



Ministry of Environment and Energy
National Environmental Research Institute

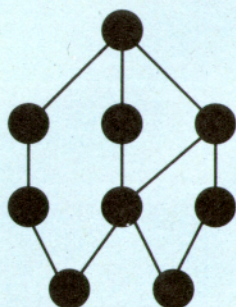
Order Theoretical Tools in Environmental Sciences

Proceedings of the Second Workshop October 21st, 1999 in
Roskilde, Denmark

NERI Technical report No. 318



?



!

[Blank page]



Ministry of Environment and Energy
National Environmental Research Institute

Order Theoretical Tools in Environmental Sciences

Proceedings of the Second Workshop October 21st, 1999 in
Roskilde, Denmark

NERI Technical report No. 318
May 2000

P.B. Sørensen¹, L. Carlsen¹, B.B. Mogensen¹, R. Brüggemann², B. Luther², S. Pudenz², U. Simon², E. Halfon³, T. Bittner⁴, K. Voigt⁵, G. Welzl⁵, F. Rediske⁵

¹)Department of Environmental Chemistry, ²)Inst. Of Freshwater Ecology and Inland Fisheries, Germany, ³)Mc Master University, Canada, ⁴)Volkswagen AG, Germany, ⁵)GSF-National Research Center for Environment and Health

Data sheet

Title: Order Theoretical Tools in Environmental Sciences
Subtitle: Proceedings of the second workshop October 21st, 1999 in Roskilde, Denmark

Authors: P.B. Sørensen¹, L. Carlsen¹, B.B. Mogensen¹, R. Brüggemann², B. Luther², S. Pudenz², U. Simon², E. Halfon³, T. Bittner⁴, K. Voigt⁵, G. Welzl⁵, F. Rediske⁵
Departments: ¹Department of Environmental Chemistry, ²Inst. Of Freshwater Ecology and Inland Fisheries, Germany, ³Mc Master University, Canada, ⁴Volkswagen AG, Germany, ⁵GSF-National Research Center for Environment and Health

Serial title and no.: NERI Technical Report No. 318

Publisher: Ministry of Environment and Energy
National Environmental Research Institute ©
URL: <http://www.dmu.dk>

Date of publication: May 2000

Referee: Lars Carlsen
Layout: Majbritt Pedersen-Ulrich

Please cite as: Sørensen, P.B. L. Carlsen, B.B. Mogensen, R. Brüggemann, B. Luther, S. Pudenz, U. Simon, E. Halfon, T. Bittner, K. Voigt, G. Welzl & F. Rediske (2000): Order Theoretical Tools in Environmental Sciences. Proceedings of the Second Workshop October 21st, 1999 in Roskilde, Denmark. National Environmental Research Institute, Roskilde, Denmark. - pp 172 -NERI Technical Report No. 318

Reproduction is permitted, provided the source is explicitly acknowledged.

Abstract: This is a proceeding form: The second workshop on Order Theoretical Tools in Environmental Sciences. The principle of partial ranking is presented and discussed as a tool for analyzing of environmental problems. The method is show to be effective for analysis of complex systems where incomparability exist among different key parameters.

Keywords: Mathematical modeling, Ecosystem analysis, Fate of exobiotics, Partial order ranking, Priority setting.

ISBN: 87-7772-548-4
ISSN (print): 0905-815X
ISSN (electronic) 1600-0048

Number of pages: 172
Circulation: 200

Price: DKK 150,- (incl. 25% VAT, excl. freight)

For sale at: National Environmental Research Institute
PO Box 358
Frederiksborgvej 399
DK-4000 Roskilde
Denmark
Tel.: +45 46 30 12 00
Fax: +45 46 30 11 14

Miljøbutikken
Information and Books
Læderstræde 1
DK-1201 Copenhagen K
Denmark
Tel.: +45 33 95 40 00
Fax: +45 33 92 76 90
butik@mem.dk
www.mem.dk/butik

Contents

R. Brüggemann, E. Halfon

Introduction to the General Principles of the Partial Order Ranking Theory 7

P.B. Sørensen, L. Carlsen, B.B. Mogensen

Uncertainty in partial order ranking 45

B. Luther, R. Brüggemann, S. Pudenz

Cluster analysis and Its Use in Hasse Diagram Technique 53

R. Brüggemann, E. Halfon, B. Luther, S. Pudenz, U. Simon

New Tools in Hasse Diagram Technique 73

S. Pudenz, T. Bittner, R. Brüggemann

Comparative Evaluation of Materials in Car Production 95

L. Carlsen, P.B. Sørensen, M. Thomsen

Estimation of Octanol-Water Distribution Coefficients 105

U. Simon, R. Brüggemann

Assessment of Water Management Strategies by Hasse Diagram Technique 117

K. Voigt, G. Welzl, F. Rediske

Environmentrational Approaches to Evaluate Internet Database 135

S. Dobel, B.B. Mogensen, P.B. Sørensen

Validation of a raking model in relation to Danish monitoring data 145

K. Voight, E. Halfon

Internet Forum for Partial Order Method -Hasse Web Site- 163

National Environmental Research Institute 171

Neri Technical Report 172

Preface

This is the proceeding from the second workshop on Order Theoretical Tools in Environmental Sciences, which was held in Roskilde at the National Environmental Research Institute (NERI), Denmark on October 21th, 1999. The first workshop was held in Berlin November 16th, 1998 and a proceeding from this workshop is available as: "Proceeding of the workshop on Order Theoretical Tools in Environmental Sciences", Berichte des IGB 1998 (Berlin), Heft 6, Sonderheft I, ISSN-Nr. 1432-508X. One way to get the proceeding from the first workshop can be by contacting e-mail: pbs@dmu.dk.

The first proceeding paper is an introduction to the principles of partial order, while the following papers analyses specific problems. The title of the papers shows clearly that a great variety of different problems can be involved in the analysis using partial order ranking. Actually the partial order ranking method as applied in the environmental science seems just to be in the initial state still having many undiscovered possibilities.

Introduction to the General Principles of the Partial Order Ranking Theory¹

Rainer Brüggemann^{a)} and Efraim Halfon^{b)}

*^{a)} Institute of Freshwater Ecology and Inland Fisheries
Müggelseedamm 310
12587 Berlin-Friedrichshagen
Germany*

*^{b)} Mc Master University
Hamilton
Canada*

Abstract

Data matrices can be analyzed by several tools. The technique of Hasse diagrams is one of these tools; it focuses on individual objects and their relation to each other. This publication explains the Hasse diagram technique (HDT) and is based on a few simple concepts of set theory, to perform a sensitivity and stability analysis: A matrix **W** is the basis to analyze the importance of criteria by which objects are characterized; in fact the ranking of a set of objects depends not only on the numerical values, but even more on the choice of criteria. Each criterion will be characterized by $\sigma(i)$ which quantifies its importance on ranking. Several examples are provided on the application and interpretation of the analysis. A brief outlook relates HDT with other methods: The cluster analysis is one of the main focus, some remarks can be found about deterministic models and their interplay with HDT. Finally Life Cycle Assessment and its relation to HDT and the use of other graphical representations is briefly explained.

¹ Parts of this text are modified and shortened versions of the GSF-report: "Theoretical base of the program 'HASSE' ". Brüggemann,R., Halfon, E. , GSF-report 20/95, 1995

1 Introduction

Sorting objects

Ranking

Often a first step in evaluation is to sort objects according to their properties, which are supposed to be important with respect to the evaluation. The assessment of priorities is a typical premise before final decisions are taken. However, often data do not allow to perform ranking unambiguously. This paper is thought of as an introduction into some elements of Discrete Mathematics to facilitate the understanding of ranking. The importance of criteria through which objects are ranked, is identified. Criteria that have not much influence on ranking may initialize further statistical investigations in order to be excluded from future measurements to save costs.

2 Partial order

2.1 Introductory remarks

Hasse diagrams show the relations of partially ordered sets (Posets). Here we explain why partial order is a useful concept on ranking.

Ordering

Ordering is a logical way to get an overview over objects: If for example pesticides are characterized by their detection frequency, DF (*P.Sørensen et al., 1998*) then these chemicals can be sorted according to increasing values of DF, their sequence corresponds to one characteristic number. Often however, only one number is not sufficient to characterize objects or -within this example- pesticides. For example not only DF, but also the observed concentration may be important for the decision to rank pesticides. Analogously other objects may be ranked, see for example databases (*Voigt and Brüggemann, 1995, Voigt, 1998*) or regions (*Brüggemann and Steinberg, 1999, Brüggemann et al., 1999a and further references therein*), etc. (for a rather recent overview, see *Brüggemann, 1998*). Common to these examples is that each object (pesticide, database, region, soil site etc.) can be characterized by more than one quantity. Objects which are characterized by several quantities (we call them “attributes“ -see later for details-) often cannot be ordered, because there are conflicts between their attributes.

An example may help to understand this:

We may have five objects $\{a,b,c,d,e\}$ characterized by two properties q_1 and q_2 . As often is the case both attributes do not behave parallel, i.e. a sorting is wanted, from which it is evident, which object is better than another with respect to all properties.

We can arrange the five objects according to q_1 and q_2

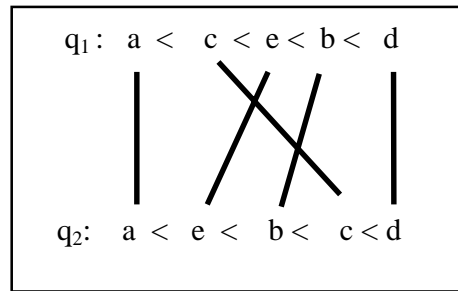


Figure 2-1: Two sequences of objects according to two different characteristics

Permutation diagram

This type of diagram is sometimes called a permutation diagram (Urrutia, 1987) or with a vertical orientation: parallel coordinates (Welzl et al., 1998). It shows that there are inversions between the two sequences. Some objects will mutually exchange their positions in dependence which quantity is used to define the sequence (for example c,e). Some other objects do not change their relationships to others, if the sequence defining quantity (here: q_1 or q_2) is changed (For example: a,d).

Partial order

Obviously some "residual order" remains, if both quantities are considered at once. This fact motivates the term "partial order". Within the given example of five objects partially ordering arises because more than one quantity is used to characterize the objects. This especially is often the case in ecology or ecotoxicology, where the complexity of nature prevents the use of a single ranking index and where utility functions (cost/benefit ratios, see for example Seip, 1994) are hardly available. Therefore the concept of partially ordered sets is useful in ecosystem theory. If the "usual" order, namely the order in which each object can be compared with each other, is to be pronounced, the term "linear" or "total" order is used.

Ecosystem theory

Permutation diagrams become confusing if many objects are included and especially if more than two quantities characterize the objects. In that case a corresponding number of sequences may arise and -to make evident the inversions- for each pair of sequences a permutation diagram must be drawn. Instead of this troublesome procedure which leads to $m \cdot (m-1)/2$ pairs of permutation diagrams (m attributes used) the technique of Hasse diagrams provides a useful tool for visualization. For example the partially ordered set of five objects is visualized in a Hasse diagram in Figure 2-2:

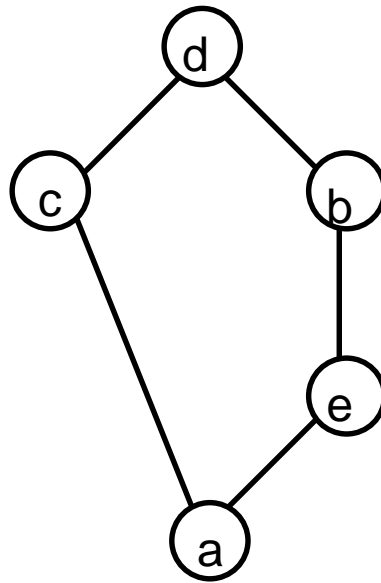


Figure 2-2: The Hasse diagram as an alternative to the permutation diagram shown in Figure 2-1

Later, we explain how to read a Hasse diagram and how to construct it. The fact that only a partial order is present, can be directly taken from the diagram (Figure 2-2). $a < c < d$ and $a < e < b < d$.

Figure 2-2 also shows that c and e , and c and b resp. change their position if either attribute q_1 or attribute q_2 is considered.

2.2 Definition of order

Firstly some simple set theoretical notations are introduced by examples.

- \subseteq : Inclusion. If $A \subseteq B$ then all elements of A are also elements of B . A may be identical with B . Example: $A = \{a,b,c\}$, $B = \{a,b,c,d,e\}$ then: $A \subseteq B$
- \cap : Intersection. If $A \cap B = C$ then the set C includes all elements which are common to A and B
Example: $A = \{a,b,c\}$, $B = \{b,d,e,f\}$
 $A \cap B = C = \{b\}$
The empty set is denoted by the symbols \emptyset or $\{\}$
If $A \cap B = \emptyset$ then A and B are disjoint sets
- \cup : Conjunction. If $A \cup B = C$ then the set C includes the elements of A and those of B
Example: $A = \{a,b,c\}$, $B = \{b,d,e,f\}$
 $A \cup B = C = \{a,b,c,d,e,f\}$
The union of disjoint sets often is denoted as " \oplus ". For example C also may be generated by $A' = \{a,c\}$ and $B = \{b,d,e,f\}$:
 $C = A' \oplus B$

\setminus : Difference of sets. If $A \setminus B = C$ then the set C includes the elements which are elements of A and not of B
 Example: $A = \{a,b,c\}$, $B = \{b,d,e,f\}$
 $A \setminus B = \{a,c\}$
 Another example: $A = \{a,b,c\}$, $B = \{a,b,c,d\}$
 $A \setminus B = \emptyset$

Family of sets: Several sets may be considered as elements of a set. For example: Let be $A=\{a,b,c\}$, $B=\{e,f\}$ and $C=\{a,f,g\}$, then a family F of sets, namely $F=\{A,B,C\}$ can be defined. F itself is a set, namely a set of sets.

Power sets: Given a set A , then the family of all its subsets is called a power set, denoted by $P(A)$. Example: $A=\{a,b\}$, then $P(A) = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}$. Often a family of sets F is defined as being a subset of a power set with A being the ground set: $F \subseteq P(A)$

Within a set E ($x,y,z \in E$) an order (partial order) on E is a relation with the following properties:

- | | | |
|---|--------------|---------|
| 1. $x \leq x$ | reflexivity | (2-1) |
| 2. $x \leq y$ and $y \leq x \Rightarrow y = x$ | antisymmetry | (2-1') |
| 3. $x \leq y$ and $y \leq z \Rightarrow x \leq z$ | transitivity | (2-1'') |

Note that $x \leq y$ or $y \leq x$ for all pairs $(x,y) \in E \times E$ is not demanded.

Reflexivity means: An object will be compared with itself.

Antisymmetry expresses the fact that for example $3 < 5$ but the reverse is not true.

Transitivity expresses the fact that for example $1 < 100$ and $100 < 1000$ then we conclude: $1 < 1000$.

A set E equipped with an order relation \leq is said to be an ordered set (or partially ordered set) or briefly "poset" and is denoted as (E, \leq) .

Poset

As already mentioned the notation "linear" or "total" order demands that all elements are comparable. Other binary relations which fulfill 1. - 3. also can be analyzed with the help of Hasse diagrams. The term "order" is not necessarily restricted to the usual comparison of numbers by their quantity. Therefore order theory is a powerful instrument which can be helpful for many scientific questions. Eigen and Winkler, in their famous book "Das Spiel - Naturgesetze steuern den Zufall", 1979, pointed out the importance of order in the nature, *Ruch and Lesche, 1978, Ruch, 1993* generalized the concept of entropy using partial orders. Ruch also applied partial orders with respect to enumeration problems of chiral molecules (*Ruch, 1972*) The increasing importance of partial order is also emphasized by *D.J.Klein, 1997*. The role of substructures to derive Quantitative Structure Relationships (QSAR) is discussed in several publications, see for a rather new one *Babic, 1996*. Substructure-structure relations fulfill the axioms of order (2-1). The determination of expansion coefficients with help of a training set is elegantly performed using the concept of Möbius functions (*Rota, 1961*).

Hasse diagram technique In environmental sciences the term "Hasse diagram technique" (HDT) becomes somewhat popular. By this term it is expressed that:

- a specific order relation, namely the product order (see below) is applied
- a diagram, which visualizes the results of sorting is of a main concern and
- methods to evaluate the ranking results are provided.

Note that a Hasse diagram can be drawn on different ways; the important and invariant information is given by the poset. The Hasse diagram is just a (very useful) visualization!

In order to establish the HDT some simple concepts are to be explained:

Attributes Attributes are quantitative, measurable data. We denote these attributes as q_1, q_2, \dots, q_m .

Object An object is the item of interest that may be characterized by attributes. Objects are ranked graphically by Hasse diagrams (see for example Figure 2-2). Generally the objects are considered to belong to a set " E ".

