Agrarian	6 - 11, 14
Upper Bavarian Tertiary Uplands	1	Agrarian	7 - 9, 11, 14
Solling	1	Forestry	7 - 9, 11, 14
Palatinate Forest BR	1	Forestry	7 - 9, 11, 14
Saarland conurbation	3	Close to conurbation	5 - 7, 9 - 12, 14, 15
Dueben Heath	2	Close to conurbation	6, 9 - 13
Upper Harz NP	1	Nearly natural	7 - 9, 14
Berchtesgaden BR/NP	1	Nearly natural	7 - 9, 14
Bavarian Forest BR/NP	1	Nearly natural	7 - 9, 14
Human	4 Universities	Urban	Blood, urine, hair, saliva

Abbreviations: BR = UNESCO Biosphere Reserve (MAB programme); NP = National Park; 1 = common bladder wrack (*Fucus vesiculosus*); 2 = common mussel (*Mytilus edulis*); 3 = eelpout (*Zoarces viviparus*); 4 = herring gull's egg (*Larus argentatus*); 5 = zebra mussel (*Dreissena polymorpha*); 6 = bream (*Abramis brama*); 7 = spruce (*Picea abies*); 8 = beech (*Fagus sylvatica*); 9 = roe deer (*Capreolus capreolus*); 10 = domestic pigeon's egg (*Columba livia f. domestica*); 11 = earth worms (*Lumbricus terrestris* or *Aporrectodea longa*); 12 = Lombardy poplar (*Populus nigra* ‚Italica'); 13 = pine (*Pinus sylvestris*); 14 = soil; 15 = sediment

Most specimens are sampled in yearly intervals. Sampling is regulated by prescribed sampling timetables and by standard operating procedures (FEM 2000). Individual samples from each sampling site are pooled to give a pooled sample of approx. 5 kg wet weight. Further processing of the pooled samples includes grinding and freeze drying, resulting in a fine homogeneous powder, which is packaged as 10 g portions. These are stored for the long term over liquid nitrogen in the gas phase at temperatures below -150 °C. Results of the measurements prior to the storage are published in comprehensive reports (e.g. FEA 1999).

Evaluation Methods

The univariate approach

Since the temporal changes are of interest, investigations of trends are performed applying the Mann-Kendall-Test and the estimator of Theil (Conquest 2000). Fig. 1 shows a simple example of a trend analysis. Thallium values from 1994 to 2003 from bream musculature (*Abramis brama*) were obtained from pooled samples from the two Saarland sampling sites Rehlingen barrage weir und Güdingen barrage weir.

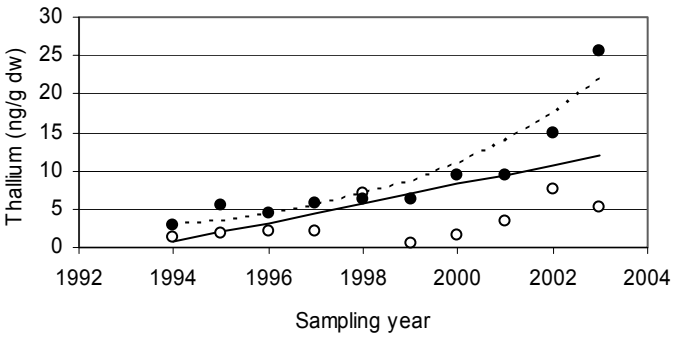

Fig. 1. Thallium measured in the musculature of the Bream (*Abramis brama*) from the Saarland conurbation. Closed circles: Rehlingen barrage weir; open circles: Güdingen barrage weir. The solid line shows the linear component of the trend as estimated with the Theil estimator and the dotted line depicts the trend estimator for data linearised prior to the analysis

Whilst the Mann-Kendall-Test found no trend in the Güdingen data ($p = 0.054$), the testing resulted in a positive trend for the Rehlingen data ($p < 0.001$) due to an input of the industrial zone of Dillingen located ca. 1 km up-stream from the Rehlingen weir. Theil estimator calculated the linear component of the trend to be an annual increase of 1.26 ng Thallium per g dry weight. Linearisation of the measurement values (by calculation of log10 values) prior to the Mann-Kendall-Test yielded a better fit of the trend line (dotted line in Fig. 1). What is much needed for the evaluation of biomonitoring data is a procedure that can evaluate the samples in terms of their contamination with environmental substances. In the above example, Rehlingen is always more strongly contaminated than Güdingen with the sole exception of 1998. For all other years one would order the both sites as Rehlingen > Güdingen when only Thallium is considered. (See also the chapter by Pudenz and Heininger p.111, where the results of ecotoxicological tests in river sediments are presented and analyzed.)

The multivariate approach

Like many other statistical methods for the evaluation of biomonitoring data, the above-depicted example of a trend analysis considers only a single variable. Although the multivariate procedures consider several measured variables at the same time, their results are often only limited meaningful. Cluster analyses can reveal structures in a given data set; principal component analyses concentrate the information contents of many variables in a set of a few latent variables, which are difficult to interpret correctly.

Fig. 2 depicts the result of a cluster analysis performed on the data in Table 2. Sampling sites are combined to clusters according to the similarity of their contamination pattern rather than the magnitude of this contamination. Consequently, four clusters are formed: cluster 1 consisted of the sites Weil, Koblenz, Cumlosen, Blankenese and Barbay; cluster 2 of Saale, Güdingen and Mulde; cluster 3 comprised the sites Iffezheim, Bimmen and Rehlingen; and the fourth cluster, more isolated, was composed by Prossen and Zehren. Even if this partition could reveal interesting structures, it was not possible to identify the most severely contaminated sites, let alone to order them according to their contamination. Similar results were obtained by principle component analysis (PCA; see Fig. 3). PCA yielded four principal components, which explained 82 % of the entire information (formerly distributed in 11 measured variables). The first component was dominated by DDE, OCS, HCB, Hg, and Pb. The second summarised the information of PCB, Cu and As. The third component was

dominated by Tl and HCH, whereas the fourth component represented Se. The clusters differed somehow from those detected by cluster analysis (Fig. 3). Additionally, the sampling sites could be ordered along the component axes. In respect to component 1, Zehren was mostly contaminated whereas Güdingen and Rehlingen appeared to be the cleanest sites. In contrast to this, Rehlingen was the most contaminated site when component 2 was considered. Thus, no unequivocal order could be achieved.

