Order Theoretical Tools for the Evaluation of Complex Regional Pollution Patterns

Rainer Brüggemann,[†] Gerhard Welzl,[‡] and Kristina Voigt[§]

Department of Ecohydrology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany, GSF - National Research Center for Environment and Health, Institute of Development Genetics, 85764 Neuherberg, Germany, and GSF - National Research Center for Environment and Health, Institute for Biomathematics and Biometry, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

Received March 24, 2003

The mathematical and statistical evaluation of environmental data gains an increasing importance in environmental chemistry as the data sets become more complex. It is inarguable that different mathematical and statistical methods should be applied in order to compare results and to enhance the possible interpretation of the data. Very often several aspects have to be considered simultaneously, for example, several chemicals entailing a data matrix with objects (rows) and variables (columns). In this paper a data set is given concerning the pollution of 58 regions in the state of Baden-Württemberg, Germany, which are polluted with metals lead, cadmium, zinc, and with sulfur. For pragmatic reasons the evaluation is performed with the dichotomized data matrix. First this dichotomized 58×13 data matrix is evaluated by the Hasse diagram technique, a multicriteria evaluation method which has its scientific origin in Discrete Mathematics. Then the Partially Ordered Scalogram Analysis with Coordinates (POSAC) method is applied. It reduces the data matrix in plotting it in a two-dimensional space. A small given percentage of information is lost in this method. Important priority objects, like maximal and minimal objects (high and low polluted regions), can easily be detected by Hasse diagram technique and POSAC. Two variables attained exceptional importance by the data analysis shown here: TLS, Sulfur found in Tree Layer, is difficult to interpret and needs further investigations, whereas LRPB, Lead in Lumbricus Rubellus, seems to be a satisfying result because the earthworm is commonly discussed in the ecotoxicological literature as a specific and highly sensitive bioindicator.

1. INTRODUCTION

Visualization of environmental pollution is essential for evaluation studies. Evaluations with ecotoxicological background frequently do not have a causally justified target function. Very often several aspects have to be considered simultaneously, for example, several chemicals entailing a data matrix with objects (rows) and variables (columns). In this paper a data set concerning the pollution of 58 regions in the state of Baden-Württemberg, Germany, which are polluted with metals lead, cadmium, zinc, and with sulfur is given. As the objects are gathered to form a set, the partially ordered set, we also use the notation “element”, especially in the context of order theory.

Different methods can be applied to analyze complex data sets. So it is possible to describe regional pollution by deterministic models. However, due to limited data availability this type of model tends to meet its bounds quite soon. Another possibility is to apply the theory of partially ordered sets to evaluations. The above-mentioned data set has been analyzed by using the Hasse diagram technique within the scope of a research project funded by the Landesanstalt für Umweltschutz in the German federal state of Baden-Württemberg. The results of this research project are published as a report.[1] Further publications concerning this project and the theory on the Hasse diagram technique can be found in international journals.[2−5] In the publication Brüggemann et al.[5] the most important theoretical aspects are outlined, exemplified by the analysis of the degraded sediments of Lake Ontario, Canada.

Analysis and evaluation of environmental pollution can also be performed by applying methods of environmental statistics. Explorative methods and graphical techniques are one way to improve the exchange of information. The development of graphical surfaces for statistical software packages is presently very much appreciated using PCs and efficient graphic software. In statistical analysis variance and distance based concepts are applied. In modern statistical techniques also order relations are taken into account (see SYSTAT[6]). For example, an order relation analysis is supported by methods of dimension reduction. An effective tool for this kind of analysis is the method of POSAC (Partially Ordered Scalogram Analysis with Coordinates) which is part of the facet theory.

The results found by the application of the Hasse diagram technique will be compared with those results found by facet theory.

We already performed such an approach of comparing multivariate statistical methods with Hasse diagram technique applied on a data set of polluted regions (herb layer) in Baden-Württemberg on the occasion of the 1. Workshop on Order Theoretical Tools in Environmental Sciences which was held on November 16th in Berlin, Germany.[7−9]

* Corresponding author e-mail: brg@igb-berlin.de.
[†] Leibniz-Institute of Freshwater Ecology and Inland Fisheries.
[‡] Institute of Development Genetics.
[§] Institute for Biomathematics and Biometry.

10.1021/ci0300083 CCC: $25.00
Published on Web 08/30/2003

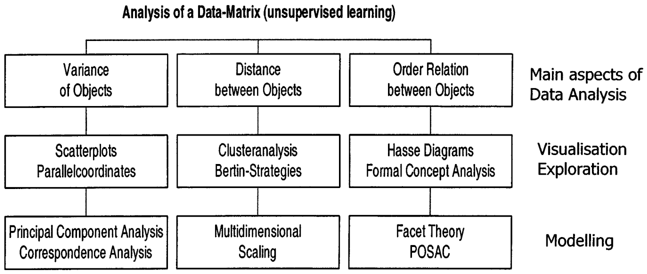

Figure 1. Main methods used in chemometrics/biometrics.

In this current paper we want to show the limits of the two methods—facet theory and Hasse diagram technique—and to demonstrate the symbiotic aspects of the methods.

2. MULTIVARIATE METHODS

In the field of chemistry and environmental sciences the terms chemometrics and environmetrics are used for scientific disciplines that use mathematical, statistical, and other methods. Those methods employ formal logic to provide a maximum of chemical or environmental information.[10] In the same context the term biometrics is understood as the discipline to apply mathematical methods, especially statistical methods in biological and related disciplines, for example, environmental sciences and agricultural sciences.[11] From a methodological point of view one can discuss three aspects within multivariate techniques:

1. Unsupervised versus supervised learning
2. Data exploration by (a) Variance analysis, (b) Distance analysis, and (c) Order analysis and
3. Modeling, that is, the kind of data reduction.

In Figure 1 the main applied methods of the analysis of environmental and chemical data matrices s are outlined.

To analyze the complex regional pollution pattern we concentrate on one given (measured or estimated by simulation models) data matrix, that is, here we apply an unsupervised learning. To perform an evaluation we apply 2c (see Figure 1), the order analysis. Finally the point of data reduction is on the focus of the whole following paper.