Element Therefore the objects are also often called "elements": An element is a member of a set.

Data Data are the numerical values corresponding to each criterion by which a given object is characterized.

Equivalent objects Equivalent objects in Hasse diagrams: Different objects that have the same data with respect to a given set of attributes. Equality with respect to a given set of attributes defines an equivalence relation, " \mathfrak{R} ". Instead of analyzing the objects $x \in E$, the set of equivalence classes may be the basis to construct a partial order. For example the set E may consist of five elements $\{a, b, c, d, e\}$, by the equivalence relation \mathfrak{R} the equivalence classes $[a, b, c]$ and $[d, e]^2$ may arise. Partial orders based on equivalence classes are often by far more comfortable. The set of equivalence sets is called quotient set and is denoted as E/\mathfrak{R} . The quotient set of the five element set, given above is then: $\{[a, b, c], [d, e]\}$. For further examples, see Brüggemann and Steinberg, 1999a and below.

Quotient set

Objects are considered to be elements of the object set E . Each object is characterized by attributes. We can create a table where the rows represent the objects and the columns the data of each object corresponding to the column-defining attribute. It is useful, but not necessary, to consider the table as being a data matrix \mathbf{Q} .

The $(n \times m)$ data matrix \mathbf{Q} contains m attributes (columns) for n elements (rows). The entry q_{ij} of \mathbf{Q} is the numerical value of the j th attribute of the i th element. Alternatively, $q_i(x)$ refers to the i th attribute of object x .

² Generally we don't use [...] brackets. However within this introductory text the use of equivalence classes as new objects, namely of the quotient set should be clarified.

Tuples

Tuples are elements of the Cartesian product $R \times R \times R \times R \dots \times R$ (m - times), whereby R is the set of real numbers. Each data for a given object and a given attribute is considered to be an element of R . Therefore a tuple of a given object is a row of the data matrix \mathbf{Q} . We prefer the notation „tuple“ instead of „vector“ or "row of a matrix", because we do not use the mathematical structure inherent in the term „vector“ on the one hand, and do not use the matrix properties of \mathbf{Q} on the other hand. The term $q(x)$ refers to the tuple $(q_1(x), q_2(x), \dots, q_m(x))$ i.e. to all its m components according to its m attributes.

Assume that we want to investigate a set E of n elements (objects) ($\text{card } E = n$). Each element is described by a set of m attributes. This set is a collection of properties, by which the objects are characterized in order to evaluate them. A single object is characterized by a m -tuple and the i .th component is the numerical value of the i .th property.

The full set of attributes is denoted by A , any subset of attributes by B , with $B \subseteq A$.

Sometimes it is useful to specify the subset of attributes. In that case we also use a subscript, for example A_i .

Information basis

A ranking can be done using all n attributes or with an arbitrary subset B of some attributes. The basis of ranking is the information collected in set A , which we therefore also call an "information basis" of the comparative evaluation of objects. If we use all attributes, i.e., all elements of A , then there can be only one possible ranking. However, if any one attribute is dropped, regarding now $m-1$ attributes, we can perform m additional ranking analyses. In total, $2^m - 1$ cases can be compared if all possible non empty subsets of A are considered. Often only the influence of one attribute is interesting. This task can be performed by comparing the case with the whole attribute set with those cases, where only one attribute is dropped. Therefore only $m+1$ comparisons are necessary.

Influence of one attribute

An example, how to get quotient sets may be helpful:

Five arbitrary objects „a,b,c,d,e“ are given and they are characterized by three attributes q_1, q_2, q_3 :

Table 2-1: Example to demonstrate equivalence classes, quotient sets

<i>objects\attributes</i>	<i>q_1</i>	<i>q_2</i>	<i>q_3</i>
a	2	3	1
b	2	3	3
c	1	3	1
d	2	3	1
e	1	3	3

We define a set E to be $\{a,b,c,d,e\}$, its cardinality , $\text{card } E$, is 5. The set of attributes $A = \{q_1, q_2, q_3\}$

- 1) Define \mathfrak{R}_1 to be equality under all three attributes, then the following four equivalence classes arise: $ec_1 = [a,d]$, $ec_2 = [b]$, $ec_3 = [c]$, $ec_4 = [e]$, one nontrivial class, because it contains more than one element, the others are singletons, containing only one element (also called trivial classes).

The quotient set E/\mathfrak{R}_1 is: $\{[a,d], [b], [c], [e]\}$ or $E/\mathfrak{R}_1 = \{ec_1, ec_2, ec_3, ec_4\}$, $\text{card } E/\mathfrak{R}_1 = 4$

Partitioning

- 2) Define \mathfrak{R}_2 to be equality under the subset of attributes $A_3 := \{q_1, q_2\}$, then the partitioning of E by \mathfrak{R}_2 is given by the following equivalence classes (which also are subsets of E): $[a,b,d]$ and $[c,e]$. The quotient set E/\mathfrak{R}_2 now contains only two elements, namely the both equivalence classes $ec_1 := [a,b,d]$ and $ec_2 := [c,e]$, i.e.: $E/\mathfrak{R}_2 = \{ec_1, ec_2\}$ ³ and $\text{card } E/\mathfrak{R}_2 = 2$

Classification

The examples show that there is a considerable simplification, if equivalence relations can be used. Therefore often not the original data are used, but a classification of data is performed just to find equivalence relations. A specific application of the use of equivalence relations is the p-transform method (*Brüggemann et al, 1999a*), in which only important objects are explicitly considered.

2.3 Product order

There are many realizations of the order axioms (2-1), see Table 2-3. A specific one is the product order: Let IB be the actual information basis of evaluation. Then two objects $x, y \in E$ are comparable, exactly if for all $q_i \in IB$ either $q_i(x) \leq q_i(y)$ or $q_i(x) \geq q_i(y)$. If for example $q_i(x) \leq q_i(y)$ for all $q_i \in IB$ with at least one attribute q_{i*} with $q_{i*}(x) < q_{i*}(y)$, then we write $x \leq y$. The demand "for all" is very important and is called the generality principle.

Generality principle

In more technical terms:

$$\begin{aligned} x, y \in E: x \leq y &: \Leftrightarrow q(x) \leq q(y) \\ q(x) \leq q(y) &: \Leftrightarrow q_i(x) \leq q_i(y) \text{ for all } q_i \in IB \end{aligned} \quad (2-2)$$

If (2-2) does not hold, i.e. there are some q_i , for which $q_i(x) < q_i(y)$ and some others for which $q_i(x) > q_i(y)$, then x and y are incomparable, and we write $x \parallel y$. If only one attribute is used, or all attributes are mutually perfectly correlated, then a total order arises, because all objects are mutually comparable. If there are no correlations, then from $q_i(x) \leq q_i(y)$

it cannot necessarily be concluded that this relation will be true for all attributes.

³ Here the equivalence sets are considered as elements of another set, namely the quotient set. Therefore here no italic letters are used.

An example may be useful: Consider four objects a,b,c,d. They are characterized by three attributes q_1 , q_2 and q_3 as the following table shows:

Table 2-2: Example of four objects with three attributes

<i>Object\attributes</i>	<i>q_1</i>	<i>q_2</i>	<i>q_3</i>
a	1	4	2
b	1.5	5	3
c	2	4	4
d	3	4	1

Obviously $a \leq b$. Actually: $a < b$. The same is true for the pair c,b: With respect to the first attribute: $a < c$, with respect to the second attribute: $a = c$. Therefore $a \leq c$. However the relation \leq does not hold for the objects b and c, because with respect to the first attribute $b < c$, however with respect to the second: $b > c$. Therefore objects which cannot be compared with each other, like b,c are called incomparable. Their attributes are contradictory, thus: $b \parallel c$. The object d cannot be compared with a,b and c. For example: $q_1(a) \leq q_1(d)$, however $q_3(a) > q_3(d)$.

The partial order is, as can be seen by the example and by equations (2-2), mainly determined by the actual information base. Therefore the product order, being the only one which is used in HDT, can also be denoted as (E, IB) . By this notation it also becomes evident that by changing the information base different partial orders arise. A poset can also be written as a set of pairs of objects, according to the order relation: In the example above:

$$(E, IB) = \{(a,b), (c,b)\}^4, \text{ according to } a \leq b \text{ and } c \leq b.$$

The axioms of order are not only fulfilled by the product order. The Table 2-3 gives an overview. Furthermore it may be useful to position the Hasse diagram technique (HDT) (see Figure 2-3).

2.4 How to draw a Hasse diagram "by hand"

Cover-relation

First of all there is to be introduced the cover-relation:

If there is no element "x" of E, for which $a \leq x \leq b$, $x \neq a, b$, $a \neq b$ holds, then a is covered by b, or b covers a.

The following notation will be useful, when the interpretation of Hasse diagrams is discussed:

The set of all elements covering the element a will be denoted $COV\uparrow(a)$, the set of all elements covered by the element a : $COV\downarrow(a)$. A corollary is: $COV\downarrow(a) \cap COV\uparrow(a) = \emptyset$ and:

$$\text{card}(COV\downarrow(a) \cup COV\uparrow(a)) = \text{grad}(a) \quad (2-3')$$

⁴ Note that sometimes the () parentheses are omitted.

Table 2-3: Some other order relations

Relation	Remarks/examples	References
Inclusion of sets	Formal concept analysis	Ganter and Wille, 1996 Bartel,H.-G., 1996
Interval order	Important in ecology (food web theory)	Lundgren, J.R.,1989
Lexicographic order	For ranking useful, when criteria can be strictly ordered for their importance	Davey, Priestley, 1990
Divisibility of natural numbers	For example transitivity: $12:6 \in IN^5, 6:2 \in IN \Rightarrow 12:2 \in IN$	
Containment	a) Shape analysis in QSAR, b) star-, sunrise-, diamond- or amoeba diagrams	a) Klein,D.J. 1997 QSAR and molecular shape analysis generally, see Mezey, P.G. 1993 b) Hartung,J., Elpelt,B.,1992
Implications	Formal concept analysis, artificial intelligence	Bartel,H.-G., John,P. 1999; Bartel,H.-G., Brüggemann, 1998 , Duquenne,V. 1987
Inclusion of subgraphs	QSAR	Klein, D.J., 1986 Essam,J.W., 1977 Schmalz et al., 1992

Valency

where the term $\text{grad}(a)$ denotes the valency of the vertex a , i.e. the number of (non-oriented) edges which are incident with element a .

With the help of the cover relation the construction of Hasse diagrams is performed as follows:

1. E may be represented by a configuration of circles.
2. Within each circle the object-name is given, or if there are more than one objects either the name of the equivalence class, or of a representative of the equivalence class or (if only two or three objects are elements of the equivalence class) the names of the objects themselves are written. Note that the program WHASSE only displays a representative within the circle; the other members of the equivalence class are shown in an extra field of the screen.
3. Each object (or each equivalence class under \Re) is drawn as a point in the two-dimensional plane.

⁵ IN the set of natural numbers 1,2,3,...

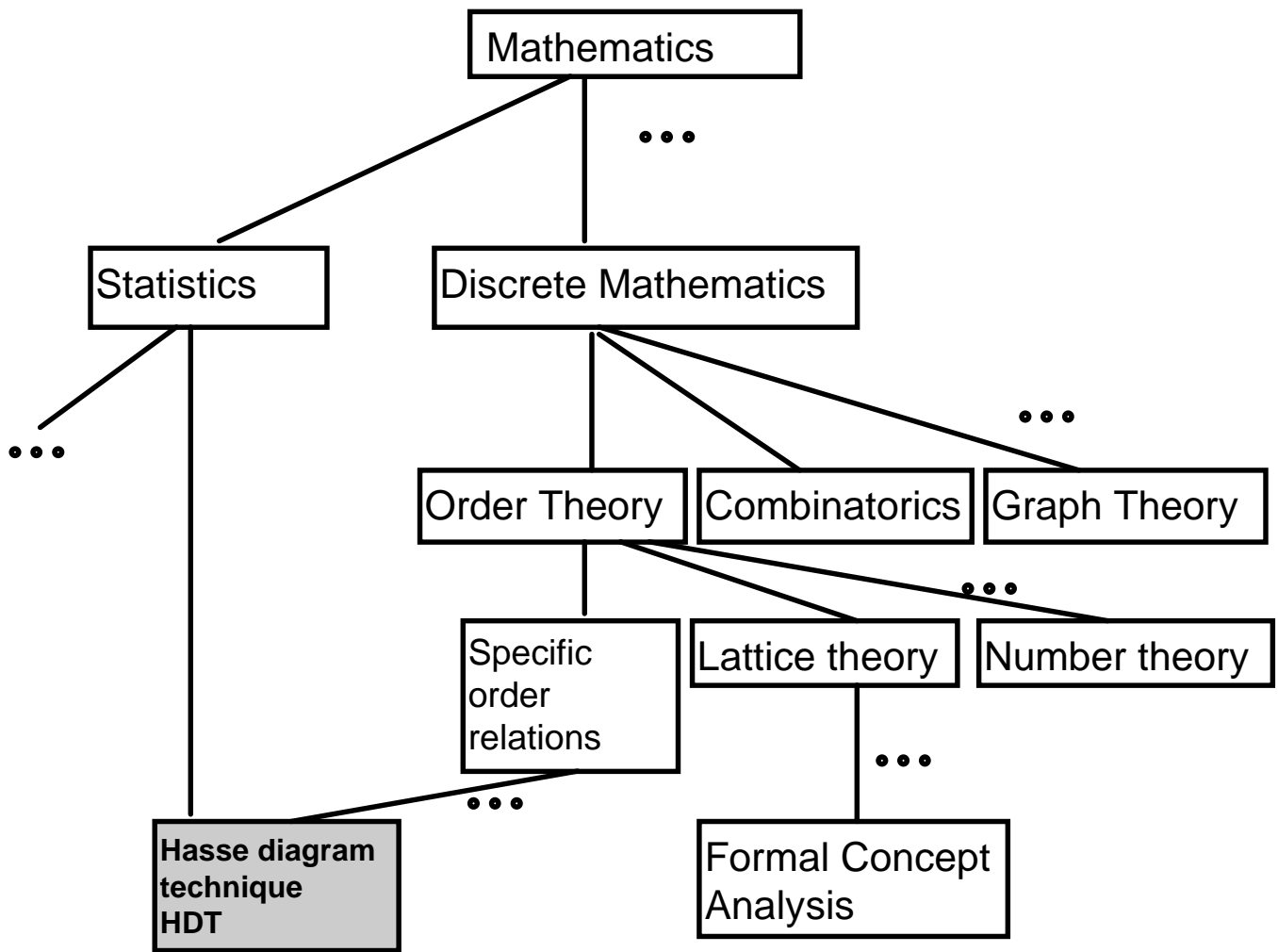


Figure 2-3: An attempt to position the HDT. Note that the same vertical position does not imply the same mathematical richness. For example the Formal Concept Analysis, developed by the school of Prof. Wille is a beautiful data exploration method, equipped with many mathematical theorems together with the charm of impressive diagrams (Ganter and Wille, 1996). An additional remark: There are many interrelations among the items shown, for example Graph theory plays an important role in Lattice theory, HDT etc. . A final remark: the classical reference for Lattice theory should be mentioned here: Birkhoff, G., 1984.

4. If a cover-relation holds, then a line between the corresponding object-pair (or pair of equivalence classes) is drawn. The covering pair is oriented corresponding to the \leq -relation.
5. The covered object in the \leq -related pair is located at a lower position on the page. (Alternatively we can, instead of the connecting line segment, draw an oriented arrow, beginning at the covering object and directed towards the covered object; in this case the locations in the two-dimensional plane of the Hasse diagram can be selected arbitrarily. In the practice it is more convenient to select the positions in the plane of the figure, according to the cover-relation.)
4. By this step the lines become an orientation.
4. Finally, not all line segments for which the \leq - relation holds are to be drawn. Because of the logical rule of transitivity (which holds by definition for partial orders) lines corresponding to the pair a,c with a

Orientation

Transitivity

$\leq b$ and $b \leq c$ concluding $a \leq c$ ⁶ are omitted. They do not present a cover-relation.

There are many ways to draw a Hasse diagram (some mathematicians are thinking about that point as art). For example the program WHASSE would draw the Hasse diagram of Figure 2-2 as shown in Figure 2-4. According to the scientific background the actual diagram may be constructed such that the results are presented as clear as possible. If there is no such specific background, the Hasse diagram is drawn as symmetric as possible. Incomparable objects are located at the same geometrical height and as high as possible on the page. For example the object c in Figure 2-4 could be located everywhere between objects d and a without hurting the order relations.