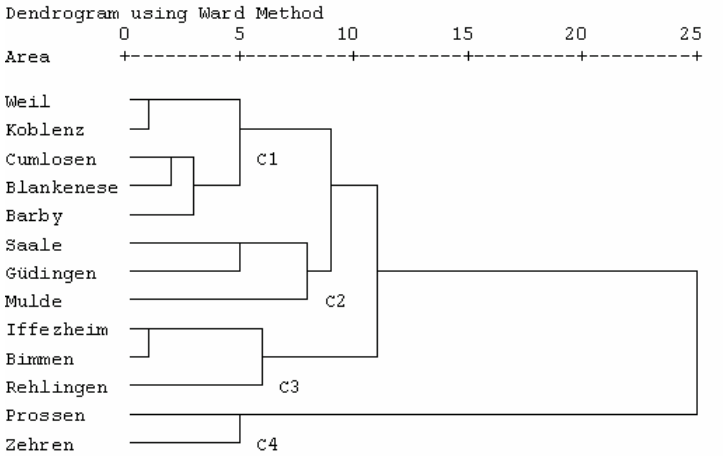

Fig. 2. Result of a hierarchical cluster analysis of the data in Table 2

The partial order approach: Evaluation of ESB data with the Hasse diagram technique

Neither cluster analysis nor PCA could order the sampling sites considering the complete information. One approach to judge more than one measured variable at the same time is the calculation of an index, which is supposed to summarise the information of all single variables. This approach, however, arises new problems, which are difficult to solve. First, the calculation of an index would require correct weighting factors for all single variables and secondly, before adding up the values all variables must be transformed to the same numeric scale. Consider, for example, the third row in Table 2. The index for Blankenese would be approx. 3 µg/g for the organic compounds when the weighting factors for all substances would be set to 1. This would mean that neither HCB nor the very much more toxic DDE will contribute equal amounts (this is, each approx. 20 %) to the index's value.

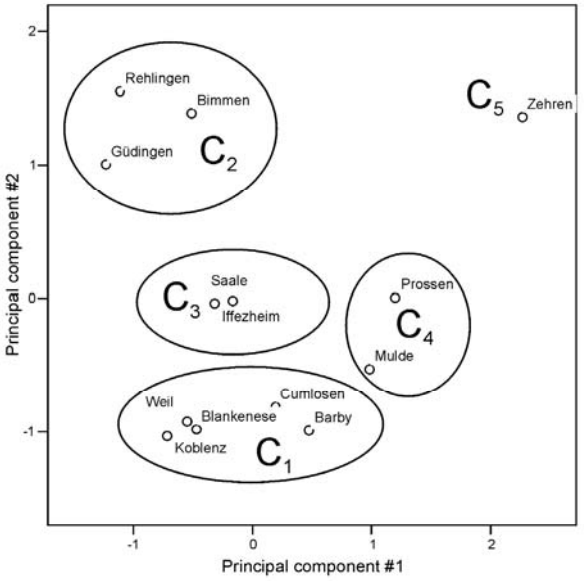

Fig. 3. Result of PCA performed on the data in Table 2

Since this is obviously unjustified we need to apply weighting factors accounting for the ecotoxicity of each compounds, but the problem is that we do not know them. An additional disadvantage of the index is the loss of information since all data are lumped together to give a stew.

The Hasse Diagram Technique (HDT) is a method out of the scale of the mathematical order theory, which was already applied productively to ecological and ecotoxicological problems (e.g. Brüggemann et al. 1999; Brüggemann 2001a; Brüggemann and Halfon 1997; Brüggemann and Steinberg 2000). More recent applications of the HDT in the environmental sciences can be seen elsewhere in this book. The principle of the technology is based on the fact that 'objects' (e.g. ecological systems, habitat diversity, field data, chemicals, databases, sampling sites of the environmental sample bank) are compared each with one another and ordered on the basis of their attributes (e.g. measured values of xenobiotics). Each comparison of objects considers all attributes at the same time. An object is classified to be 'worse' than another, if it is 'worse' or at least 'equal' in respect to each of the attributes than the other object (or, in other words, if the product order relation is fulfilled). The designation of the relation is then: '>'. In ecology the term 'better' can be replaced by 'more contaminated' since we consider higher concentrations to be 'worse'. If two objects

are equal for all attributes, the symbol for their relation is '='; they are equivalent. Controversial or inconsistent results of a comparison will result when object a is 'better' than object b in respect of one attribute but at the same time 'worse' in respect to another. This will be labelled with the symbol '||'. In terms of the HDT the two objects are 'not comparable' and can, thus, not put into any order.