2.1. Order Theoretical Approach − Hasse Diagram Technique. The basis of the Hasse diagram technique is the assumption that a ranking can be performed while avoiding the use of an ordering index.[12] In our application, Hasse diagrams present information not only on ranking, but, most of all, they show whether the criteria, characterizing the objects, lead to ambiguities in the ranking procedure: For example, an object might be ranked higher according to one criterion but lower according to another. These two objects are not ordered. This ambiguity—which is important to know for further applications—is not evident when we use an index for ranking, but it is immediately evident in a Hasse diagram. Hasse diagrams are extremely useful if several criteria are given to decide which objects are priority objects, and which objects are suboptimal and what is the reason for that. Here the polluted regions in the German state of Baden-Württemberg represent the objects, and the environmental matrices polluted by different metals and sulfur the attributes (criteria).

In this paper we investigate a method of Discrete Mathematics, namely the Hasse diagram technique to extract ranking information out of the data matrix. The ranking of a set of objects does not only depend on the numerical values but also even more on the choice of attributes (criteria). Typical results of this analysis are two matrices, D and W, that identify the main features of the structure of Hasse diagrams and quantify the influence of criteria on the ranking. As other typical results dimension analysis[5] and probability concepts (to be published) should be mentioned.

Hasse diagrams visualize the order relations within posets (*p*artially *o*rdered *sets*). Two objects x, y of a poset are ordered if all scores of x are less or equal than those of y. Hasse diagrams are oriented graphs (acyclic digraphs). A digraph consists of a set *P* of objects drawn as small circles in Hasse diagrams. In our applications the circles near the top of the page (of the Hasse diagram) indicate objects that are the “most polluted” objects according to the criteria used to rank them: These objects have no predecessors (they are not “covered” by other objects) but successors and are called maximal objects. Those objects found in the lowest part of the diagrams and are only connected by lines in the upward direction are called minimal objects. These are the “least polluted” regions according to the given data matrix. There generally may also be isolated objects which have neither successors nor predecessors.

The theoretical background of the Hasse diagram technique as well as many applications are presented in a yearly workshop “Order Theoretical Tools in Environmental Sciences”. Main theoretical articles on the development of the partial order ranking methods are given by Halfon and Brüggemann.[13,2,14,12] As shown in a publication of Lerche[15] contrary to other evaluation and decision support techniques the Hasse diagram technique is not dependent on knowledge beyond that given in the data matrix. Therefore this kind of application of partial order theory can also be seen as a tool for data analysis. Thus the evaluation by Hasse diagram technique also supports to some extent the exploration of structures in the data matrix and can be considered as a parameter-free method.

Partially ordered sets are not only useful in the application we are intending here but also of general relevance in chemistry. A good and rather recent overview about partial order in general chemistry can be obtained from the special issue of Comm. in Mathematical and Computer Chemistry, edited by Klein.[16] Valuable information about concepts in partial order theory is given by Klein.[17] There are several possibilities to define order relations. One attractive example is given by Randic.[18]

2.2. Order Theoretical Approach with Data Reduction − POSAC. The object of statistics is information, the objective is the understanding of information in data. A statistical problem is characterized by variability and uncertainty. The importance of multivariate statistics in chemistry and environmental sciences is emphasized in two main textbooks, one by Stoyan, called Environmental Statistics,[19] and one by Einax, named Chemometrics in Environmental Analysis.[10] An overview article on the chemometrical and ecometrical research of the research group Biostatistics at the GSF - National Research Center for Environment and Health, Institute of Biomathematics and Biometry is given by Welzl.[9]

Statistical methods can also be based on order information, for example, the rank correlation coefficients of Kendall or Spearman. Monjardet[20] compared these two coefficients and published some useful relationships concerning them. Order

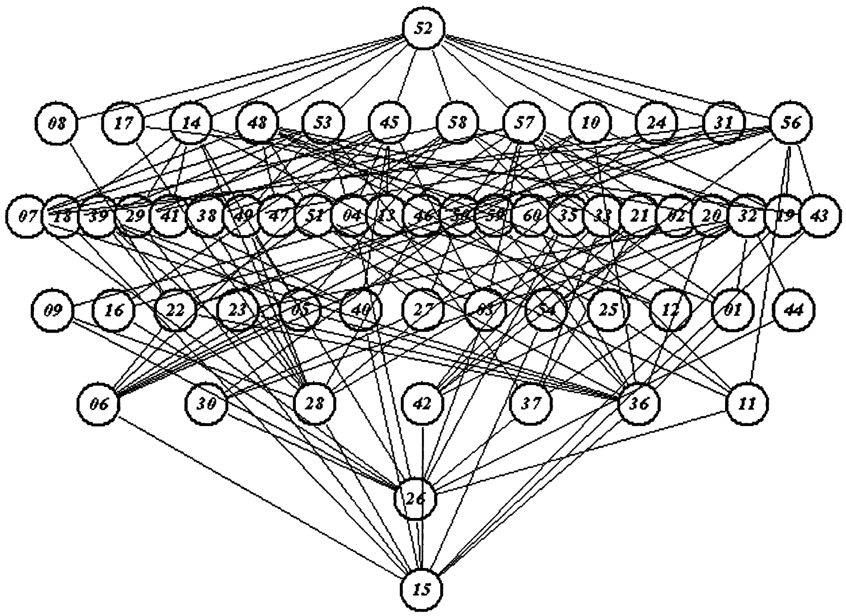

Figure 2. Hasse diagram for 58 objects with 13 variables (dichotomized variables).

information can also be found in the textbook Statistical Inference based on Ranks.[21] There are many statistical approaches of condensing a data matrix by the creation of new variables. This process—called ordination—is often used to visualize relationships in two dimensions based on the first two variables. These new variables, which are derived from the original variables, are constructed to optimize some specific criteria. For example, principal component analysis creates new axes to explain as much as possible of the variance of the data matrix. This idea can be applied when order relations (comparability as well as incomparability) are considered as the essential aspect of the data. This method—construction of new axes which presents correctly as much as possible of the order relations—is called Partially Ordered Scalogram Analysis with Coordinates (POSAC). POSAC is integrated in the program package SYSTAT [6] under the feature of statistics, data reduction. For a better interpretation of the new axes correlations between old and new variables can be calculated.[22]

The POSAC method is applied on data matrices in environmental sciences and chemistry. Welzl examined regions polluted with metals.[8] Pesticide Internet resources were analyzed with chemical and environmental evaluation criteria by Voigt et al.[23] As variance-and distance-based statistical methods intend to explore the data matrix, the aspect of evaluation (i.e., comparison of objects with respect to some protection aims) is of minor importance. We hypothesize that POSAC will combine both aspects: evaluation and exploration of data structures. This hypothesis is at the heart of our paper.