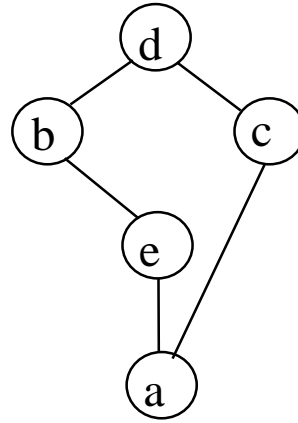


Figure 2-4: The Hasse diagram of the example of Figure 2-2 resp. drawn by the program WHASSE. Besides the vertical structure of levels the objects are located as symmetrical as possible.

Level

Because of the above mentioned convention, incomparable objects are arranged in levels. Often it is helpful to report the number of incomparable objects within an actually selected level. The concept of level can be introduced in more mathematical terms, see for example Halfon et al. 1998. Sometimes a compromise between the symmetry demand and the general clearness of the diagram is to be accepted, therefore program WHASSE offers a graphical editor to arrange the Hasse diagram manually. To describe the partial order uniquely it is recommended to use the notation of ordered pairs. The poset (E, IB) visualized by the Hasse diagram of Figure 2-4 would be written as follows:

$$(E, IB) = \{(a, e), (a, b), (a, d), (a, c), (b, d), (c, d), (e, b), (e, d)\}$$

⁶ Not to be confused with the objects of table 2-2. Here a, b, c are arbitrary objects to demonstrate the transitivity.

2.5 Hasse diagrams as mathematical objects

Digraphs

Hasse diagrams as shown for example in Figure 2-2 or 2-4 can be interpreted as mathematical graphs, i.e. they are called digraphs (*directed graph*), because of the orientation of the lines. Because of the definitions of order the digraphs are acyclic. Therefore Hasse diagrams are specific drawn examples of directed acyclic graphs (DAG). Interpreted as graphs, Hasse diagrams are triangle-free. The reason for that is the rule of transitivity, by which certain line segments can be omitted. A digraph consists of a set E (or E/\mathfrak{R} if the quotient set is to be partially ordered) of vertices (circles in Hasse diagrams) and a set of oriented edges each connecting two vertices. If the vertices are drawn in the diagram according to the rule 5 then the arrows can be simply be represented by lines. The circles are the objects of E , or elements of the set E/\mathfrak{R} to be ranked.

Maximal elements

Minimal elements

Least element

Greatest element

Hierarchies

When Hasse diagrams are used in ranking, the circles at the top of the diagram have no predecessors (they are not covered by any other object of E or E/\mathfrak{R}), and are called maximal elements or simply „maximals“. Often there are several objects which do not cover any further object. These special objects are called minimal elements or simply „minimals“. If there is only one minimal then it is also called a least element. If there is only one maximal, it is called a greatest element. Sometimes there is no path between parts of the non-directed graph, they are called (isolated) hierarchies or -if only one object (one equivalence class) constitutes a part of the Hasse diagram- an isolated element. If the partial ordering is given by (2-3), i.e. by product order, as it is presupposed in this paper, then often the objects at the top of the diagram have at least with respect to one criterion the maximal value. In ecotoxicology the criteria are often associated to the hazard. In that sense, the orientation of data is:

- small values: „good“, relatively unhazarduous
- large values: „bad“, relatively hazardous

Therefore the maximal elements are most hazardous. These and the isolated objects can be gathered to form the set of priority elements.

The new terms and some additional ones should be exemplified by Figure 2-5.

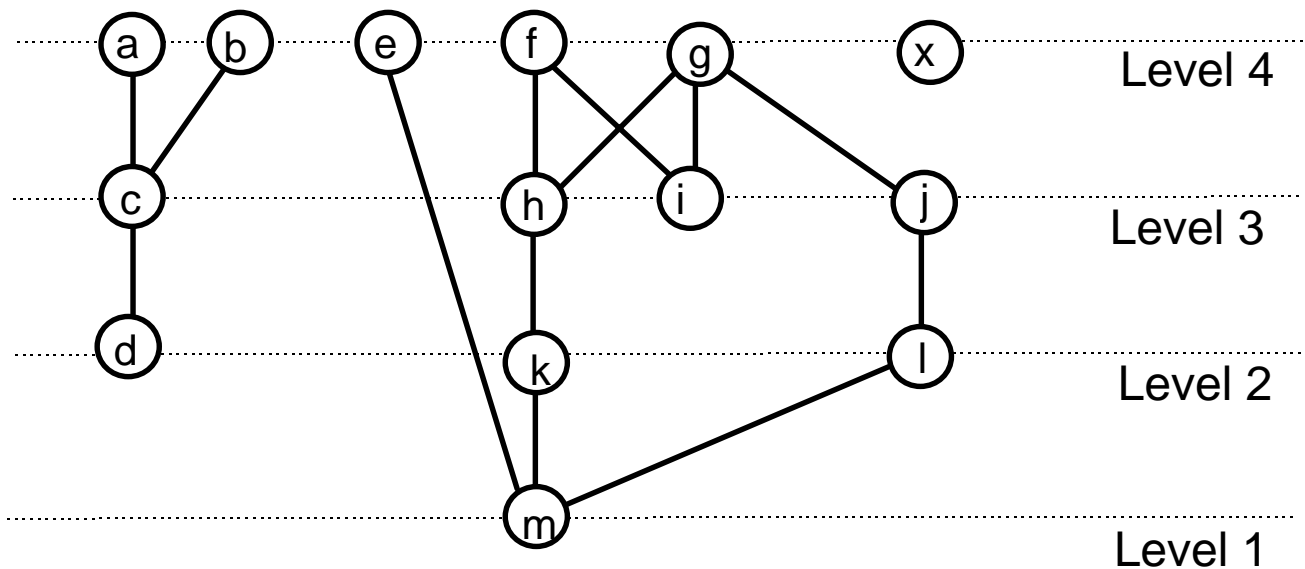


Figure 2-5: A typical Hasse diagram (see text for further explanations)

- The subsets of E , namely $\{a,b,c,d\}$, $\{e,f,g,h,i,j,k,l,m\}$, $\{x\}$ form - together with the order relations- three hierarchies. The poset is decomposable into three sub-posets, corresponding to the three subsets. Often it is a good policy to decompose the Hasse diagram, at least approximately.
- Levels: a first screening and a partitioning of set E according to increasing values of the attributes. They are defined by the longest chain within the Hasse diagram (see below).
- Level: Not unique from the point of view of order theory, but uniquely defined, if additional rules (for example: conservativity) are introduced!

Isolated element

- $(\{x\}, IB)$ is a trivial hierarchy, also called an isolated element of the poset (E, IB)
- The objects a,b,e,f,g are (proper) maximal elements (they have no upper neighbors) of the poset (E, IB)
- The objects d,m are proper minimal elements (no lower neighbors) of the poset (E, IB)
- The isolated object x can be seen as maximal and minimal element at once.

Chain

- A chain is a set of mutually comparable objects. For example $\{m,k,h,f\}$ is a chain, even the longest chain. Therefore there are 4 levels. In HDT objects are assigned to the highest possible level.

Antichain

- An antichain is a set of mutually incomparable objects. For example: $\{a,b,e,f,g,x\}$ is an antichain. $\{h,i\}$ an antichain too, but not a maximal antichain, because other objects can be added, maintaining the demand of mutual incomparability.

Articulation point

- The object m is an articulation point, because without m four hierarchies will be generated. Generally: An articulation point is a vertex of a graph whose elimination would increase the number of hierarchies.

3 Interpreting a Hasse diagram

The set E is given as follows, $E = \{a, b, c, d, e, f, g, h, i, j\}$ and we assume that there are four attributes which characterize the ten objects of the set E , namely $A = \{q_1, q_2, q_3, q_4\}$, see Table 3-1. Obviously there are no equivalence classes containing more than one element, if equality under all four attributes is demanded. The Hasse diagram will therefore contain circles in which only one object is located. The program WHASSE will draw a Hasse diagram, which can be modified by the user with the help of graphical edition facility⁷. Elements, such as "f", could be drawn at each position of the plane, however the only condition is, that the line between h and f has to have a slope which indicates that $q(f) \geq q(h)$. Similarly the element j could be positioned in different heights, again the lines connecting j with b, c and h must have such a slope that the covering relation is deduced without any doubts. There are four levels, according to longest chain. (We count the levels from the bottom to the top of the diagram). For example the third level contains (with the drawing conventions of the program WHASSE) the elements d, i, j.

Table 3-1: Example for interpreting a Hasse diagram

<i>objects\attributes</i>	<i>q₁</i>	<i>q₂</i>	<i>q₃</i>	<i>q₄</i>
a	0	5	3	2
b	5	0	4	4
c	5	0	0	5
d	0	3	1	1
e	0	0	1	1
f	0	5	5	0
g	0	2	4	1
h	0	0	0	0
i	4	0	2	1
j	2	0	0	3

⁷ Modifying the Hasse diagram by the user can only be done under the preservations of the order relations.

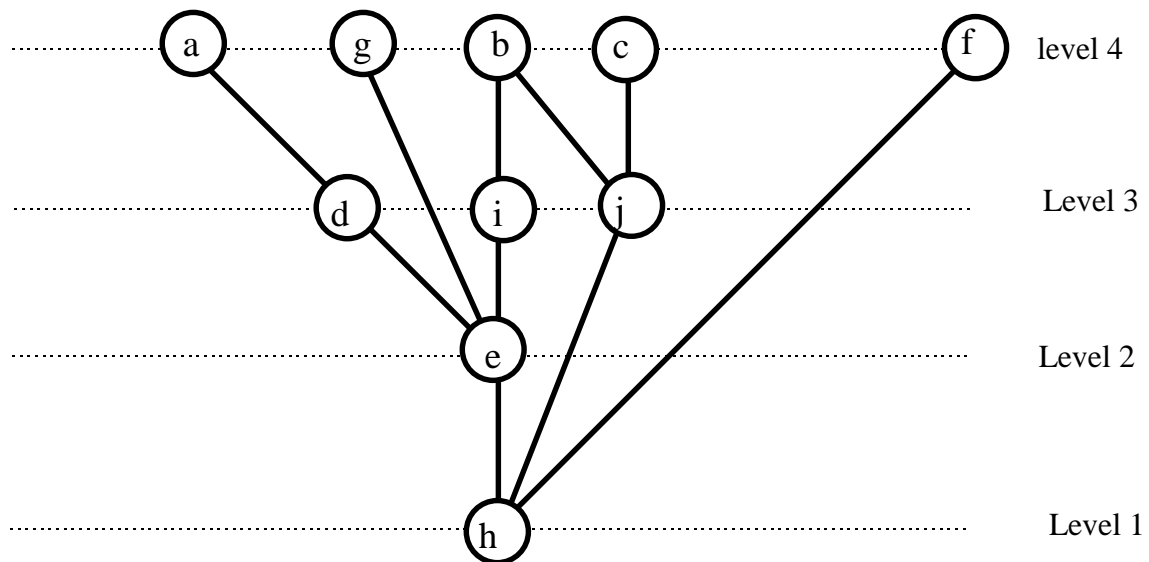


Figure 3-1: Hasse diagram of the objects given in Table 3-1

First of all it is striking that five maximal elements, namely a, g, b, f, c occur. Only one element is a minimal element. The object h therefore is the least element of this Hasse diagram.

There are many incomparabilities: Clearly the maximals are mutually incomparable. Beyond this for example: $d \parallel g$. The reason for all those incomparabilities is given by contradictory attributes. Or - to be more precise-: Given an incomparable pair of objects, then there must be at least one pair of attributes, for which their values are countercurrent. Such pairs of attributes we call antagonistic. Table 3-2 shows two incomparable objects:

Antagonistic attributes

Table 3-2: Example of incomparable objects of the set E

<i>object\attributes</i>	q_1	q_2	q_3	q_4
g	0	2	4	1
d	0	3	1	1

The following \leq -relations hold:

$$q_1(d) = q_1(g), q_4(d) = q_4(g)$$

$$q_2(d) < q_2(g) \text{ but } q_3(d) > q_3(g)$$

Therefore, for the $IB = \{q_1, q_2, q_3, q_4\}$ $d \parallel g$.

If the actual information base would consist of q_1 and q_2 , then $d > g$. Therefore, once again, the dependence of the partial order on the set of attributes is demonstrated.

An ordering index „I(i)“ can be constructed by summing up the components of each tuple (here because of four components: quadruple) $I(i) = \sum q_{ij}$ (sum to be taken over j) would not differentiate between the objects a, f and c ($I(a)=I(f)=I(c)=10$). Even if the sum is lower then 10 as it is the case for the object g, it is a maximal element showing its individuality given by the special "constellation" of attributes. We call a constellation of attribute related to a given object the pattern of attributes (of the object). The pattern of the object g which may cause special

managing activities would be hidden, if such ordering indices are introduced.

Diversity of pattern

On the one side, we can identify incomparabilities and can deduce some consequences (which clearly depend on the scientific background and cannot given in that arbitrary example); generally antichains (i.e. subsets of mutually incomparable objects) express a diversity of pattern, i.e. there are different qualities, whereas chains express that the values of attributes synchronously increase. One might say: Within a chain there are no new qualities, whereas bifurcations of chains, like {e,d,a} and {e,i,b} emerging from object e in Figure 3-1 lead to different incomparable objects with new patterns: d cannot be compared with i or b, etc.

If we ask for example whether other objects have a similar but in all components decreased values, then we simply follow the lines from the one selected object to the bottom of the diagram. For example five of ten objects have a 0 in the second component of the quadruple: b (one of the maximal element) , i, e, h and j. Only three objects have a pattern in which the second and the third component are 0: The maximal c and j and the least element h.

Priority elements

If the attributes are defined in such a manner that large numbers indicate a hazard (caused by chemicals (see for example *Sørensen, 1998*), by polluted sites (see for example *Pudenz et al., 1999*) or preferred properties (databases, see *Voigt and Brüggemann, 1995, Voigt et al., 1996*) etc. then the maximal elements are those objects which have the highest priority. If there are more than one object then different pattern may cause different strategies. If for example the single species toxicity for the species X is large and that for Y low for a given chemical, and for another chemical there is a reverse pattern and if furthermore these both chemicals are maximals then clearly both chemicals have a high priority, because in ecology no species can be preferred.

Ordinal information

Similarly: If different sites are high polluted by different chemicals and there are two, three or more maximals then there are different strategies needed to remediate the sites according to their different pattern of chemical pollution. Because of this facility the concept of partially ordered sets and their visualization by Hasse diagrams becomes important in evaluation procedures. Hasse diagrams maintain the ordinal information, which is often won by very expensive projects and only by sophisticated techniques. We like to say that partial order set theory keeps the different information about the objects "parallel" during the evaluation process.

Referring to Figure 3-1 we have five elements which are to be interpreted as high priority elements. The object f has a somewhat singular position: It is connected with the rest of the graph just by the least element, namely h. There are contradictions in the attributes such that f is not comparable with all other object besides h. If object h would be removed, then f would be isolated. Therefore object h is an articulation point, because its removal would lead to more isolated hierarchies. What are the reasons

for the singularity in f . First of all we state, that without a representation like Hasse diagrams in evaluation exercises such questions would hardly arise. Secondly we know that there must be at least one antagonistic pair of attributes which "separates" the object f from $E' = \{a, b, c, d, e, g, i, j\}$.

Using the interval technique (*Brüggemann et al, 1999b*) it is easily found that $\{q_3, q_4\}$ are antagonistic with respect to f and E' : With respect to q_3 all $x \in E'$ have smaller values than f , with respect to q_4 all $x \in E'$ have larger values than f .

4 Characterizing a Hasse diagram as a whole

4.1 Characterizing Numbers

NECA: Number of equivalence classes with more than one object, the number of nontrivial equivalence classes.

Width of a Hasse diagram

$W(E)$: The width of a Hasse diagram. It is the maximum number of elements of E/\mathfrak{R} which can be found in an antichain

The length of a Hasse diagram

$L(E)$: The length of a Hasse diagram: The number of line segments in the chain with a maximum number of equivalence classes under \mathfrak{R} .

Height of a Hasse diagram

$H(E)$: The height of a Hasse diagram. $H(E) = L(E) + 1$
Some numbers are calculated by the program WHASSE (CALCULATE-facility, submenu "Hasse Info):

NL, the number of levels = $H(E)$.

NEL, the number of elements (of E/\mathfrak{R}) in the level, which contains the most elements of E/\mathfrak{R} ; note that this number is not necessarily the same as $W(E)$

NMAX: The number of maximal elements (called : number of maximal equivalent classes because this information is related to E/\mathfrak{R})

NMIN: The number of minimal elements (notation as for the maximals)

Z: Number of all equivalence classes, including singletons (that are sets with only one element). Note that Z and NECA differ. If NA is that number of elements of E ($\text{card } E = n$), which are contained in nontrivial equivalence classes (NECA) then the following equation holds:

$$\text{card } E = NA + Z - \text{NECA} \quad (4-1)$$

In the CALCULATE-facility the equivalence classes are explicitly listed.