In short,

a > b means that object a is more contaminated than b
a = b means that object a and b are contaminated in the same way
a || b means that objects a and b cannot be compared with another

Table 2. Original data for bream from 2002 (inorganic compounds in terms of dry weight; organic compounds related to the fat contents of the musculature)

Sampling site	HCB	γ-HCH	OCS	ΣPCB	DDE	As	Pb	Cu	Hg	Se	Tl
			[µg/g]			[µg/g]	[ng/g]	[µg/g]	[ng/g]	[µg/g]	[ng/g]
Barby	1.60	0.027	0.262	2.38	2.10	0.38	86.4	1.40	1050	3.40	2.7
Bimmen	0.24	0.016	0.089	7.40	0.57	0.72	128.8	1.15	927	2.30	8.8
Blankenese	0.57	0.038	0.193	1.58	0.58	0.54	62.3	2.04	524	2.45	0.5
Cumlosen	0.98	0.025	0.301	2.17	1.48	0.43	59.0	1.93	1270	1.91	6.4
Güdingen	0.10	0.033	0.002	7.25	0.61	0.24	91.8	1.48	507	1.47	7.5
Iffezheim	0.56	0.017	0.054	4.22	0.41	0.66	87.8	1.08	1155	2.53	1.8
Koblenz	0.17	0.012	0.012	1.36	0.23	0.43	62.4	1.00	426	2.82	0.8
Mulde	1.16	0.033	0.392	2.71	5.19	0.42	98.2	2.72	1520	2.49	12.3
Prossen	1.47	0.017	0.620	6.54	5.14	0.37	150.0	1.44	1585	1.83	1.5
Rehlingen	0.06	0.043	0.003	7.16	0.34	0.56	66.9	1.31	603	3.17	14.8
Saale	0.20	0.042	0.047	2.29	2.54	0.34	149.0	1.24	1638	2.28	2.0
Weil	0.12	0.013	0.058	2.64	0.46	0.45	51.1	1.51	841	2.27	1.4
Zehren	3.15	0.022	0.711	8.54	9.28	0.47	151.5	1.29	1973	2.25	16.6

For all objects, which meet the order relation, a (partial) order can be constructed. The position of any given object within this order enables direct reading of the relative ecological or ecotoxicological load of this object. Additionally, the applied software, WHASSE (Hasse for Windows, producer: GetSynapsed GmbH, München) allows presenting the pollutant profile as bar chart, increasing the information content of the Hasse diagrams (Bücherl et al. 1995). The differences made visible in this way allow conclusions to contamination processes and pathways (Brüggemann 2001a). The most crucial advantage of the HDT is the synoptic view of

several attributes, without the need of an index or a quality function (Brüggemann 2001b). This synoptically view enables an ecosystemical evaluation. The helpfulness of the HDT to the comparison of ecological systems in respect to their environmental loading had been already demonstrated (Brüggemann et al. 1994).

Some terms often used in connection with the HDT must be explained since they are necessary to understand and interpret a Hasse diagram. Please note, that Hasse diagrams are read top-down or exclusively bottom-up. In this chapter the top-down view is preferred.

- Anti-chain: An alignment of objects, which are not comparable with one another. Elements of the same level (see chapter by Brüggemann and Carlsen, p. 61) are incomparable. They can be considered to be similarly polluted but with different pollution patterns. As sometimes the construction of levels cannot be done uniquely, their interpretation needs some care.
- Chain: An alignment of objects, which are all comparable with one another. The elements of a chain are connected with a line. Often a common mechanism is responsible for the formation of the chain, which leads to the synchronous, at least weakly monotonous increase of some attributes.
- Equivalence: Two objects are equivalent, if all attributes have equal values.
- Incomparability: Two objects are not comparably with one another if the first object is, in respect to at least one attribute, 'worse' than the other and, simultaneously, in respect to another attribute, 'better'. In the Hasse diagram incomparable objects are not connected with lines.
- Isolated object: Objects, which are comparable with none of the other objects in the data set. Because of conservatism isolated objects are always assigned to the highest diagram level. This corresponds with the assumption that high pollutant concentrations indicate a high endangerment. The software Hasse for Windows automatically adopts the correct arrangement.
- Maximal objects: Objects, for which no other objects exist in the data set, which can be classified 'worse' are called maximal objects. In WHASSE maximal objects are assigned to the uppermost level in the diagram.
- Minimal objects: Objects, for which no other objects exist in the data set, which can be classified 'better' are called minimal objects.

- Product order relation: The product order relation is fulfilled if i) an object is, in regard to at least one attribute, 'worse' than another and at the same time 'equal' in respect to all remaining attributes, or if ii) an object is, in regard to at least one attribute, 'better' than another and at the same time 'equal' in respect to all remaining attributes. The existence of the product order relation is the prerequisite for the construction of chains.
- Predecessor: An object ranked above a given object x. In our case the predecessor is more contaminated than x.
- Sensitivity: A measure for the importance of a particular attribute. How strongly will the omittance of a given attribute change the resulting Hasse diagram? Most often the matrix W, calculated in WHASSE, is used to perform sensitivity studies.
- Stability: An estimator for changes in the diagram to be anticipated, if any attributes will be added or omitted. Symbol: P(*IB*). Since P(*IB*) is normalised and can only take values between 0 and 1 it can easily be interpreted. If P(*IB*) = 1, then all objects are arranged in an anti-chain – the inclusion of additional attributes will not change the structure. If P(*IB*) = 0, then all objects are arranged in a chain or they are equivalent to each other - the chain (and/or the equivalence) remains, if attributes are omitted.
- Successor: An object ranked below a given object.

The Data Set

Specimens

The bream (*Abramis brama* Linné 1758, Cyprinidae) is a predominantly carnivorous teleost fish of the carp family inhabiting lakes and quiet parts of slowly running rivers and is widespread throughout Europe (Bond 1979). It feeds mainly on small molluscs (e.g. *Pisidium*, *Anisus*), tubifex and insect larvae (e.g. *Chironomus*), sucking in sediment with the aid of its protrucible mouth while foraging and extracting the food from the sediment, but will also take plant material occasionally.