3. APPLICATION OF HASSE DIAGRAM TECHNIQUE ON MEDIAN VALUES FOR DATA MATRIX 58 OBJECTS AND 13 VARIABLES

The given data set shows a matrix concerning the pollution of 58 regions in the state of Baden-Württemberg, Germany with metals lead, cadmium, zinc, and with sulfur. The set of regions is called P, and the set of 13 attributes is called the information base IB_{13}. It comprises a rather complex data matrix due to the different values for the concentration of the pollution. We therefore calculate the median, and then the variables are dichotomized and their range is 0 (variable less median) or 1 (variable not less than median).

The Hasse diagram of this data set is shown in Figure 2.

Looking at the diagram in Figure 2 the following conclusions can be drawn: There are no objects having the same sequence of 0's and 1's, respectively, that is, having identical binary patterns. We only find 1 maximal object (object no. 52) and 1 minimal object (object no. 15). In the case of the appearance of only 1 maximal object, we speak of the greatest element, in the case of only one minimal object of the least element.[14]

No isolated objects are found which means that every object is comparable with at least one other object. This Hasse diagram has a structure with seven levels, the number of comparabilities is only 314, and the number of incomparabilities 2678. The analysis of data representation plays an important role in the Hasse diagram technique. See, for example, the combination of cluster analysis with HDT[24] or the examination of robustness.[2]

The Hasse diagram (Figure 2) is rather complex. Therefore one may look for order ideals, generated by some key elements.[5] Order ideals generated by an element x are defined as follows: $O(x) := \{y \in P: y \leq x\}$. To get the most informative order ideals, it may be a good policy (i) to avoidkey elements whose order ideals are in a subset relation to an order ideal, already generated and (ii) to examine the matrix D. The entry D_{ij} is equal to the number of objects ordered below object i and object j simultaneously. By (i) it is clear that the greatest element should not be selected, because its order ideal will not simplify the discussion because it is the whole poset. Therefore it is better to select other elements, for example, from the sixth level (counted from the bottom). However which ones should be selected? There are 12 elements in the sixth level which are covered by element number 52, the greatest element. One may visualize all 12 order ideals by Hasse diagrams, but this would be not very efficient. In Figure 3 a fictitious example is shown, why:

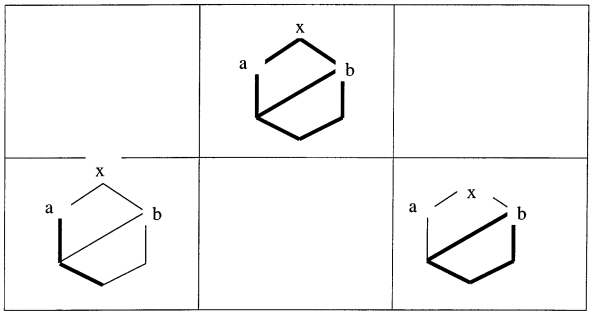

Figure 3. Example of the visualization of a poset and two order ideas O(a) and O(b). The drawing of the Hasse diagrams is slightly simplified.

The information won by O(a) is almost given by O(b) (see Figure 3). If appropriate pairs of key elements are to be selected, then 66 pairs are available. Which pair will exhibit the most information, that is, will show different objects of *P* and order ideals containing approximately the same number of elements. To find a way out, the matrix D can be examined. The software WHASSE allows for the selection of key elements, for which the matrix D is to be calculated. As a next step those key elements should be selected, for which $D_{ii} \approx D_{jj}$ and as large as possible and at the same time having a small D_{ij}. By this policy two order ideals are found and shown in Figure 4.

4. APPLICATION OF REDUCTION TECHNIQUES (POSAC)

To reduce the complexity of the dichotomized data matrix again, we apply the so-called POSAC method. This procedure reduces the data by a construction of a map φ:

$$\varphi\text{:}\ \ (P,IB_{13}) \rightarrow (P, IB_2)$$

IB_2 is the new set of attributes which implies the two-dimensional space embedded in R^2.

The map φ is constructed by optimizing the proportion of order relations of (P, IB_{13}) which are correctly represented in (P, IB_2). Note that this map is intended to be as far as possible an order preserving map. The concept of order preserving maps is an important tool in partial order theory. Therefore we give a brief explanation: Two sets A, B may be related by a map φ: φ: A $\rightarrow$ B. If any two elements of A, a, b are ordered, for example, a $<$ b and at the same time φ(a) $<$ φ(b), then φ is said to be order preserving.

We apply the so-called POSAC method which can be found in the program SYSTAT 10. As mentioned above the Partially Ordered Scalogram Analysis with Coordinates (POSAC) method reduces the data matrix in plotting it in a two-dimensional space, and a given percentage of information is lost. The new attributes are called LOV (1) and LOV (2) and based on these two attributes, that is, on IB_2 a Hasse diagram similar but clearly not identical to that of Figure 1 could be found.

In this example 76.1% of the partial order relations of the dichotomized data matrix are correctly represented that means 23.9% of the original information is lost (see Figure 5). For example the maximal object of the Hasse diagram region 52 can be found on the upper right of the POSAC plot. The minimal object of the region 15 is situated in the very lower left side of the plot.

Obviously, the number of specific objects, like the maximal and minimal object, are the same in both diagrams. This means that the "good" and "bad" position of these objects is evident. However, as the more detailed analysis shows there are slightly more comparabilities (namely 366 in (P, IB_2) compared to 314 in (P, IB_{13})) which then support a more detailed ranking of the objects.