Some other numbers are also interrelated, for example the relation

$$NL = L(E)+1 \quad (4-2)$$

Note how many elements are contained in a level depends on conventions, how to draw the Hasse diagram. The number of elements of a level should therefore be seen as an operational number within the program WHASSE.

Further quantities are discussed within the contribution of *Pudenz et al.* this issue.

4.2 Comparability and Incomparability

Within a newer application of Hasse diagrams on phospholipid fatty acids as biomarker of microbial populations, *Brüggemann 1995* it turned out that the number of comparabilities and incomparabilities are useful tools.

A difficulty arises from the elements of the equivalence class: Obviously each of those elements is comparable with each other of the same class. Therefore the comparabilities within a Hasse diagram are counted as follows:

Draw a directed comparability graph for all \leq -relations. Elements of equivalence classes are considered to be comparable with respect to \leq - and to \geq -relations. Within the notation of ordered pairs there is no difficulty: Just write each equivalence like $a \equiv b$ as (a,b) and (b,a) . Each edge of this comparability graph is counted, leading to the number $V(n,m)$, with m being the number of considered attributes) and $n = \text{card } E$. This quantity is cumbersome to enumerate, however it will be calculated by the program WHASSE.

Degree of degeneracy

Each incomparable pair of objects is counted in both directions, leading to $U(n,m)$. Finally $K(n,m)$ is the degree of degeneracy. It is calculated by: $K(n,m) = \sum n_i * (n_i - 1)$ n_i the number of elements of the i th equivalence class. (4-3)

The sum is to be performed over all equivalence classes. If singletons are included, these sets do not contribute to that sum, because of $n_i = 1$. Therefore only NECA equivalence classes are to be considered.

Then the following basic equation holds:

$$S(n) = U(n,m) + 2 * V(n,m) - K(n,m) \quad (4-4)$$

$$S(n) := \text{card } E * (\text{card } E - 1)$$

For some applications it is useful to consider normalized functions like:

$$P(n,m) = U(n,m)/S(n) \quad (4-5)$$

We discuss the equation (4-5) with a bundle of examples:

- (I) If all m objects belong to one equivalence class ("full degeneracy"), then each object has a directed line to each other. Therefore:

$$\begin{aligned} U(n,m) &= 0 \\ V(n,m) &= m*(m-1) \\ K(n,m) &= m*(m-1) \\ S(m) &= m*(m-1) \\ P(n,m) &= 0 \end{aligned}$$

- (II) If all n objects belong to a chain with the height n then:

$$\begin{aligned} U(n,m) &= 0 \\ V(n,m) &= n*(n-1)/2 \\ K(n,m) &= 0 \\ S(n) &= n*(n-1) \\ P(n,m) &= 0 \end{aligned}$$

- (III) If all n objects belong to an antichain, width= n , then:

$$\begin{aligned} U(n,m) &= n*(n-1) \\ V(n,m) &= 0 \\ K(n,m) &= 0 \\ S(n) &= n*(n-1) \\ P(n,m) &= 1 \end{aligned}$$

Standard example

- (IV) Besides chain, antichain and a fully degenerated poset another typical poset, called a "standard example" (Trotter, 1991), is to be discussed: It is called S_{2n} and has the following structure: ($n=3$)

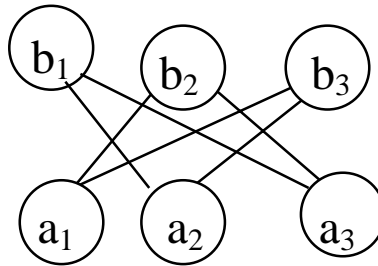


Figure 4-2: Graphical representation of the poset S_{2n} ($n=3$)

The construction principle of S_{2n} - posets is clear: Divide the set E into two disjoint sets AE and BE . The objects $a_i \in AE$, $b_i \in BE$:

$$b_k \parallel b_{k'}, a_i \parallel a_{i'} \text{ and } a_i \parallel b_j : \Leftrightarrow i = j$$

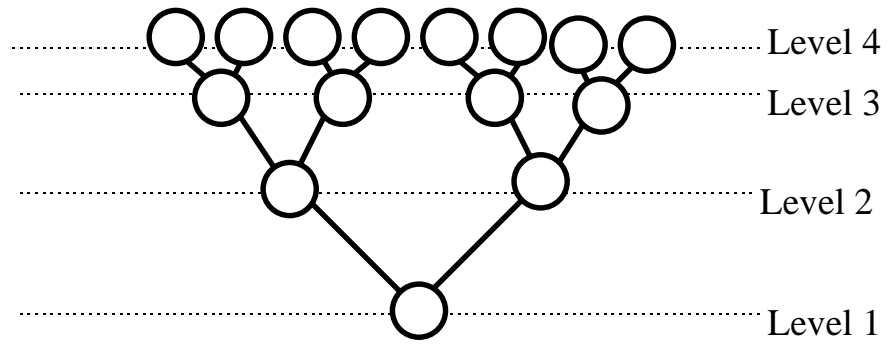
It is easily derived that:

$$\begin{aligned} V(2*n,m) &= n*(n-1) \\ U(2*n,m) &= 2*n^2 \\ K(2*n,m) &= 0 \\ P(2*n,m) &= n/(2*n-1) \end{aligned}$$

We write (in a self explaining way:)

$$P(S_{2n}) = n/(2*n - 1) \quad (4-6)$$

V) A binary (dichotomic) tree



Such trees are often good models for Hasse diagrams, therefore it is worth to consider them as an example too.

Let us select L as the Level number, i.e. $L = 1$, Level 1, $L=2$, Level 2 etc.

Then: $n = \text{card } E$ can be calculated as function of L :

$$n = 2^L - 1 = :n(L) \quad (4-7)$$

The number of maximal elements $NMAX = 2^{L-1}$

$$S(n) = n*(n-1) = (2^L-1)*(2^L-2) = :S(L)$$

The comparabilities may be calculated as function of L : $V(L) = \sum_{i=1, \dots, L-1} i*2^i$

or more easily recursively by:

$$V(L) = (L-1)*2^{(L-1)} + V(L-1) \text{ with starting value } V(1)=0 \quad (4-8)$$

We assume that $K(n,m)=0$, thus:

$$U(L) = S(L)-2*V(L) \text{ and } P(L) = 1 - 2*V(L)/S(L)$$

Some values are shown in the next table:

Table 4-1: Properties of Hasse diagrams analogously to binary (dichotomic) trees

L	n	number of maximal elements	$S(L)$	$V(L)$	$P(L)$
1	1	1	0	0	not defined
2	3	2	6	2	0.333
3	7	4	42	10	0.524
4	15	8	210	34	0.676
5	31	16	930	98	0.789
6	63	32	3906	258	0.868

For large tree-like Hasse diagrams the incomparabilities increase more than the comparabilities as functions of the number of levels.

Stability considerations

The actual attribute set may vary. Sometimes an additional property may be put into the information basis, sometimes an attribute will be omitted.

It is of interest to forecast the effect of a varying $m = \text{card } A$ on the ranking. Halfon, 1989 has explained that:

$$U(n,m) \leq U(n,m') \text{ if } m < m' \quad (4-9)$$

From the inequality $0 \leq P(n,m) \leq 1$ follows:

When $P(n,m)$ is near zero, then $U(n,m)$ must be near zero. Adding an attribute may have a big influence on ranking, and omitting an attribute may have a low influence.

Reversely: When $P(n,m)$ is near 1, then $U(n,m)$ must be near $S(n)$. Adding an attribute may have a low influence on ranking, and omitting an attribute may have a big influence.

When $P(n,m)$ is near 0.5, then both modifications of the information basis will have moderate effects. Table 4-2 summarizes these prognoses:

Table 4-2: Stability and Instability of the information basis

	<i>Adding an attribute</i>	<i>Omitting an attribute</i>
$P(n,m) \approx 0$	unstable	stable
$P(n,m) \approx 1$	stable	unstable

The quantity $P(n,m)$ does not differentiate between fully degeneracy and a chain, because in both cases no incomparability appears. A diagram shows this schematically:

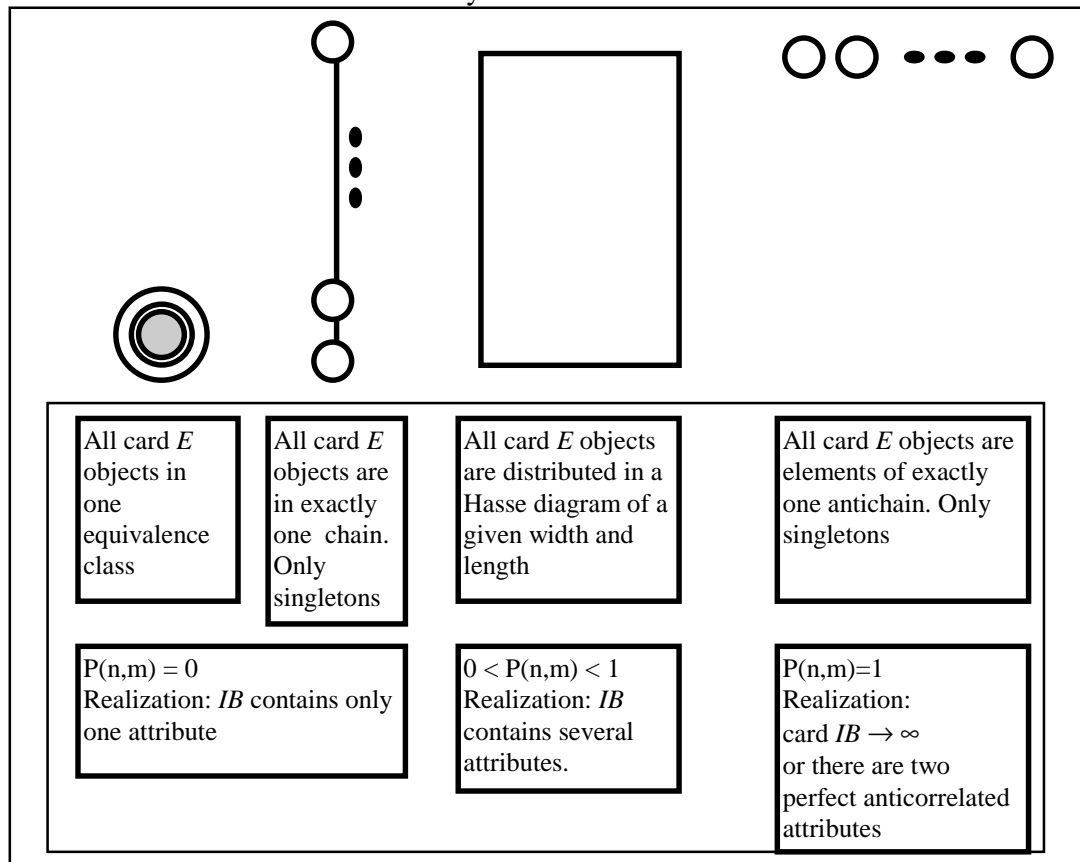


Figure 4-3: Schematical representation of the index $P(n,m)$ A discussion in more mathematical terms is given in Brüggemann, Bartel, 1999.

5 Sensitivity analysis

5.1 Mathematical Notation and background

5.1.1 Definition of key elements and successor sets

The rationale of this section is the geometrical analysis of a Hasse diagram to investigate substructures, i.e. relations among elements that might be hidden because of extensive number of circles and lines.

Starting point

Preferably a maximal element should be chosen as a starting point for the analysis. This choice, however, is not mandatory; other elements of E or E/\mathfrak{R} could be chosen too. This selected element is called "key element". We also can select simultaneously more than one key element even all elements (no restrictions apply here). For the sake of convenience all key elements are supposed to form a set $\mathbf{K} (\subseteq E)$.

Key element

Successors

The analysis of a key element implies a search of all elements located lower than that of the key element, i.e. all elements that can be reached from the key element by a path, a sequence of connecting edges. (Therefore the selection of maximal elements rather than other elements is more meaningful). These elements together with elements equivalent but not identical to the key element are called successors. The set of all successors of the key element "k" is denoted as $G(k)$. Note the similar concept of "down-sets" in *Davey and Priestley (1990)*: The order ideal (or down set) , generated by the key element will be denoted by $O(k)$, the principal order ideal.

Principal order ideal

Then it is valid:

$$G(k) = O(k) \setminus \{k\} \quad (5-1)$$

By definition $G(k)$ does not include the key element itself. The successor sets and their cardinalities are at the heart of the Hasse analysis shown here. The analysis of the structure of a Hasse diagram includes the analysis of different successor sets $G(k)$, arising from different subsets of A .

5.1.2 Warning

A crucial assumption for the whole analysis, which is based on the cardinality of successor sets, is the following: In our applications, the partial ordering of the set E is induced by the " \geq "-relation. Thus it has to be assured by appropriate data handling that any two elements ordered by " $>$ " can be considered as physical and numerical significantly different. <numerical significant differences of data> For example differences within the statistical noise or within numerical uncertainties or within experimental errors are considered as being physically meaningless, but the Hasse diagram technique considers such objects as different. If there are differences between objects, which are physically meaningless then the cardinality of the successor sets also would be meaningless. It is a

Numerical significant differences of data

good practice to use statistical methods , for example cluster analysis, to come up to well differentiated "pseudo-objects". Conventional cluster analysis leads to a partitioning of the set E into mutually disjoint subsets of similar objects. Formally an equivalence relation is defined, which operates on E and generates the partitioning into subsets. This generalized equivalence relation is simply stated as follows:

Let $a, b \in E$ then $a \sim b : \Leftrightarrow a, b$ belong to the same cluster. A discussion of cluster analysis will be performed in the contribution of B.Luther et al, this proceeding.

The concept of key elements and successor sets is important for the sensitivity analysis:

A set of objects can be ranked using a set of attributes. The full set, or a subset of attributes can be used. Since this choice is quite subjective we have developed a technique to analyze the effect of including or removing any attributes. As mentioned in the introduction, to assess the importance of an attribute we compare the results of ranking when different subsets of attributes are used. In practice this effort implies a comparison of Hasse diagrams. Results are stored in a matrix \mathbf{W} . This means: The matrix \mathbf{W} contains the mutual comparisons of the Hasse diagrams which arise, if all interesting cases are to be studied. In more general terms: If the power set $P(A)$ is analyzed, then a set of posets $\{(E, A_i)\}$ arises. The matrix \mathbf{W} is one of several possibilities to define distances between the elements of $\{(E, A_i)\}$.

5.2 Residual sets

5.2.1 Definition of successor set extended

The notation of successor set must be expanded to include all Hasse diagrams that are drawn when any combinations of attributes are used. Remember that the successor set depends not only on the key element but on the attributes used. For simplicity we will use the following notation:

$$G(k, A) \text{ or } G(k, B)$$

where G is the successor set, k denotes some arbitrary chosen key element, A is the full set of attributes and B, C, \dots are subsets of attributes, $B \subseteq A$, and $C \subseteq A$ resp.

5.2.2 Symmetric differences

To assess the influence of each attribute on ranking, we compare Hasse diagrams that arise from each subset B of A . A straightforward method to perform this task is to choose a key element and quantify the effect of each attribute set on its successor set.

For this purpose the residual set, R is now introduced.

$$R(k, B, C) := (G(k, B) \setminus G(k, C)) \quad (5-2)$$

In general $R(k,B,C) \neq R(k,C,B)$. Therefore the symmetric difference set " $W(k,B,C)$ " of the sets $G(k,B)$ and $G(k,C)$ is introduced:

$$W(k,B,C) := R(k,B,C) \cup R(k,C,B) = [G(k,B) \setminus G(k,C)] \cup [G(k,C) \setminus G(k,B)] \quad (5-3)$$

Symmetric difference

The set operation: $[G(k,B) \setminus G(k,C)] \cup [G(k,C) \setminus G(k,B)]$ is called a symmetric difference of $G(k,B)$, $G(k,C)$, and is denoted as " Δ ".

If the cardinality of $W(k,B,C)$ is small (compared with the min $[|G(k,B)|, |G(k,C)|]$) then subsets B and C lead to not very different Hasse diagrams. If the difference is large then the two corresponding Hasse diagrams are dissimilar to each other. Those attributes by which B and C differ, play a key role in ranking.

This finding motivates the introduction of the matrix \mathbf{W} .

5.3 Definition of the matrix \mathbf{W}

First of all it is to be stated that now only the set E is important. Because different cases would induce different equivalence relations the different quotient sets are no more comparable. If different subsets A_i of attributes will be considered then in general equivalence sets will "absorb" the objects, if on the one side attributes are dropped (which is consistent with equation (4-9)). On the other side there will be a splitting into smaller equivalence classes, if the number of considered attributes will increase.