Data were obtained for the bream's swimming musculature as collected for the ESB in 2002. Different numbers of 4 to 19 years old bream (n = 16 – 28) have been captured after the breeding season in each of the 13 different sampling areas. HCB, γ-HCH, octachlorostyrole (OCS), PCB 101,

PCB 118, PCB 138, PCB 153, PCB 180, and 4,4'-DDE were measured via GC-MS. The 5 PCB congeners were added up to give the sum parameter ΣPCB. With the exception of mercury, which was measured via DMA, all metal analyses were carried out using ICP-MS.

Fig. 4. The bream (*Abramis brama*), a specimen for the ESB (Photograph by Dr. Roland Klein, University of Trier)

Sampling areas

Bream were collected in 13 sampling areas belonging to the three riverine zones of Elbe, Rhine and Saar. ESB samples of bream were collected at 5 different sampling sites along the river Elbe; these are – in down-stream direction: Prossen (river km 13), Zehren (km 93), Barby (km 296), Cumlosen (km 470), and Blankenese near the Port of Hamburg (km 633). Additionally, bream from the Elbe tributaries Saale and Mulde (the latter near the mouth), were sampled too. In 2002, the year of the Elbe flood, at all sites (with the exception of Blankenese) bream were collected during the first days of the announcing flood. The Blankenese specimens were collected after the flood (in September). River Rhine was represented by sampling areas Weil (river km 174), Iffezheim (km 334), Koblenz (km 590,3), and Bimmen near the German-Dutch border (km 865). Bream from the river Saar stemmed from the barrage weirs of Güdingen (km 93) and Rehlingen (km 54).

Hasse Diagrams or POSETs

Hasse diagram for the entire data set

The entire raw data set, as listed in Table 2 (13 objects, 11 attributes), was used to construct a Hasse diagram. The result (Fig. 5A) is a single antichain, due to very individual contaminant patterns of the 13 sampling areas. None of the sampling areas is comparable to any other. This result, being somewhat unpromising, leads us to the development of a double fold stratagem (Helm 2002). Firstly, since the incomparableness of the objects is, in many cases, due to only minor differences of the measured values, many of them being in the magnitude of the measurement uncertainty, an appropriate, carefully performed rounding of the data will better the diagram. Take, for example, the As values for the Elbe sampling areas Cumlosen (0.43 [µg/g dw]) and tributary Mulde (0.42 [µg/g dw]). Since 0.43 is greater than 0.42, Cumlosen can be considered to me more contaminated than Mulde. This difference, however, lies within the measurement uncertainty of approx. 10% and is, therefore, not justified. (Note the similar ideas of "smoothing" of data (noise deficient QSAR) in chapter by Carlsen, p. 163).

Secondly, splitting the entire data set into two or more sub-sets, each with a lower number of attributes, will yield more comparabilities since a greater number of attributes will reduce the probability that the product order relation is fulfilled: $P((a_n > b_n) \cap (a_m > b_m)) < P(a_n > b_n)$.

Let us first focus on the rounding procedure. Desirable as it may be, rounding of the data according to the overall data error is not possible, due to the simple fact that we do not know it. The complete data error is composed by i) the simple (or multiple) measurement uncertainty, ii) the long-term lab error and iii) the sampling error. The simple measurement uncertainty is – depending on the contaminant and its concentration – about 10 to 20 %. The long-term lab error is even more important because the series of measurement in biomonitoring can easily span many years (like in ESB). Unfortunately the latter is unknown and the same is true for the sampling error. Thus we better adjust the rounding of the data to the scale of the measured values.

Pre-processing: Rounding of the raw data

A good measure for the scale (or magnitude) of data is the median. Other than the arithmetic mean the median is insensitive to the distribution of the data. After calculation of the median for every data column of Table 2, data are to be rounded to the *k*. decimal place.

$$k = \text{trunc}(log10(\text{median})) - 1 \quad \text{if median} < 1 \tag{1}$$
$$k = \text{trunc}(log10(\text{median})) + 1 \quad \text{if median} \geq 1 \tag{2}$$

If k is negative, the data will be rounded to k^{th} position after the decimal point; if k is positive, the data will be rounded to k^{th} position prior to the decimal point. Two examples illustrate the method (HCB and Hg from sampling area of Prossen, river Elbe).

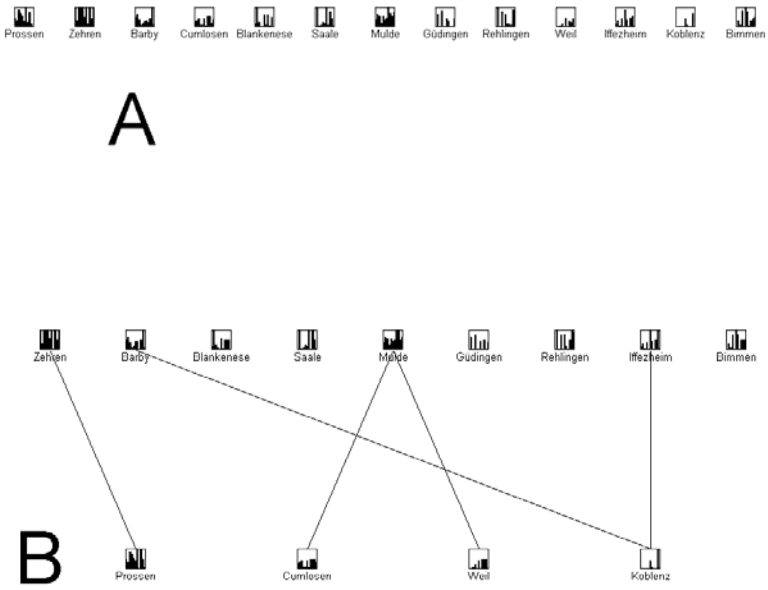

Fig. 5. Hasse diagram for the raw data (A) and for the rounded data (B)

Table 3. Stepwise implementation of the rounding process

Step #	Example 1	Example 2
1. Measured value	1.47	1585
2. Median of the data column	0.559	1050.00
3. The logarithm of the median	-0.253	3.021
3. Truncation of the decimal	0	3
4. decrement, if median < 1, or,	-1	
increment, if median ≥1		4
5. *k*	-1	4
6. Resulting rounded value	1.5	2000

In example 1 *k* is -1, which means that the HCB values will be rounded to the first digit after the decimal point, and the mercury values will be rounded to the fourth digit before the decimal point, because *k* is 4 in this case. The rounded data are listed in Table 4.