It can be stated that on the one side the diagram of Figure 5 does not reveal any new structure (as one might assume—taken the 76% correctness into account), on the other side the possibility is opened that for the two dimensions an interpretation can be found. These two dimensions can be considered as in PCA as latent order variables (LOVs, in figures also called "dim" to account for dimensional aspect of the procedure). In that case the loadings are important because they support the interpretation. For example: To explore the influence of the attributes on the whole analysis, we perform a correlation analysis of the two latent order

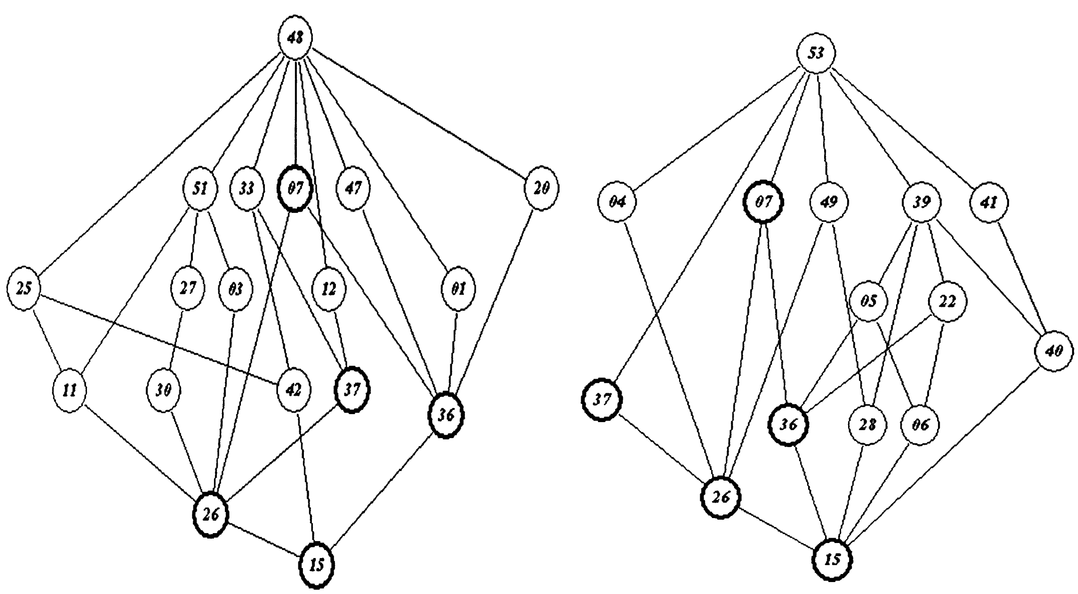

Figure 4. Two order ideals, selected by examination of the matrix D ($D_{53,53} = 15$, $D_{48,48} = 18$, $D_{53,48} = 5$). Note that these automatically generated Hasse diagrams are slightly graphically modified to ensure clear readability.

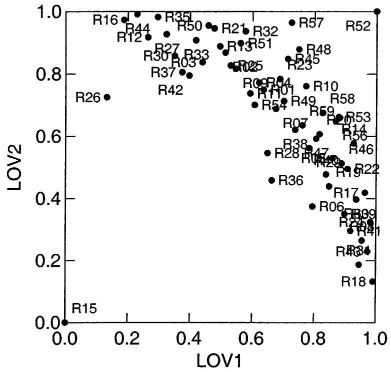

Figure 5. POSAC plot of the data matrix 58 × 13 (median values).

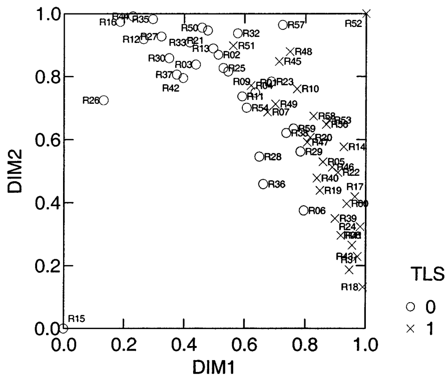

Figure 6. Scatter plot of the latent variables considering TLS.

variables given in the POSAC plot. This is done by applying the analyses of variance (ANOVA) with LOV (1) and LOV (2) as dependent variables and the (dichotomized) attributes as factors.

The following F-statistics are calculated:

LOV (1): TLS: 67.596, HLS: 22.768, LRPB: 19.346, TLPB: 17.522

LOV (2): LRPB: 85.043, TLS: 14.222, LRCD: 8.502, HLS: 5.6620

The attribute TLS (Sulfur found in Tree Layer) is high correlated with LOV (1). The so-called polar item for LOV (1) is TLS. The attribute LRPB (Lead in Lumbricus Rubellus) is high correlated with the LOV (2), that is, the polar item for LOV (2) is LRPB. It may be worth noting that—following these results—a new IB_2' could be generated, with only TLS, and LRPB as binary attributes. The resulting Hasse diagram, however, will represent an ordered set with four equivalent classes, which is the image of an order preserving map from (P,IB) to (P,IB_2'). Therefore it becomes clear that by LOV(1) and LOV(2) indeed a new aspect is introduced.

To visualize the high correlation of LOV (1) with TLS the scattered plot is given in Figure 6.

As it can easily be seen the data points marked due to TLS = 1 are located at higher values of LOV (1).

Applying the same procedure to the attribute LRPB, the scattered plot in Figure 7 is found.

5. THE ROLE OF SIMPLIFICATION

The simplification by use of order ideals is that many order relations are omitted namely of those elements which do not belong to the specific order ideal.

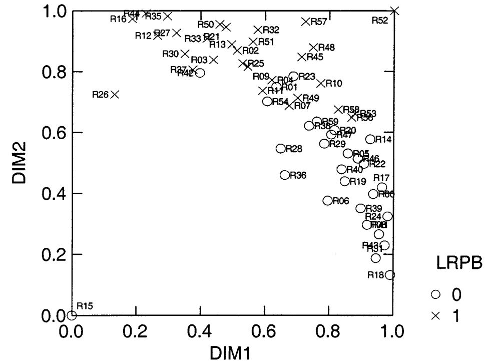

Figure 7. Scatter plot of the latent variables considering LRPB.

Now the evaluative structure can be made quite clear, but with the price that several diagrams (several order ideals) have to be discussed at once. The simplification by applying the POSAC procedure is the projection onto a two-dimensional plane which is often only possible in an approximate manner. The Hasse diagram technique can predict by its dimension analysis when exact results are to be expected.[5] The simplification by POSAC must be in some way consistent with that of the Hasse diagram technique: This can be easily seen if the order ideals and the Figure POSAC are considered more closely:

1. The order ideal $O(48)$ is located on the upper half of the plane where as the $O(53)$ is more located at the right side of the POSAC plot.