The matrix $\mathbf{W}(k)$ assesses the difference of Hasse diagrams induced by the two subsets of attributes with respect to a key element k . **This matrix is at the heart of the analysis**, we call it the "dissimilarity-matrix", because the larger the matrix-entries are, the greater is the difference between the successor sets for the element k and hence between the Hasse diagrams.

We define the entry $W(k,B,C)$ to be:

$$W(k,B,C) := \text{card } W(k,B,C) \text{ or } W(k,B,C) = \text{card } [R(k,B,C) \cup R(k,C,B)] \quad (5-4)$$

For any key-element k the residual sets $R(k,B,C)$ and $R(k,C,B)$ are determined, their elements counted and summed.

The entries of the matrix \mathbf{W} are calculated by adding the cardinalities of the R -sets. From

$$\mathbf{W}(k,B,C) = \text{card } [(R(k,B,C) \cup R(k,C,B))] \text{ and } R(k,B,C) \cap R(k,C,B) = \emptyset$$

follows

$$W(k,B,C) = \text{card } R(k,B,C) + \text{card } R(k,C,B) \quad (5-5)$$

Dissimilarity-matrix

By definition the dissimilarity-matrix is symmetric and its diagonal elements are zero. From the definition of the residual sets we can derive:

$$W(k,B,C) = \text{card}\{ [G(k,B) \setminus G(k,C)] \cup [G(k,C) \setminus G(k,B)] \}$$

The equivalent form

$$G(k,B) \Delta G(k,C) = [G(k,B) \cup G(k,C)] \setminus [G(k,B) \cap G(k,C)] \quad (5-6)$$

is easier to evaluate.

Hamming-distance

The quantity $\text{card}[G(k,B) \Delta G(k,C)]$ is called the Hamming-distance (Bollobás, 1986) between the successor sets given by B, C .

To simplify notation, we now write $W(k,i,j)$ for $W(k,B,C)$. We note that

$$W(k,i,j) \geq 0 \text{ and } W(k,i,j) = W(k,j,i) \text{ for all } i,j \quad (5-7)$$

according to the defining equation. The matrix itself is denoted by $\mathbf{W}(\mathbf{k})$. Thus we have:

$$\begin{aligned} & \text{W(k)} = \frac{\begin{matrix} W(k,1,1), W(k,1,2), ..., W(k,1,p) \\ W(k,2,1), W(k,2,2), ..., W(k,2,p) \\ \\ \\ W(k,p,1), W(k,p,2), ..., W(k,p,p) \end{matrix}}{(5-8)} \\ & \text{with } p = 2^m - 1 \end{aligned}$$

5.4 Search for the important attributes

5.4.1 Theoretical considerations

Additivity property

The additivity property holds for the elements of $\mathbf{W}(k)$, $\mathbf{W}(k,i,j)$. Namely, given subsets of attributes $B \subset C \subset D$ or generally $A_r \subset A_s \subset A_t$ then it is valid

$$W(k,r,t) = W(k,r,s) + W(k,s,t). \quad (5-9)$$

Equation (5-9) can be generalized to:

$$A_1 \subset A_2 \dots \subset A_n \Rightarrow W(k,1,u) = W(k,1,2) + W(k,2,3) + \dots + W(k,u-1,u).$$

The following statements hold:

$$W(k,i,i) = 0 \text{ for any } i \text{ and } W(k,i,j) \geq 0 \text{ for } i \neq j$$

$$W(k,i,j) \leq W(k,i,s) + W(k,s,j) \text{ (the triangle inequality)}$$

Metric distance

Therefore $W(k,i,j)$ is a metric distance (Hamming-distance).

Furthermore, given three subsets for which the following is valid

$$A_i \cup A_j = A_r$$

then the equation (5-10) holds:

$$W(k,i,j) = W(k,r,i) + W(k,r,j). \quad (5-10)$$

Consistency

We call this equation "consistency".

The importance of Eq.s (5-10) and (5-11) is the fact that attribute sets can be related to successor sets (and thus to some features of Hasse diagrams) by

$$A_i \subseteq A_j \Rightarrow G(k,A_i) \subset G(k,A_j) \quad (5-11a)$$

and

$$A_i \cup A_j = A_r \Rightarrow G(k,A_r) = G(k,A_i) \cap G(k,A_j). \quad (5-11b)$$

The following two procedures provide insight on the importance of attributes:

- 1) The determination of $W(k,i,j)$ to see whether the ranking of the key element k , changes when we compare A_i with A_j (fixed k).
- 2) The comparison of several $W(k,i,j)$'s, $k \in K$ (K is any set of key elements) to see how a change in attributes affects a set of several key elements.

To perform this second analysis it is important to introduce:

$$W(K,i,j) = \sum W(k,i,j); k \in K \subset E \quad (5-12)$$

or in a shorter notation

$$W(K) = \sum W(k); k \in K$$

Again $W(K)$ is a symmetrical matrix.

$W(E)$ is the total dissimilarity matrix of the set of E . Let be $n := \text{card } E$. Of theoretical interest is the fact that there can exist n $W(k)$ -matrices for a total of $2^n - 1$ different $W(K)$ -matrices, with $K \in P(A)$, $K \neq \emptyset$. *In practice, however, mainly the $W(k)$ and the $W(E)$ matrices are useful.*

Furthermore, as mentioned earlier often not all possible cases are of interest. If the importance of attributes is to be studied, then it suffices to consider the case in which all attributes (case No. 1) are taken in to regard and consecutively all those cases, in which exactly one attribute is dropped. We arrive then to a matrix W which only contains the first row. The numbers $W(E,1,i)$ $i=2,\dots,m+1$ indicate the relative importance of each attribute. Note that the first case refers to the case where all attributes are taken into regard, then there remain m cases, where exactly one attribute is omitted. The use of the matrix $W(E,1,i)$ as sensitivity measure is summarized in the next section:

5.4.2 Steps towards a sensitivity

The original matrix W is important for several applications. Because of the symmetry not all p^2 entries of this matrix are to be analyzed.. $p = 2^m -$

1. It is sufficient to consider the upper or lower triangle part, i.e. only $p^*(p-1)/2$ entries.. Because the diagonal elements are zero, even the diagonal of this matrix can be disregarded. If a sensitivity analysis is to be performed, only m entries are of importance, as mentioned in the section above. Here the steps are restated, in order to help the user of the program WHASSE. The matrix \mathbf{W} is the key for the sensitivity analysis of ranking, each entry of \mathbf{W} is the cardinality of the symmetrized difference of two successor sets which are constructed from a given key element k and the two Hasse diagrams induced by two attribute subsets. Thus the rows and columns of this matrix are indicated by those two given subsets A_i and A_j :

- 1) Since we are interested only in comparisons of the full attribute set A with subsets A_i , only one row of the matrix \mathbf{W} is of interest. Since this is an example, we can choose the first one without loss of generalization, thus we are left with $W(k,1,1), W(k,1,2), \dots, W(k,1,p)$, where the index 1 denotes the full attribute set A and $p=2^m - 1$.
- 2) To see the influence of attributes on a Hasse diagram we compare the Hasse diagrams induced by A with those induced by the attribute sets with only $m-1$ attributes. Therefore the effect of dropping exactly one attribute is given by the remaining m entries of the first row, $W(k,1,2), \dots, W(k,1,m+1)$.
- 3) The remaining m matrix elements of the first row $W(k,A,A_1), \dots, W(k,A,A_m)$ are put together to form a "sensitivity tuple" of the key element k , $s(k) := [W(k,A,A_1), \dots, W(k,A,A_m)]$. Note that the enumerations of the subset A_i are as follows:

$$\begin{aligned} A_i &= \{q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_m\} & (i \text{ skipped}) \\ A_1 &= \{q_2, \dots, q_m\} \\ A_m &= \{q_1, \dots, q_{m-1}\} \end{aligned} \tag{5-13}$$

- 4) $s(k)$ can also be written as $[s_1, \dots, s_m]$. The larger s_i the larger is the symmetrized difference between $G(k,A)$ and $G(k,A_i)$ and correspondingly the larger the influence of attribute q_i on the position of key element k within the Hasse diagram under A compared with that under A_i .
- 5) The matrix $\mathbf{W}(k)$ depends on the selection of the key element k . If however, more objects are to be analyzed we generalize according to eq. (5-14):

$$W(K,i,j) = \sum W(k,i,j); k \in K \subseteq E \tag{5-14}$$

where K is any set of key elements, in a shorter notation $\mathbf{W}(K) = \sum \mathbf{W}(k)$.

- 6) All objects are selected as key elements. Therefore instead of $W(k)$, $\mathbf{W}(E)$ is to be investigated. $\mathbf{W}(E)$ is the total matrix of the set E . We note that a crude upper limit of $W(E,i,j)$, with $\text{card } E = n$, can be found simply by comparing a poset of solely non comparable elements (case "i") with a poset where all elements are equivalent to each other (case "j"). Together with Eq. (5-7):

$$0 \leq W(E,i,j) \leq n(n-1) \tag{5-15}$$

Sensitivity tuple

7) $W(E)$ will be used as a measure of sensitivity. Accordingly we suggest to quantify the sensitivity by:

$$\sigma(i) := W(E, A, A_i) \quad 1 \leq i \leq m \quad (5-16)$$

with the enumeration scheme of (5-13). According to (5-15), $\sigma(i)$ has values between 0 and $n(n-1)$.

6 Concluding remarks

6.1 Statistics and partial order

Statistics, especially cluster analysis, delivers important tools for preparing the data matrix Q . Therefore it is recommended, first to analyze the data set by statistical methods and then to perform a ranking by the concept of partial order. A detailed description, how to use fuzzy clustering methods can be found in Pudenz et al, 1998; Pudenz et al., 1999; one lecture (Luther et al.) within this workshop '99 deals with mathematical questions behind the application of cluster analysis in order theory. The problem is that each partitioning will lead to another poset and there seems to be no logical relations among the Hasse diagrams if different partitions are selected. For example will the order relations be preserved, if the partitioning is refined?

Fuzzy clustering methods

After ranking, one wants to interpret the results. Here again statistical methods are appropriate, for example the use of Bertin matrices, as Welzl et al. pointed out (Welzl et al. 1998, 1999). Voigt et al, this workshop '99 showed how statistical software can be applied and related it to approaches of order theory.

Bertin matrices

Often the dendrograms of a hierarchical cluster method are thought of as being Hasse diagrams. On the one side, this is true, because the clusters are related to each other by an inclusion, which fulfills the order axioms, on the other side this is wrong, because the clusters are not necessarily ordered with respect to their attributes, Figure 6-1 shows this schematically.

Dendrograms

6.2 Models and HDT

There are different steps to include deterministic mathematical models into the comparative evaluation on the basis of HDT:

- Model results may be used as attributes. Examples can be seen in the literature (Brüggemann, 1988, Brüggemann, 1994). Models may help to find attribute sets, which correspond to some protection goals. For example instead of ranking chemicals by a sorption coefficient and persistence, the resulting concentration of the sorbed chemical may be

the better descriptor to rank chemicals and may better correspond to a criterion, which aims to protect a sorbing medium.

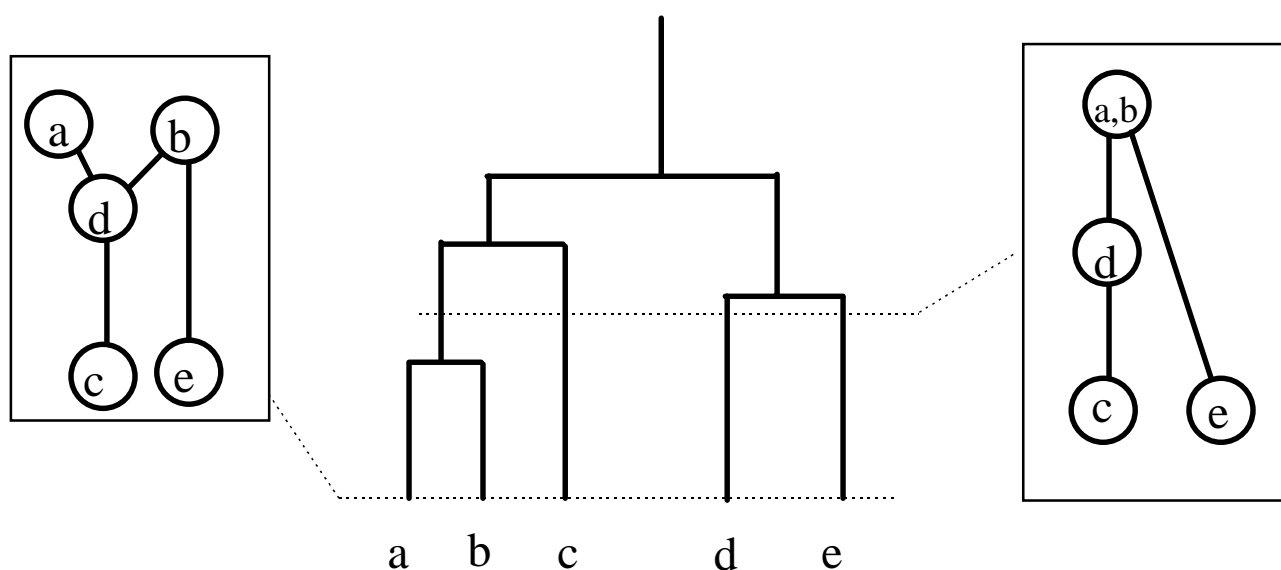


Figure 6-1: Schematical relation between dendrograms of hierarchical cluster analysis and Hasse diagrams.

- The input data of models may be used as attributes and the resulting Hasse diagram may be compared with that, which arose from field monitoring studies. A first attempt is made by *Galassi et al 1996* and *Halfon et al. 1996*. A full established comparison was performed by *Sørensen et al., 1999* (*submitted at Environmental Science & Technology*).
- Even if a deterministic model delivered the one fitness or goal function, which leads to a unique decision, the contributions to this goal function may be of interest for a ranking exercise, to understand the final result.

6.3 Other evaluation techniques

There is no problem to evaluate objects, if the protection goal is known and a corresponding goal function is at hand. However, often there are more protection goals at once, and even worse, there is no deterministic goal function available. Typically the decision makers look for an operationally defined function, which aggregates the important properties. By the aggregation the individual preferences are important. If -for example- toxicity is seen as more important than mobility of a chemical, than within a linear function the toxicity data will get a higher weight. The arbitrariness can be avoided by HDT; (see also the contribution of *Brüggemann et al, this proceeding*), however at the cost, that incomparabilities arise. To come up to an unique decision further steps are necessary, an example will be presented by *Simon et al, this proceeding*: First attempts to apply elements of the so-called tournament theory will be applied on the evaluation of watershed management strategies.

Life Cycle Assessment

Inventory analysis

If products are to be evaluated, then the concept of Life Cycle Assessment (LCA) is often applied. *Klöpffer and Volkwein 1995, Volkwein et al. 1996* showed, how to overcome the difficulty of the different dimension which are to be handled, when LCA is applied. He and his coauthors used HDT for a graphical display of LCA results. In terms of HDT, the LCA, especially the inventory analysis, is a formalized technique to define and to find the object set and the information base which are to be considered. Thus not only the properties of the product should be used as evaluating attributes, but also those of accompanying chemical substances, of energy balances, of emissions etc. With other words: The inventory analysis includes the definition of system boundaries, before the proper Life cycle assessment begins. In *Brüggemann and Steinberg 2000* some more remarks can be found.

For other evaluation techniques, see also *Schneeweiß C., 1991 and Poschmann et al., 1998*.

Finally a comment should be given to different graphical representations: Bar diagrams: If a Hasse diagram is small enough, then the values of the attributes of each object can be drawn as a bar diagram. This would be the most comfortable graphical presentation of a partially ordered set:

- The numerical values can be derived and
- the bar diagrams of the objects are logically connected by lines, according to the order relation.

Sunrise-, star-, diamond- or amoeba diagrams⁸. The relation to a Hasse diagram is at best shown by an example:

⁸ Slightly different graphics, no unique use in the literature.