Table 4. Rounded data for bream from 2002 (inorganic compounds in terms of dry weight; organic compounds related to the fat contents of the musculature)

Sampling site	HCB	γ-HCH	OCS	ΣPCB	DDE	As	Pb	Cu	Hg	Se	Tl
			[µg/g]			[µg/g]	[ng/g]	[µg/g]	[ng/g]	[µg/g]	[ng/g]
Barby	1.6	0.03	0.26	2	2.1	0.4	90	1	1000	3	3
Bimmen	0.2	0.02	0.09	7	0.6	0.7	130	1	1000	2	9
Blankenese	0.6	0.04	0.19	2	0.6	0.5	60	2	1000	2	1
Cumlosen	1.0	0.02	0.30	2	1.5	0.4	60	2	1000	2	6
Güdingen	0.1	0.03	0.00	7	0.6	0.2	90	1	1000	1	7
Iffezheim	0.6	0.02	0.05	4	0.4	0.7	90	1	1000	3	2
Koblenz	0.2	0.01	0.01	1	0.2	0.4	60	1	0	3	1
Mulde	1.2	0.03	0.39	3	5.2	0.4	100	3	2000	2	12
Prossen	1.5	0.02	0.62	7	5.1	0.4	150	1	2000	2	2
Rehlingen	0.1	0.04	0.00	7	0.3	0.6	70	1	1000	3	15
Saale	0.2	0.04	0.05	2	2.5	0.3	150	1	2000	2	2
Weil	0.1	0.01	0.06	3	0.5	0.4	50	2	1000	2	1
Zehren	3.1	0.02	0.71	9	9.3	0.5	150	1	2000	2	17

The Hasse diagram constructed for the rounded data is depicted in Fig. 5B. As a consequence of the rounding, the number of levels increased from 1 to 2, the number of incomparabilities was reduced from 156 to 146.

Five chains have been formed, from which three do represent intra-river segments, indicating an increase of pollution for the river Elbe from Prossen towards Zehren, a decrease from tributary Mulde towards Cumlosen, and, for the river Rhine, a decrease of the pollution from Iffezheim towards Koblenz. However, the chains are very short (each comprising only two elements) and there are still 5 isolated objects (Blankenese, Saale, Güdingen, Rehlingen, Bimmen). The stability, P(*IB*), is 0.94, indicating, that the diagram is very near an anti-chain and that the inclusion of additional attributes will not change the structure. On the other hand, the exclusion of attributes may very well change the structure. This leads to the second stratagem, the splitting of the data set. The data set will be splitted into two sub-sets; one containing the inorganic components and the other containing the organic substances.

Hasse diagram for the inorganic compounds

The resulting Hasse diagram or Partial Order for the six inorganic compounds was similarly poorly structured as the Hasse diagram for all 11 pollutants when the original raw data were used (Fig. 6A). The diagram consisted of only two levels; 9 objects were assigned to an anti-chain. Only two short chains were formed, each of which consisted of two objects and none of these short chains represented intra-river segments. Application of rounded data produced an additional level and the number of incomparabilities was reduced from 152 to 130. Two three-link chains and 6 two-link chains were formed; some of them corresponded to river segments, thus allowing deriving a partial order for these streams with an enriched degree of comparabilities (Fig. 6B). Within the first chain for river Elbe the sampling site of Zehren appeared to be most contaminated, followed by Prossen and tributary Saale. The second Elbe chain showed site Mulde to be stronger contaminated than Cumlosen and within the Rhine segment site Koblenz was cleaner than site Iffezheim. Most important for the order achieved are Cu ($W = 14$) and Se ($W = 10$); of no importance is Hg ($W = 0$). The stability value (P(*IB*)) was reduced from 0.97 to 0.83 by the rounding of the data, but is still near that of an anti-chain.

Hasse diagram for the organic compounds

When the same scheme was applied to the sub-set of the five organic compounds, the improvement achieved by the rounding of the data was far less pronounced. The number of incomparabilities was reduced from 116 to 114; the number of levels remained 4 and that of the isolated objects re-

mained 2 for both, the raw data and the rounded data. This remarkably lesser improvement can be attributed to the fact that the organic compounds are stronger correlated with each other than the inorganic compounds with each other and thus are more often increased simultaneously. The maximum chain length was 4.

In terms of inter-fluvial segments, both Bimmen and Iffezheim were recognised to be more contaminated than Koblenz (River Rhine), additionally Bimmen was more strongly contaminated than Weil. Within the Elbe segment the order Zehren > Prossen > Cumlosen was found. Elbe tributary Mulde was more strongly contaminated than Cumlosen. Güdingen and Rehlingen from river Mosel proved to be incomparable.

In general, the Elbe sites and tributaries appeared in the upper part of the diagram, whereas the river Rhine sites appeared in the lower half, allowing a tentative assessment of the contamination for both streams. The most important substances for the partial order shown in Fig. 7 are HCH and PCB-sum (for each: $W = 17$). The stability P(IB) is 0.73, indicating that the order achieved is still near an anti-chain.