2. Corresponding to the demand that the two order ideals should be well separated the location of the regional sites is quite well separated too.

3. The common property of the elements of $O(48)$ can be deduced from the POSAC plot namely that they have a high value of LOV (2) and a considerable variation in LOV (1) which means that the elements of $O(48)$ are mainly elements of high values in LRPB and with a strong variation in TLS with the exception of the minimal element R15.

4. The common property of the elements of $O(53)$ can once again be deduced from the POSAC plot. They all have a high value in LOV (1), that is, TLS and a considerable variation in LOV (2), that is, LRPB (exception R15).

5. Clearly the approximative nature of POSAC leads in some cases to reverse orders, that is, in the original Hasse diagram (see Figure 2) the relation 39 > 22 is found, whereas corresponding to the POSAC analysis the reverse relation is found. This fact should lead to cautious interpretation if element-wise evaluation is wanted. However, this kind of discrepancy may be acceptable if general structures within our data set are to be explored. If this finding is to be discussed more closely it is necessary to take into consideration that there are two steps of data handling: A. original data (with noise) are transformed into dichotomized attributes and B. the dichotomized attributes are projected to the POSAC plane of LOV (1) and LOV (2).

6. DISCUSSION AND OUTLOOK

(1) Concerning the whole set of objects: The region 52 is the city of Schwetzingen in the German state of Baden-Württemberg which is the highest polluted region in the two different comparative evaluation analyses Hasse diagram technique and Partially Ordered Scalogram Analysis with

Table 1: Data Matrix of 58 Regions Polluted with Environmental Chemicals in Different Matrices (Original Data)[a]

region	HLPb	HLCd	HLZn	HLS	TLPb	TLCd	TLZn	TLS	MLPb	MLCd	MLZn	LRPb	LRCd
06	1	0.07	29	1750	0.6	0.04	22	1620	11.1	0,207	31	2	4.4
08	1.5	0.07	27	1750	0.9	0.1	33	1890	19.8	0.361	55	2.3	6.5
07	1.2	0.09	28	1600	0.9	0.07	22	1750	14.4	0.304	41	4.5	5.3
17	0.6	0.06	36	1820	1.1	0.1	39	1800	12.77	0.295	63.4	2.9	4
09	0.09	0.2	850	580	0.1	0.12	30	1070	17.5	0.309	45.4	23.5	7
16	1	0.12	32	1520	0.5	0.23	34	1570	13.4	0.36	50.5	5.2	6
22	1	0.03	28	2150	0.9	0.03	43	1780	14.5	0.312	41.3	2.2	3.5
18	0.5	0.42	28	4030	1	0.1	32	1700	11.3	0.346	46.7	2.4	2.9
30	0.8	0.08	27	1610	0.8	0.09	26	1530	14.25	0.347	52.8	10.2	5.1
23	1.1	0.04	42	2000	1.1	0.12	36	1620	17.1	0.265	56	4.1	5.8
15	0.9	0.1	24	1670	0.6	0.08	23	1510	13.7	0.362	43.2	2.6	3.8
14	1	0.17	34	1830	0.7	0.14	40	1790	12.04	0.591	41.4	2	12.5
05	1.1	0.1	32	1990	0.9	0.08	26	1890	14.13	0.424	45.2	4.3	9.1
28	0.9	0.05	34	1670	0.5	0.06	29	1310	11.5	0.277	36.1	1.8	5.9
39	1	0.1	38	1740	0.9	0.08	42	2090	16.7	0.417	40.9	3.3	10.3
40	0.7	0.06	34	1770	0.6	0.04	28	1730	11.2	0.367	42.3	1.5	8.3
29	0.6	0.14	27	1680	0.4	0.13	31	1550	9.22	0.404	29.5	2.5	10.9
41	0.7	0.17	39	1840	0.8	0.08	26	1670	11.22	0.54	43.2	2.3	11.2
42	0.7	0.1	33	1690	0.4	0.08	28	1510	10.74	0.491	63	3.7	4.7
27	0.1	0.12	26	1600	1	0.11	28	1560	14.3	0.364	54.4	5.5	4.9
38	1.7	0.18	34	1720	0.4	0.12	34	1620	16.1	0.313	49.7	3.8	6.2
49	0.8	0.11	37	1680	0.8	0.04	42	1710	9.6	0.337	49.4	5.8	4.7
37	0.6	0.12	33	1580	0.5	0.06	26	1440	13.1	0.338	47.3	5.2	7.6
47	1.1	0.11	25	1650	1.1	0.13	24	1730	21.3	0.419	135	3.9	6.6
48	2.3	0.42	33	1600	0.7	0.1	21	1740	23.5	0.682	72.3	43.3	7.8
51	0.8	0.14	22	1640	1	0.09	25	1880	19.2	0.489	55.1	18.8	8.7
04	0.8	0.02	26	1790	0.6	0.01	32	1820	10.87	0.294	34.5	5.9	10.8
03	0.8	0.14	31	1710	0.6	0.29	28	1600	13.34	0.468	47.5	10.4	12.1
13	0.18	0.18	1160	350	0.1	0.09	34	1040	14.4	0.276	48.2	11.3	12.8
26	0.8	0.05	19	1620	0.4	0.03	21	1420	12.16	0.262	43.9	6.5	3.2
36	1.2	0.05	31	1570	0.7	0.04	26	1460	12.9	0.352	38.6	1.8	3.7
46	0.8	0.09	33	1680	0.8	0.04	34	1770	15.1	0.333	74	2.7	7.4
50	1.4	0.13	29	1730	0.9	0.12	29	1660	17.8	0.308	142	4.9	12
53	1	0.12	36	1750	0.7	0.07	37	1770	16.9	0.403	46.3	4.7	7.7
45	1.5	0.17	45	1780	0.6	0.09	29	1780	17.3	0.433	48.8	5.1	14.2
54	0.7	0.1	26	1750	0.7	0.09	24	1110	9.08	0.285	42.6	2.1	7.6
59	1.3	0.13	26	1470	0.7	0.13	31	1550	14.8	0.368	54.4	3.4	7.1
60	1	0.2	32	2160	0.8	0.09	27	2000	11.65	0.244	47.3	3.2	7.2
58	1	0.11	28	1980	0.7	0.07	29	1860	19.4	0.402	57.4	157.2	8.1
57	1.7	0.15	39	1850	0.9	0.11	36	1650	30.1	0.475	105	85.2	8.9
35	0.08	0.24	720	1960	0.2	0.08	27	830	17.8	0.409	62	87.3	5.7
33	0.16	0.26	800	530	0.1	0.07	23	860	51	0.54	84.8	122	14.3
25	0.9	0.09	35	1460	0.8	0.05	28	1430	18.4	0.592	74.9	469.5	11.4
12	0.16	0.23	910	1460	0.1	0.09	26	790	9.78	0.231	39.9	47.2	9.2
21	0.06	0.24	830	620	0.1	0.04	31	860	14	0.474	63.3	193.5	9
11	0.9	0.08	27	1720	0.7	0.05	24	1660	23.4	0.43	63.1	260.9	5.2
02	0.7	0.14	27	1770	0.7	0.18	28	1540	12.9	0.351	68.7	7.1	17.3
01	1	0.04	21	1540	1	0.08	25	1640	21.7	0.396	76.2	3.7	7.8
10	1	0.03	29	1780	0.7	0.13	34	1710	13.35	0.527	62.2	4.5	6.3
20	1.5	0.14	32	1730	0.9	0.12	27	1680	16.2	0.324	58.4	2.6	4.1
24	1.7	0.18	39	1740	0.6	0.03	41	1670	13.7	0.254	56.7	2.2	7.9
31	1.1	0.15	28	1740	1.2	0.16	36	1690	22.5	0.423	49.6	3.8	4
32	1.2	0.03	35	1820	1	0.15	47	1610	18.6	0.434	81.2	11.5	10.5
19	0.8	0.01	18	4030	0.6	0.01	21	3390	11.33	0.384	83.9	1.6	2.6
43	0.5	0.11	39	4030	0.8	0.01	28	1770	11.92	0.265	89.5	1.7	4.5
44	0.8	0.08	38	1800	0.6	0.07	28	1610	16.9	0.282	56.7	9.3	7.8
52	2	0.23	36	4030	1	0.13	30	2330	22.4	0.379	57.2	7.9	7.8
56	1	0.11	34	1970	0.8	0.08	30	1840	40.2	0.631	118	7.4	5.3