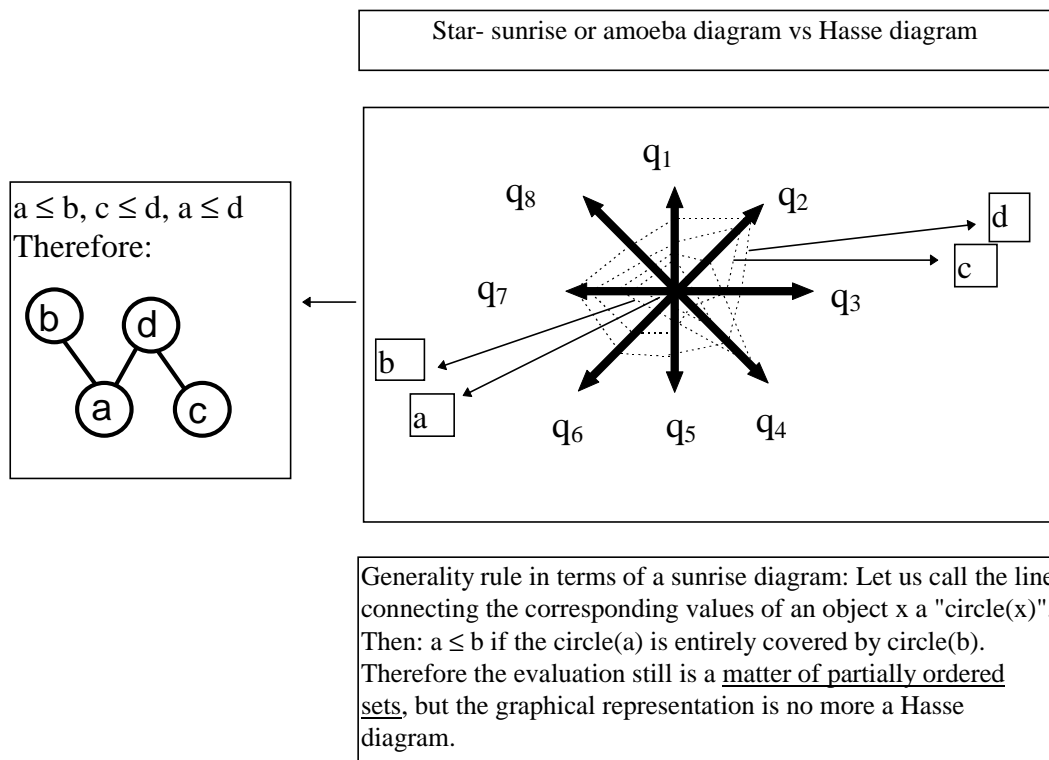


Figure 6-2: Star-, Sunrise- or amoeba diagram and their relation to a Hasse diagram

Containment order

Instead of the product order a "containment" order can be defined, for example to perform an ordering of molecules based on their shape. See for example *D.J.Klein, 1997*.

One may say: The partial order is the central concept, just the graphical representations vary. Among them the Hasse diagrams seem to be most valuable for evaluation purposes.

It is difficult, to deduce such algebraic relational arguments for the rather famous Chernov-Faces. They are useful to show the multivariate properties of some few objects, but because of rather unrestricted possibilities to model faces, a relation to partial order is hardly to establish.

Some more literature about multivariate graphical display can be found in *Hartung, Elpelt, 1992*.

7 References

Babic, D., N. Trinajstić (1996): Möbius inversion on a poset of a graph and its acyclic subgraphs. *Discrete Appl Math* 67, 5-11.

Bartel, H.-G., R. Brüggemann (1998): Application of formal concept analysis to structure-activity relationships. *Fresenius J Anal Chem* 361, 23-28.

Bartel, H.-G. (1996): *Mathematische Methoden in der Chemie*. Oxford: Spektrum Akademischer Verlag.

Bartel, H.-G., P.E. John (1999): Formalbegriffsanalytische Untersuchung von Struktur-Eigenschafts-Beziehungen an Stannocanen und Stannotranen unter Verwendung von ^{119}Sn -NMR-Daten. *Zeitschr für physikalische Chemie* 209, 141-158.

Birkhoff, G. (1984): *Lattice Theory*. Providence, American Mathematical Society; Vol. 25, 1940 (reprint 1984).

Bollobás, B. (1986): *Combinatorics*. Cambridge University Press.

Brüggemann, R. (1998): Workshop on: Order Theoretical Tools in Environmental Sciences. Introduction. In Group Pragmatic Theoretical Ecology (Ed.), *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences*. Berichte des IGB Heft 6, Sonderheft I, 1998. (pp. 1-9). Berlin: IGB.

Brüggemann, R., C. Steinberg (1999): Ein Ansatz zur Methodik der vergleichenden Bewertung von Bodenbelastungen. *UWSF - Z Umweltchem Ökotox* 11, 135-143.

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

Brüggemann, R., H.-G. Bartel (1999): A Theoretical Concept to Rank Environmentally Significant Chemicals. *J Chem Inf Comp Sc* 39, 211-217.

Brüggemann, R., K. Voigt (1996): Stability of Comparative Evaluation, - Example : Environmental Databases. *Chemosphere* 33, 1997-2006.

Brüggemann, R., S. Pudenz, K. Voigt, A. Kaune, K. Kreimes (1999a): An algebraic/graphical tool to compare ecosystems with respect to their pollution. IV: Comparative regional analysis by Boolean arithmetics. *Chemosphere* 38, 2263-2279.

Brüggemann, R., C. Bücherl, S. Pudenz, C. Steinberg (1999b): Application of the concept of Partial Order on Comparative Evaluation of Environmental Chemicals. *Acta hydrochim hydrobiol* 27, 170-178

Brüggemann, R., H. Behrendt, K.-P. Seiler, S. Trapp, K. Voigt (1994): Einsatz der Hasse-Diagrammtechnik zum Vergleich von Chemikalien hinsichtlich ihrer Umweltgefährdungspotentiale. *DGM* 38, 122-127

Brüggemann, R., J. Benz, (1988): EDV-gestützte Bewertung der Umweltgefährlichkeit von Chemikalien. In J. Gasteiger (Ed.), *Software-entwicklung in der Chemie* 2. (pp. 331-367). Berlin: Springer Verlag.

Brüggemann, R., L. Zelles, Q.Y. Bai, A. Hartmann (1995): Use of Hasse Diagram Technique for Evaluation of Phospholipid Fatty Acids Distribution in Selected Soils. *Chemosphere* 30(7), 1209-1228.

Davey B.A., H.A. Priestley, (1990): Introduction to Lattices and Order. *Cambridge University Press*, Cambridge

Duquenne, V. (1987): Contextual implications between attributes and some representation properties for finite lattices. In B. Ganter, R. Wille, & K.E. Wolff (Eds.), *Beiträge zur Begriffsanalyse*. (pp. 213-239). Mannheim: BI Wissenschaftsverlag.

Eigen, M., R. Winkler, (1979): *Das Spiel - Naturgesetze steuern den Zufall* -. Piber & Co.

Essam, J.W. (1977): The Graph-like State of Matter. Part 8.-LCGI Schemes and the Statistical Analysis of Experimental Data. *J Chem Soc Faraday Trans II* 73, 1289-1307.

Galassi, S., A. Provini, E. Halfon (1996): Risk Assessment for Pesticides and Their Metabolites in ater. *Intern J Environ Anal Chem* 65, 331-344

Ganter, B., R. Wille, (1996): *Formale Begriffsanalyse Mathematische Grundlagen*. Berlin: Springer-Verlag.

Halfon, E. (1989): Comparison of an Index Function and a Vectorial Approach Method for Ranking Waste Disposal Sites: *Environ. Science Technol.* 23, 600-609.

Halfon, E., S. Galassi, R. Brüggemann, A. Provini (1996): Selection of Priority Properties to Assess Environmental Hazard of Pesticides. *Chemosphere* 33, 1543-1562.

Hartung, J., B. Elpelt, (1992): *Multivariate Statistik Lehr- und Handbuch der angewandten Statistik*. München: R.Oldenbourg Verlag.

Klein, D.J., D. Babic (1997): Partial orderings in Chemistry. *J Chem Inf Comp Sc* 37, 656-671.

- Klein, D.J. (1986):** Chemical Graph-theoretic Cluster Expansions. *Int J Quant Chem* 20, 153-171.
- Klöpffer, W., S. Volkwein, (1995):** Bilanzbewertung im Rahmen der Ökobilanz. In Thom'e-Kozmiensky & K. Joachim (Eds.), *Management der Kreislaufwirtschaft*. (pp. 336-340). EF-Verlag für Energie- und Umwelttechnik GmbH.
- Lundgren, J.R. (1989):** Food webs , Competition graphs, Competition-common enemy graphs, and niche graphs. In F. Roberts (Ed.), *Applications of Combinatorics and Graph theory to the Biological and Social Sciences*. (pp. 221-243). New York: Springer-Verlag.
- Mezey, P.G. (1993): Shape in Chemistry An Introduction to Molecular Shape and Topology.** Weinheim: VCH.
- Poschmann, C., C. Riebenstahl, E. Schmidt-Kallert, (1998): Umweltplanung und -bewertung.** Gotha: Klett-Perthes.
- Pudenz, S., R. Brüggemann, B. Luther, (1998):** Combination of Fuzzy Clustering and Hasse Diagrams in Comparative Regional Evaluation. In Group Pragmatic Theoretical Ecology (Ed.), *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences , Berichte des IGB , Heft 6, Sonderheft I, 1998*. (pp. 95-99). Berlin: IGB.
- Pudenz, S., R. Brüggemann, D. Komossa, K. Kreimes (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* 36, 441-450.
- Pudenz S., R.Brüggemann, A. Kaune, K. Kreimes 1999:** Submitted to Chemosphere.
- Rota, G.-C. (1964):** On the foundations of Combinatorial Theory I: Theory of Möbius functions. *Z Wahrscheinlichkeitstheorie* 2, 340-368.
- Ruch, E., B. Lesche (1978):** Information extent and information distance. *J Chem Physics* 69, 393-401.
- Ruch, E. (1993):** Der Richtungsabstand. *Acta Applicandae Mathematicae* 30, 67-93.
- Ruch, E. (1975):** The diagram lattice as Structural Principle. *Theor Chim Acta* 38, 167-183.
- Ruch, E. (1972):** Algebraic Aspects of the Chirality Phenomenon in Chemistry. *Accounts of the Chemical Research* 5, 49-56.
- Schmalz, T.G., D.J. Klein, B.L. Sandleback. (1992):** Chemical Graph-Theoretical Cluster Expansion and Diamagnetic Susceptibility. *J Chem Inf Comp Sc* 32, 54-57.

Schneeweiss, C. (1991): *Planung 1 - Systemanalytische und entscheidungstheoretische Grundlagen*. Berlin: Springer-Verlag.

Seip, K.L. (1994): Restoring Water Quality in the Metal Polluted Soerfjorden, Norway. *Ocean & Coastal Management* 22 19-43.

Sørensen, P.B., B.B. Mogensen, S. Gyldenkaerne, A.G. Rasmussen. (1998): Pesticide Leaching Assessment Method for Ranking both Single substances and Scenarios of Multiple Substance Use. *Chemosphere* 36, 2251-2276.

Sørensen, P.B., B.B. Mogensen, L. Carlsen, M. Thomsen (1998): The role of uncertainty in Hasse diagram ranking. In Group Pragmatic Theoretical Ecology (Ed.), *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences*, *Berichte des IGB*, Heft 6, Sonderheft I, 1998. (pp. 71-84). Berlin: IGB.

Sørensen, P.B. (1999): publication in preparation.

Trotter, W.J. (1992): *Combinatorics and Partially Ordered Sets Dimension Theory*. Baltimore, Maryland: The Johns Hopkins University Press.

Urrutia J. (1987): Partial Orders and Euclidian Geometry, in Rival I. (Ed.), *Algorithms and Order*, NATO ASI Series, Series C: Mathematical and Physical Sciences Vol. 255, Kluwer Academic Publishers, Dordrecht, S. 387-434.

Voigt, K., J. Benz, R. Brüggemann, R. (1996): The Internet as an Information Source for Environmental Chemicals, -first results of the evaluation of the metadatabase of Internet resources. In D.I. Raitt & B. Jeapes (Eds.), *Online Information* 96. (pp. 151-159). Oxford: Learned Information Europe Ltd.

Voigt, K., R. Brüggemann (1995): Toxicology Databases in the Metadatabank of online Databases. *Toxicology* 100, 225-240.

Voigt, K. (1998): Hasse Diagram Technique as an Effective Tool in Environmental Information Management. In Group Pragmatic Theoretical Ecology (Ed.), *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences*, *Berichte des IGB*, Heft 6, Sonderheft I. (pp. 55-66). Berlin: IGB.

Voigt, K., R. Brüggemann (1998): Evaluation Criteria For Environmental and Chemical Databases. *Online & CD-ROM Review* 22, 247-261.

Volkwein, S., R. Gühr, W. Klöpffer (1996): The Valuation Step within LCA Part II: A Formalized Method of Prioritization by Expert Panels. *Int J LCA* 1(4), 182-192.

Welzl, G., K. Voigt, K., G. Rediske (1998): Visualisation of Environmental Pollution - Hasse diagram technique and Explorative Statistical Methods. In Group Pragmatic Theoretical Ecology (Ed.), **Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences**, *Berichte des IGB, Heft 6, Sonderheft I.* (pp. 101-110). Berlin: IGB.

Welzl, G., K. Voigt, K., G. Rediske (1999): Methodische Fragen aus der Sicht der Multivariaten Statistik zur Bewertung von Umweltobjekten. In Wasserforschung e.V. (Ed.), **Methodische Ansätze in der Umweltbewertung**, *Dokumentation eines Workshops vom 20. November 1998 im Rahmen des Forschungsvorhabens "Nachhaltige Wasserwirtschaft - Entwicklung eines Bewertungs- und Prüfsystems".* (pp. 59-70). Berlin: IFV Wasserforschung e.V.

Uncertainty in partial order ranking

P. B. Sørensen, L. Carlsen and B. B. Mogensen

*The National Research Institute in Denmark
Department of Environmental Chemistry
Frederiksborgvej 399
P.O. Box 358
DK-4000 Roskilde
Denmark*

Abstract

Environmental data are often associated with a significant degree of uncertainty which enters into the ranking analysis. This paper investigates the influence of the input uncertainty using a simple example from which some general "rules of thumb" can be derived. The focus is the relationship between two elements, which are compared to each other. The approach is a simple mathematical analysis, assuming equal distributed uncertainty and a constant uncertainty interval, however, the analysis seems to give some general relationships. It is demonstrated that a partial order ranking will be rather robust against uncertainties in the input in relation to the identified ranks in the Hasse diagram. If the Hasse diagram predicts a rank between two elements it appears unlikely that this rank should have been inversely ($d > c$ instead of $c > d$) due to the uncertainty of input values. The uncertainty may to some extent affect the question about comparable/incomparable.

1 Introduction

Complex information

Uncertainty

The method of partial order ranking has been used within the environmental area for a variety of purposes as an attractive way of handling complex information (*e.g. Brüggemann and Barthel, 1999; Brüggemann et al., 1995; Brüggemann et al., 1994; Halfon and Reggiani, 1986; Halfon et al., 1996; Galassi et al., 1996 and Sørensen et al., 1998*). However, environmental data are often associated with a significant degree of uncertainty which enters into the ranking analysis. The uncertainty may original from three different sources:

1. Uncertainty due to the assumed relationship between the attributes used and the phenomenon described by the ranking
2. Uncertainty from incomplete ranking originating from limited number of elements. Thus, since a given element are compared to a fraction of the total number of elements only an uncertainty of the actual rank of that element may be introduced.
3. Uncertainty due to input uncertainty

The different sources to uncertainty have to be managed differently. The type 1 uncertainty can be minimized by increasing the number of different attributes, so a large number of various aspects will be taken into account in the ranking. However, the type 1 and type 2 uncertainties are to a certain extend inversely related, since an increasing number of attributes will lead to a decreased number of comparisons in the Hasse diagram and thus a larger uncertainty of the type 2 hence, it is necessary to make a compromise between the demand of a minimum number of attributes in order to avoid a large type 1 uncertainty on the one hand, and on the other hand a demand of a limited number of attributes in order to minimize the type 2 uncertainty. Actually, what we are facing here is formulated in more general terms from the Statistical Learning Theory (*Vapnik, 1995*) as a main principle for solving problems using a restricted amount of information:

When solving a given problem, try to avoid solving a more general problem as an intermediate step.

The uncertainty from input

The uncertainty from input (type 3) is simply the uncertainty induced from the variability in the input parameters. Such variability may obviously be due to a true variability or to errors in the procedure used to determine the values. This paper will investigate the influence of the input uncertainty using a simple example from which some general “rules of thumb” can be derived. The focus will be on the relationship between two elements. Relationships in overall ranking results involving many elements is investigated in an other paper (*Sørensen et al., 2000*)

2 A simple example using two attributes

Uncertainty related to the rank of one element

The scope of the paper is to evaluate the ranking uncertainty due to uncertainties in the input parameters (attribute values). However, the influence from the input uncertainties is further affected by the uncertainty due to the limited number of comparisons. It seems obvious that the uncertainty related to the rank of one element (in the following denoted a) as a result of uncertainties in the input is a consequence of both the actual uncertainty of the attribute values for a and the uncertainty of the attribute values for all the elements which are compared to a . A further complicating factor can be explained by considering a second element in the Hasse Diagram denoted b . In the diagram b may be predicted to be incomparable to a , but by considering the uncertainties in the attributes values for both a and b , we can conclude that the incomparability between a and b it is not necessarily true. This is illustrated by a simple example in Figure 1, where 4 elements are compared using two attributes.