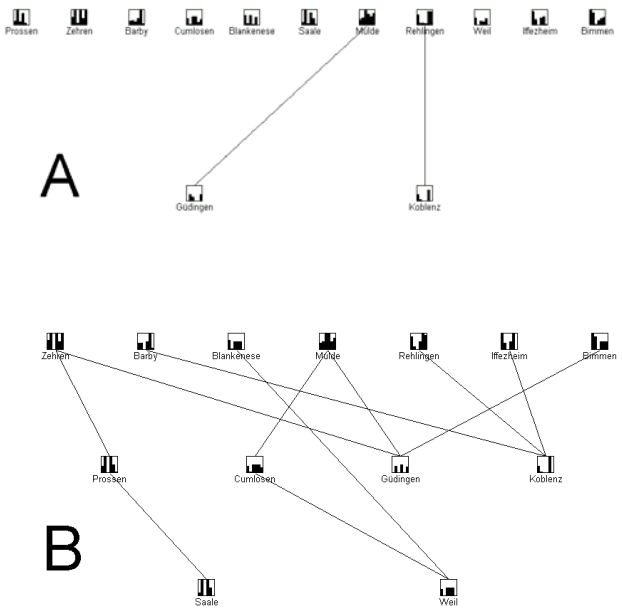

Fig. 6. Hasse diagram for 6 inorganic compounds; A: raw data;
B: rounded data

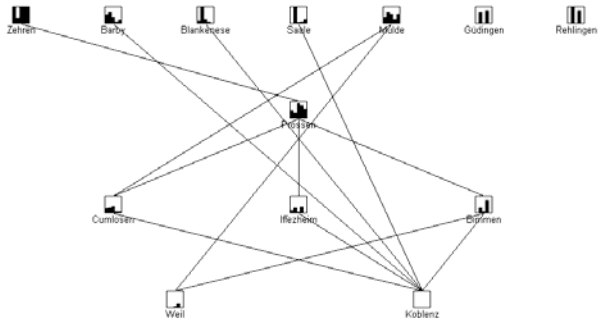

Fig. 7. Hasse diagram for 5 organic compounds; rounded data

From partial to total order: Calculation of averaged ranks

Though some improvement have been achieved, the Hasse diagrams or partially ordered sets shown in Fig.'s 6B and 7 are not entirely satisfactory. To overcome this, linear extensions can be constructed for each partially ordered set of objects. A linear extension is a total order where all comparabilities will be conserved (Brüggemann et al. 2004; see also chapters by Brüggemann and Carlsen, p. 61 and Carlsen, p. 163). This will yield a number of linear extensions for a given partially ordered set, taking all possible locations of the incomparable objects into account. Take e.g. the Hasse diagram for the inorganic compounds (Fig. 6B). To simplify the matter we will only consider the subgraph consisting of Zehren, Prossen, Saale and Güdingen. The possible linear extensions for this subgraph, when all comparabilities are preserved, are as follows.

Zehren > Güdingen > Prossen > Saale
Zehren > Prossen > Güdingen > Saale
Zehren > Prossen > Saale > Güdingen

Please note, that in all cases the orders Zehren > Prossen > Saale and Zehren > Güdingen, respectively, have been preserved. In a likewise manner all possible linear extensions for any partially ordered set can be constructed. Most unfortunately the number of linear extensions will increase exceedingly with the number of objects and incomparabilities in the given set. In case of Fig. 6B not less than 13,414,830 linear extensions can be constructed. When all possible linear extensions are known the most prob-

able rank for each object can be calculated, thus yielding a totally ordered set (Brüggemann et al. 2004) without the use of weights or assuming a rank index function. Since, however, the calculation of all possible linear extensions will require an impracticably long computing time this approach cannot be taken in most cases. To solve this obstacle, a method based on Monte Carlo calculations was developed (Sørensen et al. 2001, Lerche et al. 2003) and Brüggemann et al. (2004) gave a simple and straightforward equation for the calculation of averaged ranks which are very close estimations of the 'true' ranks obtained by the examination the cumbersomely calculated full set of all linear extensions (Brüggemann et al. 2004), counting from bottom to top.

$$R_{kav(x)} = (S_{(x)} + 1) \cdot (N + 1) / (N + 1 - U_{(x)}) \quad (3)$$

where $R_{kav(x)}$ is the averaged rank of object x, $S_{(x)}$ is the number of successors of x (objects ranked below x), $U_{(x)}$ is the number of objects incomparable to x and N is the total number of objects in the set. It must be mentioned that an object with many successors will tend to get a higher averaged rank than one, which has instead many predecessors. Therefore the partial order must be well justified, and only with significant comparabilities and incomparabilities the averaged rank approach is a reasonable one (Brüggemann et al. 2004).

Application of the averaged ranks approach to the partially ordered set of inorganic compounds (Fig. 6B) yielded the totally ordered set depicted in Fig. 8A. In words, the order is:

{Zehren, Mulde} > {Barby, Blankenese, Rehlingen, Iffezheim, Bimmen} > Prossen > Cumlosen > Saale > {Güdingen, Koblenz, Weil};

corresponding to the averaged ranks 2.8, 4.6, 7.0, 8.4, 10.5 and 11.2, respectively. (Please note that in Fig. 8 equivalent objects are 'hidden beneath' the first object of each equivalence group.)

Likewise a totally ordered set can be obtained for the organic data (Fig. 8C). The resulting order is:

Mulde > {Zehren, Prossen} > {Saale, Barby, Blankenese} > {Rehlingen, Bimmen, Güdingen} > Iffezheim > Cumlosen > Weil > Koblenz;

corresponding to the ranks of 2.8, 3.5, 4.6, 7.0, 8.4, 9.3, 11.6 and 12.7, respectively.