[a] Abbreviations: HL = herb Layer, TL = leaves tree layer, ML = moss layer, LR = *Lumbricus Rubellus*, Pb = lead, Cd = cadmium, Zn = zinc, S = sulfur.

Coordinates. This fact can be explained by the high traffic pollution in this area. On the other side the minimal object found in both evaluation methods, the region 15 is the area of Stockach around which practically no industry and little traffic is found. This is an area with a lot of woods and a low population density in this area.[1]

(2) Relationships among the attributes: Two variables got an exceptional importance by the data analysis shown here: TLS and LRPB. TLS is difficult and needs further investigations, whereas LRPB seems to be a satisfying result because the earthworm is commonly discussed in the ecotoxicological literature as a specific and highly sensitive bioindicator.

(3) Final statements concerning the applicability of Hasse diagram technique and POSAC as tools in data exploration: From the data analysis point of view one has to emphasize that the original data set had to be categorized. In this case the dichotomized values using the median are calculated. The method, Hasse diagram technique, focuses on individual

Table 2: Data Matrix of 58 Regions Polluted with Environmental Chemicals in Different Matrices (Median Values)[a]

region	HLPb	HLCd	HLZn	HLS	TLPb	TLCd	TLZn	TLS	MLPb	MLCd	MLZn	LRPb	LRCd
R06	1	0	0	1	0	0	0	0	0	0	0	0	0
R08	1	0	0	1	1	1	1	1	1	0	1	0	0
R07	1	0	0	0	1	0	0	1	1	0	0	1	0
R17	0	0	1	1	1	1	1	1	0	0	1	0	0
R09	0	1	1	0	0	1	1	0	1	0	0	1	0
R16	1	1	0	0	0	1	1	0	0	0	0	1	0
R22	1	0	0	1	1	0	1	1	1	0	0	0	0
R18	0	1	0	1	1	1	1	1	0	0	0	0	0
R30	0	0	0	0	1	1	0	0	0	0	0	1	0
R23	1	0	1	1	1	1	1	0	1	0	1	0	0
R15	0	0	0	0	0	0	0	0	0	0	0	0	0
R14	1	1	1	1	1	1	1	1	0	1	0	0	1
R05	1	0	0	1	1	0	0	1	0	1	0	0	1
R28	0	0	1	0	0	0	1	0	0	0	0	0	0
R39	1	0	1	1	1	0	1	1	1	1	0	0	1
R40	0	0	1	1	0	0	0	1	0	1	0	0	1
R29	0	1	0	0	0	1	1	0	0	1	0	0	1
R41	0	1	1	1	1	0	0	1	0	1	0	0	1
R42	0	0	1	0	0	0	0	0	0	1	1	0	0
R27	0	1	0	0	1	1	0	0	0	0	1	1	0
R38	1	1	1	0	0	1	1	0	1	0	0	0	0
R49	0	0	1	0	1	0	1	1	0	0	0	1	0
R37	0	1	1	0	0	0	0	0	0	0	0	1	1
R47	1	0	0	0	1	1	0	1	1	1	1	0	0
R48	1	1	1	0	1	1	0	1	1	1	1	1	1
R51	0	1	0	0	1	1	0	1	1	1	1	1	1
R04	0	0	0	1	0	0	1	1	0	0	0	1	1
R03	0	1	0	0	0	1	0	0	0	1	0	1	1
R13	0	1	1	0	0	1	1	0	1	0	0	1	1
R26	0	0	0	0	0	0	0	0	0	0	0	1	0
R36	1	0	0	0	1	0	0	0	0	0	0	0	0
R46	0	0	1	0	1	0	1	1	1	0	1	0	0
R50	1	1	0	0	1	1	1	0	1	0	1	1	1
R53	1	1	1	1	1	0	1	1	1	1	0	1	1
R45	1	1	1	1	0	1	1	1	1	1	0	1	1
R54	0	0	0	1	1	1	0	0	0	0	0	0	1
R59	1	1	0	0	1	1	1	0	1	1	1	0	0
R60	1	1	0	1	1	1	0	1	0	0	0	0	0
R58	1	0	0	1	1	0	1	1	1	1	1	1	1
R57	1	1	1	1	1	1	1	0	1	1	1	1	1
R35	0	1	1	1	0	0	0	0	1	1	1	1	0
R33	0	1	1	0	0	0	0	0	1	1	1	1	1
R25	0	0	1	0	1	0	0	0	1	1	1	1	1
R12	0	1	1	0	0	1	0	0	0	0	0	1	1
R21	0	1	1	0	0	0	1	0	0	1	1	1	1
R11	0	0	0	0	1	0	0	0	1	1	1	1	0
R02	0	1	0	1	1	1	0	0	0	0	1	1	1
R01	1	0	0	0	1	0	0	0	1	1	1	0	1
R10	1	0	0	1	1	1	1	1	0	1	1	1	0
R20	1	1	0	0	1	1	0	1	1	0	1	0	0
R24	1	1	1	1	0	0	1	1	0	0	1	0	1
R31	1	1	0	1	1	1	1	1	1	1	0	0	0
R32	1	0	1	1	1	1	1	0	1	1	1	1	1
R19	0	0	0	1	0	0	0	1	0	1	1	0	0
R43	0	0	1	1	1	0	0	1	0	0	1	0	0
R44	0	0	1	1	0	0	0	0	1	0	1	1	1
R52	1	1	1	1	1	1	1	1	1	1	1	1	1
R56	1	0	1	1	1	0	1	1	1	1	1	1	0