The elements:

$$a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \quad d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$$

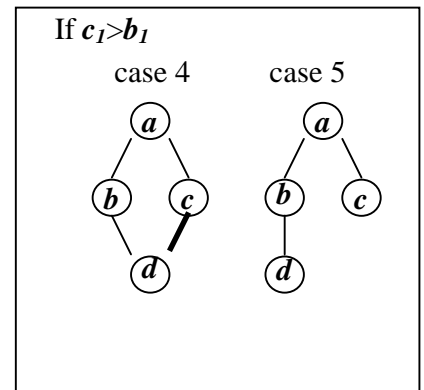
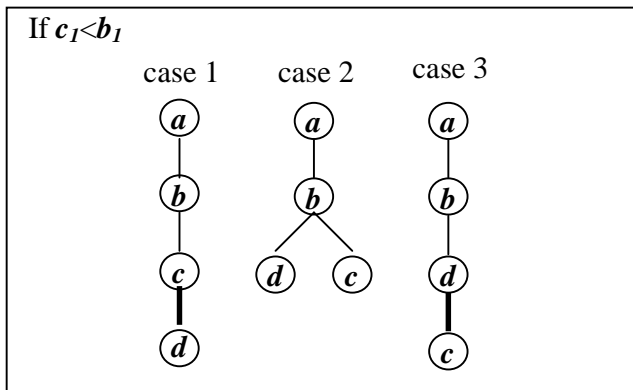
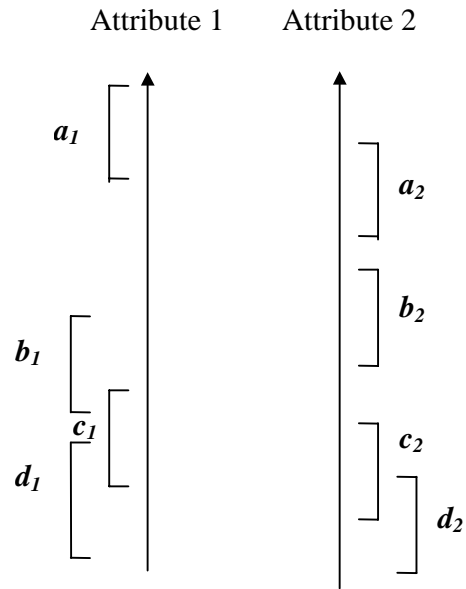


Figure 1. The uncertainty intervals of each attribute are shown on the axes and possible overlaps are identified between the attributes. All possible Hasse diagrams are shown in the boxes and the solid line shows the comparisons with can conflict due to the uncertainty (c, d).

In Fig. 1, 5 different rankings are possible as a result of the uncertainty intervals of the attribute values. Only one pair of parameters (c , d) are subject to a potential contradiction in the ranking, since both rankings c above d and d above c are possible. However, it is more likely that c is above d than opposite because the attribute values for the element c are most likely to be above d . How much likely c will be above d will be illustrated in the following.

Fig. 2 shows the relationship between the elements c and d in Fig. 1

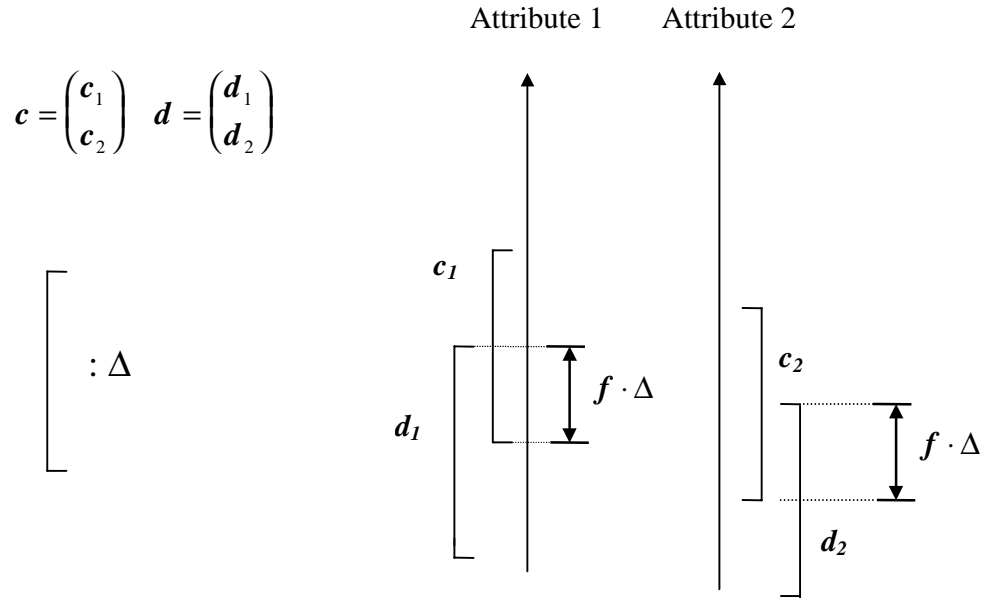


Figure 2. Scale up of the relationship between the elements c and d (c.f. Fig. 1). The overlapping value of the uncertainty interval (Δ) is denoted $f \cdot \Delta$, where the f is the fraction of overlap between the two intervals.

An overlapping factor f

In Fig. 2 the uncertainty intervals are overlapping for attribute 1 and attribute 2 respectively the overlap being quantified by an overlapping factor f , which in this illustrative example for simplicity is assumed equivalent for the two attributes. Thus, the uncertainty intervals of the attributes c_1 and d_1 are assumed to overlap to the same degree as the uncertainty intervals between c_2 and d_2 . The uncertainty intervals are further scaled up in Fig. 3 for the attribute 1 only.

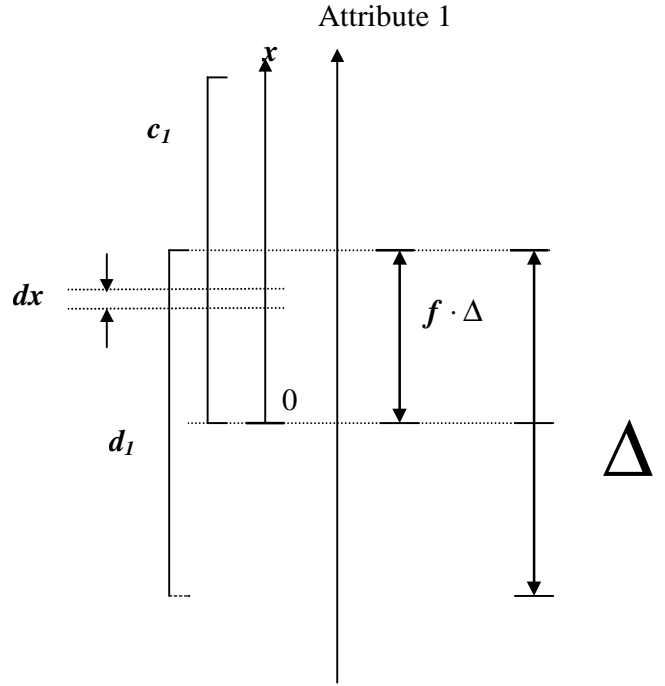


Figure 3. A detailed description of the uncertainty between the elements d_I and c_I . The minimum value of c_I is defined as zero on the attribute 1 axis (x).

Attribute values equally distributed

Assuming the attribute values equally distributed in the uncertainty interval (same probability for any number in the interval) the probability for a specific smaller interval inside the uncertainty interval is:

$$p_{dx} = \frac{dx}{\Delta} \quad (1)$$

where dx denotes a “small” interval inside the uncertainty interval and p_{dx} is the probability to enter that interval. The probability for selecting a c value between zero and $f \cdot \Delta$ can be calculated using a simple ratio or by integration of Eq. 1, the latter being used in the following derivations as:

$$p_{c_1 < f \cdot \Delta} = \frac{f \cdot \Delta}{\Delta} = f = \int_0^{f \cdot \Delta} \frac{dx}{\Delta} \quad (2)$$

where $p_{c_1 < f \cdot \Delta}$ is the probability for selecting a c value between zero and $f \cdot \Delta$. The argument for the integral equation is that the probability for selecting a number in the interval between zero and $f \cdot \Delta$ equals to the sum of all the “small” probabilities for selecting a “little” interval dx between zero and $f \cdot \Delta$. If a number is selected on the x axis in Fig. 3 (denoted x), in the interval between zero and $f \cdot \Delta$ and a second number (denoted d_I) is selected in the uncertainty interval for d_I then the probability for the second number to be above the first number is

$$p_{d_1 > x} = \frac{f \cdot \Delta - x}{\Delta} \quad (3)$$

The probability for selecting one number for c_I in the interval between zero and $f \cdot \Delta$ and a number for d_I which is above the selected c_I number can be calculated by combining Eq. 2 and Eq. 3:

$$p_{d_1 > c_1} = \int_0^{f \cdot \Delta} \frac{f \cdot \Delta - x}{\Delta} \cdot \frac{dx}{\Delta} \quad (4)$$

The solution of Eq. 4 is

$$p_{d_1 > c_1} = \frac{1}{2} \cdot f^2 \quad (5)$$

The complementary probability for $c_I > d_I$ is given by minus the probability for $d_I > c_I$:

$$p_{c_1 > d_1} = 1 - \frac{1}{2} \cdot f^2 \quad (6)$$

The probability number

The equations 5 and 6 can be used to calculate the probability number with respect to two attributes as the relationships between the second attribute values obviously are the same as the relationships between the first attribute values as long as the two attributes are assumed uncorrelated. The probability for the element d to be above the element c is then (using Eq. 5)

$$p_{d > c} = p_{d_1 > c_1} \cdot p_{d_2 > c_2} = \frac{1}{4} \cdot f^4 \quad (7)$$

where f is assumed equivalent for both attributes. Eq. 6 can similarly be used to calculate the probability for $c > d$ as

$$p_{c > d} = p_{c_1 > d_1} \cdot p_{c_2 > d_2} = \left(1 - \frac{1}{2} \cdot f^2\right)^2 \quad (8)$$

The probability for c and d to be incomparable is given as (using Eqs. 7 and 8)

$$p_{c \text{ incom } d} = 1 - p_{d > c} - p_{c > d} = f^2 - \frac{1}{2} \cdot f^4 \quad (9)$$

The Eqs. 7, 8 and 9 are shown as curves for all possible f intervals in Fig. 4. and it is clearly demonstrated that the f value needs to be rather large before both the rankings $d > c$ and $c > d$ are likely to happen. For f values below 0.5 it is virtually unlikely that $d > c$, whereas, there may exist some uncertainty covering the statement: comparable/incomparable.

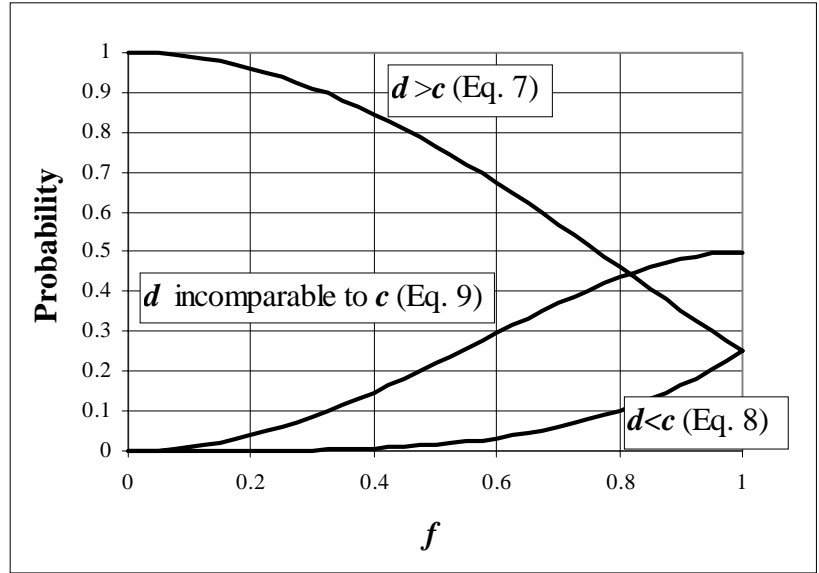


Figure 4. A graphically presentation of the Eqs. 7, 8 and 9.

3 Conclusion

The present paper has demonstrated in simple terms by an example that a partial order ranking will be rather robust against uncertainties in the input in relation to the identified ranks in the Hasse diagram. If the Hasse diagram predicts a rank between two elements it appears very unlikely that this rank should have been inversely ($d > c$ instead of $c > d$) due to the uncertainty of input values. The uncertainty may to some extent affect the question about comparable/incomparable.

4 References

Brüggemann, R. and Bartel, H. -G. (1999): A Theoretical Concept to Rank Environmentally Significant Chemicals, *J. Chem. Inf. Comp. Sc.*, Vol. 39, pp. 211-217.

Brüggemann, R., Schwaiger, J. and Negele, R.D., (1995): Applying Hasse Diagram Technique for the Evaluation of Toxicological Fish Tests. *Chemosphere*, Vol. 30, pp.1767-1780.

Brüggemann, R., Voigt K. and Münzer B., (1994): The Technique of Hasse Diagrams Applied on Environmental Online Databanks. *Chemosphere*, Vol. 29, pp. 683-691.

Halfon, E. and Reggiani, M. G., (1986): On ranking chemicals for environmental hazard, *Environmental and Science Technology*, 23, 1173-1179

Halfon, E. ,Galassi, R., Brüggemann, R. and Provini, A., (1996): Selection of priority properties to assess environmental hazard of pesticides. *Chemosphere*, Vol. 33, pp. 1543-1562

Galassi, S., Provini, A., and Halfon, E.(1996): Risk Assessment for Pesticides and Their Metabolites in Water, (1996), *Intern. J. Environ. Anal. Chem.* Vol. 65, pp. 331-344.

Sørensen, P. B., B. B. Mogensen, S. Gyldenkaerne and A. G. Rasmussen, (1998): Pesticide Leaching Assessment Method for Ranking Both Single Substances and Scenarios of Multiple Substance Use. *Chemosphere*, Vol. 36. pp. 2251-2276.

Sørensen, P.B., Mogensen, B.B., Carlsen L. and Thomsen, M., (2000): The Influence on Partial Order Ranking From Input Parameter Uncertainty, *Definition of a robustness parameter*, Accepted for publication in *Chemosphere*

V. N. Vapnik, *The Nature of Statistical Learning Theory*, 1995, Springer-Verlag New York, Inc.

Cluster analysis and Its Use in Hasse Diagram Technique

B. Luther, R. Brüggemann*, S. Pudenz

Institute of Freshwater Ecology and Inland Fisheries

Department: Ecohydrology

E-Mail (Brüggemann): Brg@IGB-Berlin.de

Abstract

The starting point is the assessment of a collection of objects which are described by multiple criteria. The traditional way of assessment for this kind of data is the assessment with a ranking index which is in most cases derived by the construction of **weighted sums**. We want to replace this line of action by the application of **order theory** and (in consequence) **Hasse diagram technique**. The concept of partially ordered sets (abr. posets) is well known and used for quite some time (think for example about scheduling problems) but on the field of multicriteria data analysis it is relative new and so complete new problems arise.

Basically we need a partial order on \mathbb{R}^n (our objects are associated to tuples of real numbers), the most simple one seems to be the product order which is our starting point (we will discuss a few alternatives). Here we see the first problem, the problem of **robustness** or **stability**: *criteria* are mostly real numbered values (with a continuous distribution) therefore it makes no sense to claim a continuous connection between the criteria of the objects and the derived partial order. But we *have* to guarantee somehow, that slight, insignificant differences of our data (noise) do not lead to big changes of the results. This is one reason for the importance of **cluster analysis** on the field of order theory. Another reason is the wish to reduce the "complexity" of a poset, which is no longer readable or in other words, the wish to find a "coarse" structure of complex posets.

We need a theoretical description of both posets and cluster analysis and objective methods to assess our assessments. (I.e. an answer to the question: Is the poset (the clustering) a "good" one for a given set of dates?).

1 Introduction

This text disperses into four parts: *this* section, the introduction, shall motivate our line of action and gives in the end a formal description of the basic model. In the second section we will describe briefly the most important things of order theory, in the third section we will go into the field of cluster analysis. Finally the last section will show some strategies to combine methods of both fields.