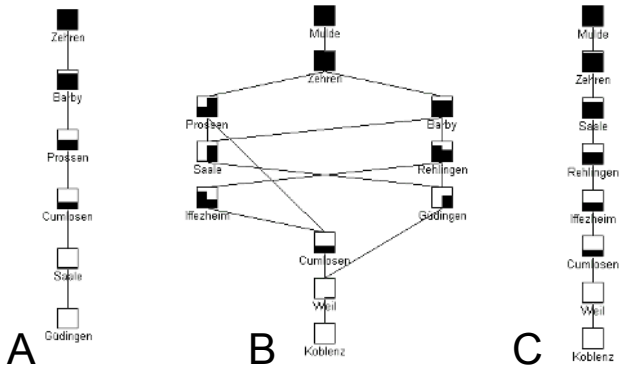

Fig. 8. Complete order sets for the inorganic (A), the organic compounds (C), and a synopsis of both (B), achieved by the calculation of averaged ranks. Note that there are equivalent objects (see text)

Synopsis of inorganic and organic data

We have now two different totally ordered sets; each one for the organic and the inorganic compounds which parallel in same respects. In both cases, the sampling sites of Mulde, Zehren, Barby and Blankenese appear at or near the top and can thus be considered to represent the mostly contaminated sites. In contrast, Koblenz, Weil and Cumlosen are assigned to the bottom end of the order. When defining the averaged ranks of both groups of compounds as new attributes the order as shown in Fig. 8B can be obtained. The stability (P(*IB*)) of this poset is 0.14, indicating nearness to a chain. This poset has 8 levels, only one minimal and one maximal object, and the number of comparabilities is 69 (the maximum number). The corresponding total order, achieved by the calculation of averaged ranks, is Mulde > Zehren > {Barby, Blankenese} > Prossen > {Rehlingen, Bimmen} > Saale > Iffezheim > {Güdingen, Cumlosen} > Weil > Koblenz.

This means that the Elbe sites are stronger contaminated than those of river Rhine, if all 11 compounds are considered. Within river Elbe and its tributaries we found the order: Mulde > Zehren > {Barby, Blankenese} > Prossen > Saale > Cumlosen. The order for the river Rhine sites is Bimmen > Iffezheim > Weil > Koblenz. Since none of these orders corresponds to the course of the streams we must assume a complex interplay of contamination and subsequent dilution and degradation processes. Coming from the Czech Republic the Elbe is medium strongly contaminated. Up to the mouth of tributary Mulde this contamination is strongly increased, but tributary Saale, bringing cleaner water, will dilute this contamination.

Barby, located at the mouth of tributary Saale, and Cumlosen have decreasingly less contamination, whereas Blankenese (located in close proximity and down-stream from the Port of Hamburg) shows a new increase of contamination. Whether this is due to local inlet of contaminants or the sampling time (after the flood) cannot be decided here. The river Rhine, too, starts with medium contamination, which is increased between Weil and Iffezheim, whereas the contamination is diluted (or otherwise reduced) in down-stream direction until to Koblenz. But the industrial regions located at the Lower Rhine seem to increase the contamination anew, so that location Bimmen, near the German-Dutch border, is the most strongly contaminated site of the river Rhine. River Saar, represented by only two sampling sites, shows the order Rehlingen > Güdingen; an order which reverses the direction of the stream. As already mentioned above, this is due to an input of the industrial zone of Dillingen located ca. 1 km up-stream from the Rehlingen weir.

Conclusions and Outlook

The construction of a partial or even total order is a valuable means for the evaluation of biomonitoring data. A critical point is the number of incomparabilities. To reduce this number the data should be pre-processed. For this contribution pre-processing was done by rounding the data according to the magnitude of the median. A somewhat more sophisticated approach could be a rounding procedure, which considers the measurement uncertainty, thus expressing the data as multiples of the measurement uncertainty, taking into account different uncertainties for different chemicals. Another possibility for pre-processing is the cluster analysis (Luther et al. 2000), resulting in a smaller number of objects, which can be used as input for the Hasse Diagram Technique. Disadvantage of this approach is, however, the loss of information.

It should be stressed that there are more possible applications for posets in biomonitoring. One promising application for the Hasse Diagram Technique is the ranking of chemicals. When chemicals are expressed as fractions of the corresponding guide values or limit value, a partial order as shown in Fig. 9 will result. Prior to the ranking, the data of six chemicals obtained from the sampling sites of river Elbe, including the two tributaries Mulde and Saale, have been transformed according to guide values of the EC. Then the data table was transposed so that the rows became columns and vice versa, thus making the chemicals to objects and the sam-

pling sites to their attributes. The resulting partial order reveals that mercury is the most relevant chemical, whereas γ-HCH is of least importance.

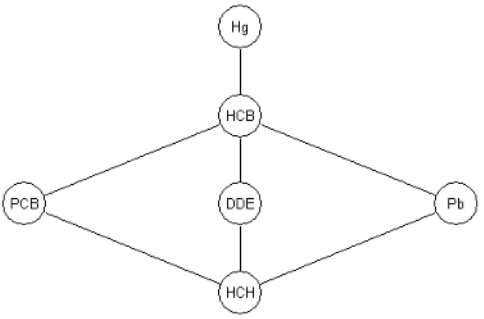

Fig. 9. Ranking of chemicals for the sampling sites of river Elbe

Analysis of the temporal course of contaminations is another example for further applications (Fig. 10). Based on essentially the same chemicals it seems that human beings experience an increasingly better protection than wild live. For bream sampled from 1997 to 2003 at the barriage weir of Rehlingen the improvement of the environment did not parallel the time course (Fig. 9A), since the order of contamination is, obviously, 1999 > {1997, 2001} > 2000 > {2002, 2003} > 1998. In contrast, samples taken from students of the Münster University show a clear parallel between time and bettering of the body burden. This would mean, that man is better protected against the chemical released into the environment by him than wild life is. But since man is an inseparable part of the environment contaminations will always backfire on the mankind.