[a] Abbreviations: HL = herb layer, TL = tree layer, ML = moss layer, LR = *Lumbricus Rubellus* Pb = lead, Cd = cadmium, Zn = zinc, S = sulfur.

objects, in this case the regions of a state in Germany and their comparability relation to each other. Partially Ordered Scalogram Analysis reduces the data set into a two-dimensional scale. Hence by this method more clarity for the interpretation of the data is received at the cost of representing all comparabilities correctly. Namely some small part of information is lost by this method. Important priority objects, like maximal and minimal objects, can easily be detected by Hasse diagram technique and POSAC.

If data sets become too large, it seems that the Hasse diagram technique cannot be applied in the first step of the data-analysis. To reduce the complexity of the data matrix, additional multivariate methods such as a supervised cluster analysis should be performed initially.

Both multivariate methods have their pros but also their cons:

The Hasse diagram technique can help answering questions like the following: Which region is higher/lower polluted

Table 3: Reduced Data Matrix 58 × 2[a]

region	LOV (1)	LOV (2)	region	LOV (1)	LOV (2)
52	1.000	1.000	46	0.889	0.513
57	0.725	0.964	09	0.546	0.816
53	0.879	0.662	38	0.737	0.621
32	0.577	0.937	60	0.937	0.397
48	0.749	0.879	18	0.991	0.132
45	0.713	0.848	05	0.858	0.530
56	0.869	0.649	01	0.635	0.749
14	0.927	0.577	44	0.229	0.991
58	0.827	0.675	29	0.784	0.562
31	0.946	0.187	03	0.439	0.838
39	0.898	0.350	40	0.838	0.478
50	0.459	0.955	12	0.265	0.918
10	0.772	0.761	16	0.187	0.973
51	0.562	0.898	07	0.675	0.688
08	0.955	0.265	11	0.592	0.737
59	0.761	0.635	43	0.973	0.229
24	0.982	0.324	27	0.324	0.927
23	0.688	0.784	04	0.621	0.772
20	0.816	0.607	49	0.701	0.713
33	0.419	0.908	54	0.607	0.701
13	0.496	0.889	37	0.375	0.806
25	0.530	0.827	19	0.848	0.439
17	0.964	0.419	30	0.350	0.858
35	0.296	0.982	42	0.397	0.795
21	0.478	0.946	28	0.649	0.546
41	0.918	0.296	36	0.662	0.459
02	0.513	0.869	06	0.795	0.375
47	0.806	0.592	26	0.132	0.725
22	0.908	0.496	15	0.000	0.000

[a] Abbreviation: DIM = dimension, LOV = latent order variable.

than another region? With the aid of this technique comparisons of objects (regions) are possible. The ranking of objects belongs to the most important aims of this technique. Hence it can be used for evaluations in environmental sciences. One problem is that the Hasse diagram visualizes the partially ordered set. As a consequence the diagram might be very complex, because even small numerical differences are interpreted as comparability and incomparability, respectively. In general the order-relation is not a robust one; however, see ref 2. By reduction of the data matrix for example by POSAC the detailed ranking information was lost; however, structural information is won.

Multivariate explorative statistical methods offer simple and effective tools for graphical analyses of data matrices. In general they give *no* answer to the question if region i is higher polluted than region j. But the visualization of a data matrix by parallel coordinate plots[25] which in turn are just another expression of partial order[26] and Bertin matrices[27−29] may well be the basis for successful evaluation. To find and interpret patches in the Bertin matrix is an important task in data analysis. Also, parallel coordinate plots can be used to demonstrate maximal or minimal objects. These methods were applied on a reduced data set of the polluted regions in Baden-Württemberg.[7,8]

Hence a combination of Hasse diagram techniques and explorative statistical methods is a very promising approach to future tasks in chemometrics and environmetrics and will be followed as a main research topic by the authors. Special focus should be put on generalization to more than two dimensions and to the quantification of the importance of each additional dimension (like the procedure in PCA). Additionally it seems to us a very promising way to introduce the concept of probability in order theory.[30]

REFERENCES AND NOTES

(1) Brüggemann, R.; Friedrich, J.; Kaune, A.; Komossa, D.; Pudenz, S.; Voigt, K. *Vergleichende ökologische Bewertung von Regionen in Baden-Württemberg*; GSF-Bericht 20/98. GSF − Forschungszentrum für Umwelt und Gesundheit: Neuherberg, Germany, 1998; pp 1−148.

(2) Brüggemann, R.; Bartel, H. G. A Theoretical Concept to Rank Environmentally Significant Chemicals. *J. Chem. Inf. Comput. Sci.* **1999a**, *39*, 2, 211−217.