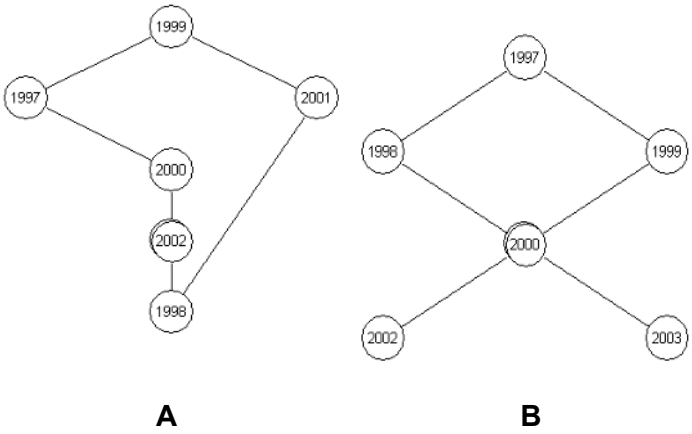

Fig. 10. Temporal course of contamination for environmental samples (A: bream from sampling site Rehlingen) and human samples (B: students of the University of Münster)

References

Bond CE (1979) Biology of fishes. Philadelphia, Saunders

Brüggemann R (2001a) Einsatzmöglichkeiten des Konzepts partiell geordneter Mengen für Bewertungen in ökosystemarem Kontext. In. Steinberg CEW, Brüggemann R, Kümmerer K, Ließ M, Pflugmacher S, Zauke GP (Hrsg.). Streß in limnischen Ökosystemen. Neue Ansätze in der ökotoxikologischen Bewertung von Binnengewässern 173-185

Brüggemann R (2001b) Ansätze zur vergleichenden Bewertung unter spezieller Berücksichtigung der Hassediagrammtechnik. In. Neumann-Hensel H, Ahlf W, Wachendörfer V (Hrsg.): Nachweis von Umweltchemikalien. Auswerte- und Interpretationsmethoden für Toxizitätsdaten aus einer ökotoxikologischen Testkombination. Initiativen zum Umweltschutz 29:27-46

Brüggemann R, Sørensen PB, Lerche D, Carlsen L (2004) Estimation of averaged ranks by a local partial order model. J Chem Inf Comput Sci 44:618-625

Brüggemann R, Halfon E (1997) Comparative analysis of nearshore contaminated sites in Lake Ontario: Ranking for environmenatl hazard. J Environ Sci Health A32:277-292

Brüggemann R, Steinberg C (2000) Einsatz der Hassediagrammtechnik zur vergleichenden Bewertung von Analysendaten – am Beispiel von Umweltuntersuchungen in den Regionen Baden-Württembergs. Analytiker Taschenbuch 21:3-33

Brüggemann R, Münzer B, Halfon E (1994) An algebraic/graphical tool to compare ecosystems with respect to their pollution – the German river 'Elbe' as an example – 1: Hasse-diagrams. Chemosphere 28:863-872

Brüggemann R, Bücherl C, Pudenz S, Steinberg CEW (1999) Application of the Concept of Partial Order on Comparative Evaluation of Environmental Chemicals. Acta hydrochimica et hydrobiologica 27:170-178

Bücherl C, Brüggemann R, Halfon E (1995) Hasse, Ein Programm zur Analyse von Hasse-Diagrammen, Projektgruppe Umweltgefärdungspotentiale von Chemikalien, ISSN 0721-1694, GSF-Bericht 19/95

Buxbaum F (1958) Cactus Culture: Based on Biology, Blandford Press, London

Carson R (1962) The Silent Spring, Greenwich, Connecticut, 1962

Conquest L (2000) Environmental monitoring: Investigating associations and trends. In: Statistics in Ecotoxicology (Edt. T. Sparks), Chichester, New York

FEA – Federal Environmental Agency (1999) Umweltprobenbank des Bundes – Ausgabe 1999. Ergebnisse aus den Jahren 1996 und 1997, Texte 61/99, Werbung & Vertrieb, Berlin

FEM – Federal Environment Ministry (2000) Environmental Policy: Environmental Specimen Bank. Concept, Federal Environmental Agency, Berlin

Hasse H (1967) Vorlesungen über Klassenkörpertheorie, Universität Marburg. Physica-Verlag

Helm D (2003) Bewertung von Monitoringdaten der Umweltprobenbank des Bundes mit der Hasse-Diagramm-Technik. UWSF – Z. Umweltchem Ökotox 15:85-94

Lerche D, Sørensen PB (2003) Evaluation of the ranking prob-abilities for partial orders based on random linear extensions. Chemosphere 53:981-992

Lerche D, Sørensen PB, Brüggemann R (2003) Improved estimation of the ranking probabilities in partial orders using random linear extensions by approximation of the mutual ranking probability. J Chem Inf Comput Sci 43:1471-1480

Luther B, Brüggemann R, Pudenz S (2000) An approach to combine cluster analysis with order theoretical tools in problems of environmental pollution. MATCH 42:119-143

Sørensen PB, Lerche DB, Carlsen L, Brüggemann R (2001) Statistically Approach for Estimating the Total Set of Linear Orders. Pudenz S, Brüggemann R and Lühr HP 14/2001, 87-97. 2001. Berlin, IGB Berlin. Berichte des IGB

Sørensen PB, Vorkamp K, Thomsen M, Falk K, Møller S (2004) Persistent organic Pollutants (POPs) in Greenland environment - Long-term temporal changes and effects on eggs of a bird of prey. National Environmental Research Institute - NERI Technical Report 509 pp 1-126

von Storch H, Costa-Cabral M, Hagner C, Feser F, Pacyna J, Pacyna E, Kolb S (2003) Four decades of gasoline lead emissions and control policies in Europe: A retrospective assessment. The Science of the Total Environment (STOTEN) 311:151-176