(3) Brüggemann, R.; Pudenz, S.; Voigt, K.; Kaune, A.; Kreimes, K. An Algrebraic/Graphical Tool to Compare Ecosystems with Respect to Their Pollution, IV: Comparative Regional Analysis by Boolean Arithmetics. *Chemosphere* **1999b**, *38*, 19, 2263−2279.

(4) Brüggemann, R.; Bücherl, C.; Pudenz, S.; Steinberg, C. Application of the Concept of Partial Order on Comparative Evaluation of Environmental Chemicals. *Acta Hydrochim. Hydrobiol.* **1999c**, *27*, 3, 170−178.

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

(6) SPSS Science. Systat 10, 2001, http://www.spssscience.com/SYSTAT/index.html.

(7) Welzl, G.; Voigt, K.; Rediske, G. 1998. In *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences held on November 16th, 1998 in Berlin, Berichte des IGB 1998, Heft 6, Sonderheft I*, Institute of Freshwater Ecology and Inland Fisheries, Ed.; Institut für Gewässerökologie und Binnenfischerei: Berlin, 1998; pp 101−110.

(8) Welzl, G. In *Chemische Stressfaktoren in aquatischen Systemen, Symposium am 13. und 14. April 2000 in Berlin, Schriftenreihe Wasserforschung*; Weigert, B., Steinberg, C., Brüggemann, R., Eds.; Wasserforschung e.V.: Berlin, 2000; pp 121−132.

(9) Welzl, G.; Faus-Kessler, T.; Scherb, H.; Voigt, K. In *Encyclopedia of Life Support Systems (EOLSS)*; Sydow, A. Ed.; EOLSS Publishers Co.: Oxford, 2003; in print.

(10) Einax, J. W.; Zwanziger, H. W.; Geiss, S. *Chemometrics in Environmental Analysis*; VCH: Weinheim, 1997; pp 1−16.

(11) Lorenz, R. J. *Grundbegriffe der Biometrie*; Gustav Fischer Verlag: Stuttgart, 1996; pp 1−238.

(12) Halfon, E.; Reggiani, M. G. On Ranking Chemicals for Environmental Hazard. *Environ. Sci. Technol.* **1986**, *20*, 11, 1173−1179.

(13) Halfon, E.; Brüggemann, R. In *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences held on November 16th, 1998 in Berlin, Berichte des IGB 1998, Heft 6, Sonderheft I*; Institute of Freshwater Ecology and Inland Fisheries, Ed.; Institut für Gewässerökologie und Binnenfischerei: Berlin, 1998; pp 11−48.

(14) Brüggemann, R.; Halfon, E. In *Order Theoretical Tools in Environmental Sciences, Proceedings of the Second Workshop October 21st, 1999 held in Roskilde, Denmark, NERI − Technical Report No. 318*; Sørensen, P. B., Carlsen, L., Mogensen, B. B., Brüggemann, R., Luther, B., Pudenz, S., Simon, U., Halfon, E., Voigt, K., Welzl, G., Rediske, G., Eds; National Environmental Research Institute: Roskilde, Denmark, 2000; pp 7−44.

(15) Lerche, D.; Brüggemann, R.; Sørensen, P. B.; Carlsen, L.; Nielsen, O. J. A Comparison of Partial Order Technique with three Methods of Multicriteria Analysis for Ranking of Chemical Substances. *J. Chem. Inf. Comput Sci.* **2002**, *42*, 1086−1098.

(16) Klein, D. J.; Brickmann, J. Partial Orderings in Chemistry. *Commun. Math. Comput. Chem.* **2000**, *42*.

(17) Klein, D. J.; Babic, D. Partial orderings in Chemistry. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 656−671.

(18) Randic, M. In Search of Structural Invariants. *J. Math. Chem.* **1992**, *9*, 97−146.

(19) Stoyan, D.; Stoyan, H.; Jansen, U. *Umweltstatistik, Statistische Verarbeitung und Analyse von Umweltdaten*; B. G. Teubner Verlagsgesellschaft: Stuttgart, Germany, 1997; p 53.

(20) Monjardet, B. Concordance between Two Linear Orders: The Spearman and Kendall Coefficients Revisited. *J. Classif.* **1997**, *14*, 269−295.

(21) Hettmannsperger, T. P. *Statistical Inference based on Ranks*; John Wiley and Sons: New York, 1984.

(22) Borg, I.; Shye, S. *Facet Theory, Form and Content*; Sage Publications: Thousand Oaks, U.S.A., 1995; pp 111−112.

(23) Voigt, K.; Welzl, G.; Rediske, G. In *Order Theoretical Tools in Environmental Sciences, Proceedings of the Second Workshop October 21st, 1999 held in Roskilde, Denmark, NERI − Technical Report No. 318*; Sørensen, P. B., Carlsen, L., Mogensen, B. B., Brüggemann, R., Luther, B., Pudenz, S., Simon, U., Halfon, E., Voigt, K., Welzl, G., Rediske, G., Eds.; National Environmental Research Institute: Roskilde, Denmark, 2000; pp 135−144.

(24) Luther, B.; Brüggemann, R.; Pudenz, S. An Approach to Combine Cluster Analysis with Order Theoretical Tools in Problems of Environmental Pollution. *Match, Commun. Math. Comput. Chem.* **2000**, *42*, 119−143.
(25) Wegmann, E. J. Hyperdimensional data analysis using parallel coordinates. *J. Am. Stat. Assoc.* **1990**, *85*, 664−675.
(26) Urrutia, J. In *Algorithms and Order*; Rival I., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989; pp 387−434.
(27) Falguerolles de, A.; Friedrich, F.; Sawitzki, G. *A Tribute to J. Bertin's Graphical Data Analysis*; 1996, http://www.statlab.uni-heidelberg.de/projects/bertin.
(28) Lawler, E. L. *The traveling salesman problem*; Wiley: Chichester, Great Britain, 1985.
(29) Sawitzki, G. Extensible Statistical Software: On a Voyage to Oberon. *J. Comput. Graphical Stat.* **1996**, *5*, 3, 263−283.
(30) Sørensen, P. B.; Mogensen, B. B.; Gyldenkaerne, S.; Rasmussen, A. G. Pesticide Leaching Assessment Method for Ranking both Single Substances and Scenarios of Multiple Substance Use. *Chemosphere* **1998**, *36*, 2251−2276.

CI0300083