Often one is faced with the evaluation of objects: What is the price of an article, what are the costs of some engineering constructions? In environmental sciences typically the evaluation has to be performed regarding several criteria [3]. For example, objects of such an environmental evaluation may be chemicals (as shown in [3]), or geographical areas. In any case, in order to evaluate objects, a tuple of data is needed. Those data are considered as helpful to describe the objects with respect to the criteria by which they are to be evaluated. The evaluation requires a comparison of objects, therefore often a **ranking index** (see 1.2) is introduced (in most cases a **weighted sum**). However the use of a ranking index implies that the combination of criteria can be quantitatively described. Typically in environmental sciences there is no consensus on how to do this. Therefore the concept of a partial order appears to be extremely helpful, to perform at least a comparative evaluation. A simple variant is done by ordering the objects corresponding to the componentwise order of their tuples of data (product order) – this has the advantage, that every ranking index delivers a linear extension of our so-called product order (hence we really have a generalization). (See for details: "Introduction to the General Principles of the Partial Order Ranking Theory" (this issue) and "New Tools in Hasse Diagram Technique" (especially for an explanation of linear extensions; also this issue))

Ranking index

No consensus on how to do this

In short we arrive at the following agreements:

1.1 Agreements / Model

1. Let $O = \{o_1, \dots, o_N\}$ a finite set of objects, the so called **object set** or **ground set**, which can be seen (in some cases) as subset of a **totality** G ($O \subseteq G$).
1. $IB = \{\alpha_1, \dots, \alpha_{|IB|}\}$ forms the **attribute set** or **information base** (a finite set of $|IB|$ attribute names is simply enumerated). We call the elements of **IB criteria** as well. With **IB** we have a first description about the objects from O .
2. $\mathcal{A} = \{\mathcal{A}_1, \dots, \mathcal{A}_{|IB|}\}$ is the family of sets of possible values – so \mathcal{A}_i denotes the set of possible values of attribute α_i (for example: $\alpha_i = \text{"age"}$, $\mathcal{A}_i = \mathbb{N}$ with some subset O of all humans). We take in all cases $\mathcal{A}_i \subseteq \mathbb{R}$ for granted in order to guarantee that
 - O is totally ordered relating to one *single* criterion and
 - all attributes can be expressed numerically.

That does not mean, that we can only deal with *quantitative* data! (For

Elements of IB criteria

example: $\alpha_i = \text{"investment risk"}$ with $\mathcal{A}_i = \{0, 1, 2\}$ instead of $\mathcal{A}_i = \{\text{"low"}, \text{"middle"}, \text{"high"}\}$.

3. In consequence we assign a matrix $X \in \mathbb{R}^{N \times |B|}$ with attribute values to O. For X we have:

$$(\forall i \in \{1, \dots, |B|\})(\forall k \in \{1, \dots, N\}) \quad x_{ki} \in \mathcal{A}_i$$

where x_{ki} denotes the value of the i -th attribute α_i from object o_k – one row of X refers to one object of O (of course the rules of rows and columns are interchangeable, in many approaches of HDT the objects correspond to the rows so we do here the same just for the sake of identity).

1. Furthermore we do a standardization to X, i.e. we define a matrix $Y \in \mathbb{R}^{N \times |B|}$ to O, by

$$(\forall i \in \{1, \dots, |B|\})(\forall k \in \{1, \dots, N\}) \quad y_{ki} = \frac{x_{ki} - \mu_i}{\sigma_i}$$

where μ_i is the (estimated) mean and σ_i is the (estimated) standard derivation of attribute α_i . Then the euclidian distance between any two rows of Y reproduces the dissimilarity between their corresponding objects in O. We further need this standardization to guarantee *scale invariance* for cluster analysis but since it is order preserving, it does not touch the order theoretical aspects of our examination.

Scale invariance

Further we have the following simple description of a ranking index and a weighted sum.

1.2 Ranking index and weighted sum (definition)

Let $r : \mathbb{R}^n \rightarrow \mathbb{R}$ a monotonous increasing mapping (i.e. *componentwise monotonous*). Then we call r a **ranking index**. If we have $(\forall x \in \mathbb{R}^n) r(x) = c^T \cdot x = \sum_{i=1}^{|B|} c_i \cdot x_i$ (with $c_i > 0$ for all i) we call r a **weighted sum** (with positive weights).

We are describing our objects by multiple real criteria. For a cluster analysis we have to precise the *dissimilarity* of the objects and perhaps (for some methods) a suitable representation in the euclidian space. The *assessment* of the objects under *all* criteria will be done by *partial orders*.

2 Order Theory

2.1 Posets Ia: Base Concept

2.1.1 Partially ordered sets (definition)

Let $\mathbf{P} := (O, P)$ a digraph. If it has the three properties:

1. $(\forall a \in O) \quad (a, a) \in P$ (reflexivity)
2. $(\forall a, b \in O) \quad (a, b), (b, a) \in P \Rightarrow a = b$ (antisymmetry)
3. $(\forall a, b, c \in O) \quad (a, b), (b, c) \in P \Rightarrow (a, c) \in P$ (transitivity)

we call the relation P a **partial order** (or briefly **order**) and \mathbf{P} a **partial ordered set** (briefly: **poset**). Instead of " $(a, b) \in P$ " (" $(a, b) \notin P$ ") we write " $a \leq b$ in \mathbf{P} " (" $a \not\leq b$ in \mathbf{P} ") – in words: " a is (not) less or equal b in \mathbf{P} ". If $a \not\leq b$ and $b \not\leq a$ in \mathbf{P} we write in contrary $a \parallel b$ (in words: " a and b are incomparable"). An order is called **total** or **linear** if we have in addition

4. $(\forall a, b \in O) \quad a \leq b \text{ or } b \leq a \text{ in } \mathbf{P}$ (totality)

– a total or linear order contains no pair a, b with $a \parallel b$.

2.1.2 Extensions (definition)

Let $\mathbf{P} := (O, P)$ and $\mathbf{Q} := (O, Q)$ two posets with the same object set O . If $P \subseteq Q$ we call \mathbf{Q} an extension of \mathbf{P} (and Q an extension of P). Hence it is clear what a **linear extension** is.

2.1.3 Intervals, Ideals and Filters

Let $\mathbf{P} := (O, P)$ a partial ordered set and $a, b \in O$. Then we call the set $[a, b] := \{x \in O \mid a \leq x \leq b\}$ an **interval**, the set $[a] := \{x \in O \mid x \leq a\}$ a **principal ideal** and the set $[a) := \{x \in O \mid a \leq x\}$ a **principal filter**.

Principal ideal

Principal filter

2.2 Posets Ib: Assessment of orders

If we want to "assess our assessments" which are appearing as posets, we need a measure which describes the quality of a poset itself. Our idea is to assume the existence of something as a "natural poset" $\mathbf{P}^* := (O, P^*)$ on our object set and compare each other partial order $\mathbf{P} := (O, P)$ on the same object set with \mathbf{P}^* , i.e. the more "similar" an order is to \mathbf{P}^* the nearer it is considered to the natural one. In consequence we have to quantify the similarity (or *dissimilarity*) of posets which have the same object set.

Similarity (or dissimilarity) of posets

We therefore introduce one dissimilarity measure (the W-distance) and one similarity measure (the Tanimoto Index) for posets.

2.2.1 The W-distance – a dissimilarity measure for posets

Let $\mathbf{P} := (O, P)$ and $\mathbf{Q} := (O, Q)$ two posets. Then we call

$$W(\mathbf{P}, \mathbf{Q}) := W(P, Q) := \sum_{o \in O} |(o]_P \Delta (o]_Q| \text{ the } \mathbf{W}\text{-distance between } \mathbf{P} \text{ and } \mathbf{Q}$$

(i.e. between P and Q). Here " Δ " describes the symmetrical difference of two sets, so $(o]_P \Delta (o]_Q$ denotes the set of all $o' \in O$ with:

- $o' \leq o$ in \mathbf{P} and $o' \not\leq o$ in \mathbf{Q} or
- $o' \leq o$ in \mathbf{Q} and $o' \not\leq o$ in \mathbf{P}

or in shorter words: The dissimilarity of the ideals is seen as the dissimilarity of the orders. A more detailed description can be found in the "Introduction to the General Principles of the Partial Order Ranking Theory", this issue.

2.2.2 The Tanimoto-Index – a similarity measure for posets

Let $\mathbf{P} := (O, P)$ and $\mathbf{Q} := (O, Q)$ two posets. Then we call

$$T(\mathbf{P}, \mathbf{Q}) := T(P, Q) := \frac{|P \cap Q|}{|P \cup Q|} \text{ the } \mathbf{Tanimoto}\text{-index of } \mathbf{P} \text{ and } \mathbf{Q}. \text{ It is a simi-}$$

larity measure of directed graphs – we compare the sets of edges and obviously the quotient above is 0 iff both graphs have no common edges and 1 iff the edges sets are equal. Sometimes the comparison of $|P \cap Q|$ with $|P \cup Q|$ is not appropriate. Soerensen et al. suggest other modified variants.

2.2.3 Remark

There is a connection between this two measures: we have

$$T(P, Q) = 1 - \frac{W(P, Q)}{|P \cup Q|}, \text{ i.e. we could introduce a } \textit{normed} \text{ dissimilarity}$$

measure with $\bar{W}(P, Q) := \frac{W(P, Q)}{|P \cup Q|}$. Furthermore we can leave one of

this two measures, because one of them is the "mirror image" of the other (for example: minimizing one of them means maximizing the other). We will concentrate on the W-distance.

2.3 Posets Ic: Preorders and equivalence

Partial orders may express all possible ordinal relations between objects we want to show, but it is not possible to express *equivalences* between objects in a partial order. But even in the evaluation of an object set with a ranking index some *different* objects may get the *same* value, may be "equally good". (A special case of this "quality equivalence" arises, when our X has two (or more) equal rows. Of course this is feasible.) Since we are furthermore interested in cluster analysis, we have to think about modeling equivalences in a suitable graph theoretical way. Obviously we

*Tanimoto-index of P
and Q*

Weaken the demand of antisymmetry

have to take the axioms of partial orders and weaken the demand of antisymmetry. This leads us to preorders (or quasiorders).

2.3.1 Preordered sets (definition)

Let again $\mathbf{P} := (O, P)$ a digraph. If P is reflexive and transitive *but* not necessarily antisymmetric, we call P a **preorder** and \mathbf{P} a **preordered set** (briefly: **preset**). Instead of " $(a, b) \in P$ " (" $(a, b) \notin P$ ") we write " $b \lesssim a$ in \mathbf{P} " (" $a \not\lesssim b$ in \mathbf{P} ") – the meaning of " $a \parallel b$ " is the same as in posets, " $a \sim b$ " means $a \lesssim b$ and $b \lesssim a$ (orders are exact the preorders with $a \sim b \Rightarrow a = b$). (See for further details about presets: *Brüggemann, R. , Bartel, H.-G., 1999*; they use the term quoset instead preset).

2.3.2 Splitting of preordered sets

A preorder P ever includes an **equivalence relation** $\alpha(P)$ and a **strict order**¹ $\beta(P)$:

$$\alpha(P) := \{ (a, b) \in O \times O \mid a \lesssim b \text{ and } b \lesssim a \text{ in } \mathbf{P} \}$$

$$\beta(P) := \{ (a, b) \in O \times O \mid a \lesssim b \text{ and } b \not\lesssim a \text{ in } \mathbf{P} \}$$

Both $\alpha(P)$ and $\beta(P)$ are transitive and for $a \sim b$ and $b < c$ ($a < b$ and $b \sim c$) in \mathbf{P} we have $a < c$ in \mathbf{P} .

2.4 Posets II: A Preorder $\mathbf{P} = (\mathbb{R}^{|\mathbf{IB}|}, ?)$ and the role of domination sets

We assume that O is mapped on $\mathbb{R}^{|\mathbf{IB}|}$, therefore we can (and will) use geometric properties of this vectorspace to form a preorder on $\mathbb{R}^{|\mathbf{IB}|}$. For this we take up an idea of Yu [7] who describes his so-called "criteria space"² $Y \subseteq \mathbb{R}^{|\mathbf{IB}|}$ with **domination sets**, i.e. if we assume $Y = \mathbb{R}^{|\mathbf{IB}|}$ (for the sake of simplicity), he constructed a mapping $D : \mathbb{R}^{|\mathbf{IB}|} \rightarrow \wp(\mathbb{R}^{|\mathbf{IB}|})$ where $D(x) := \{ y \in \mathbb{R}^{|\mathbf{IB}|} \mid x+y \text{ is "better" (or worse) than } x \}$ is called the set of **domination factors** of x (all "directions" that lead to a dominated point) and the set $\{x\} + D(x) := \{x+y \mid y \in D(x)\}$ is called the **domination set**. In the case of a preorder P we have the relationship: $D_P(x) := \{ y \in \mathbb{R}^{|\mathbf{IB}|} \mid x+y \lesssim x \}$.

That all criteria have the same "orientation"

The starting point of working with domination sets was the minimal assumption that all criteria have the same "orientation", i.e. a higher value of any criterion is always considered as "bad". So we directly arrived at the pareto rule in opposite to the weighted sums which denote two extreme possible domination sets.

¹ A strict order is an order without its reflexive pairs.

² The "criteria space" is the set of all possible tuple of criteria. In contrary the "decision space" is the set of corresponding "decision variables". In an optimization problem our criteria space is given with all possible values of the objective (special case of 1-dimensional criteria space). It is important to see, that we have still not discussed any decision variables.

2.4.1 Weighted sums

The most disseminated method for generating a ranking index is the (positive) **weighted sum** – it corresponds with the choice $D_P(x) := H(c) := \{ y \in \mathbb{R}^{|B|} \mid c^T \cdot y \leq 0 \}$ where $c \in (0, \infty)^{|B|}$ is a vector of positive weights, i.e. we have the choice of a constant cone halfspace $H(c)$.

- Advantage: easy to calculate, linear order, numerical values of $c^T \cdot x$ yield information about stability
- Disadvantages: different quantities are offset against each other (problem of comparing "apples with pies")

"Apples with pies"

2.4.2 Pareto rule

The method we are preferring is the so called **pareto rule** (or in HDT also called "generality rule"), that leads to the constant cone $D_P(x) := \Lambda := \{ y \in \mathbb{R}^{|B|} \mid (\forall i \in \{1, \dots, |B|\}) y_i \leq 0 \}$.

- Advantage: avoiding arbitrariness in weighting the criteria
- Disadvantage: Hence we have to deal in the end with a preordered set, we loose information about stability

We like to mention expressly, that in both cases the set $D_P(x)$ of domination factors refer to x is independent of x – that is what we mean when we talk about a *constant* cone (note that every halfspace $H(c)$ has to include Λ). The problem of stability is then that no assumption is made referring to the "gap" between two different objects. Alternatively to cluster analysis, we may introduce a "gap vector" $d \in [0, \infty)^{|B|}$ which contains for every attribute a "significance threshold". We now propose 2 alternatives, which shall pay tribute to (numerical) significance of ordinal relations.

"Gap vector"

2.4.3 Improved Pareto rules

Let $d \in [0, \infty)^{|B|}$ our "gap vector". To take nonsignificant fluctuations into consideration we see two basic alternatives:

- We may define $D(x) := \{x - d\} + \Lambda = \{x - d + \lambda \mid \lambda \in \Lambda\}$, i.e. we only assume the ordinal relation " $y < x$ " if $y_i \leq x_i - d_i$ for *all* attributes (threshold has to be exceeded in *all* attributes).
- $D(x) := \{x\} + \Gamma_d$ with $\Gamma_d = \left\{ y \in \mathbb{R}^{|B|} \mid \begin{cases} (\forall i \in \{1, \dots, |B|\}) & y_i \leq 0 \\ (\exists j \in \{1, \dots, |B|\}) & y_j \leq -d_j \end{cases} \right\}$,
i.e. we only assume the ordinal relation " $y < x$ " if $y_i \leq x_i - d_i$ for *one* attribute (we still have to insist on the relation $y_i \leq x_i$ for *all* attributes in order to guarantee transitivity).

Since it is enough to discriminate the objects refer to *one* attribute we have tested the 2nd alternative (later we will give an example). The advantage of this robustification of the old pareto rule is, that it is more "careful": Contrary to the assumption that similar objects may belong to an equivalence class (the point of view of traditional cluster analysis) we consider similar objects as formally incomparable to avoid an arbitrary

Robustification

clustering. This is a realistic point of view if we see our preorder as pre-processing for a further examination (maybe a form of cluster analysis). A disadvantage: We need a method to choose suitable "gap vectors" d , so we have to think about its statistical properties. The variants discussed are schematically shown in Fig.1 (for the 2-dimensional case).

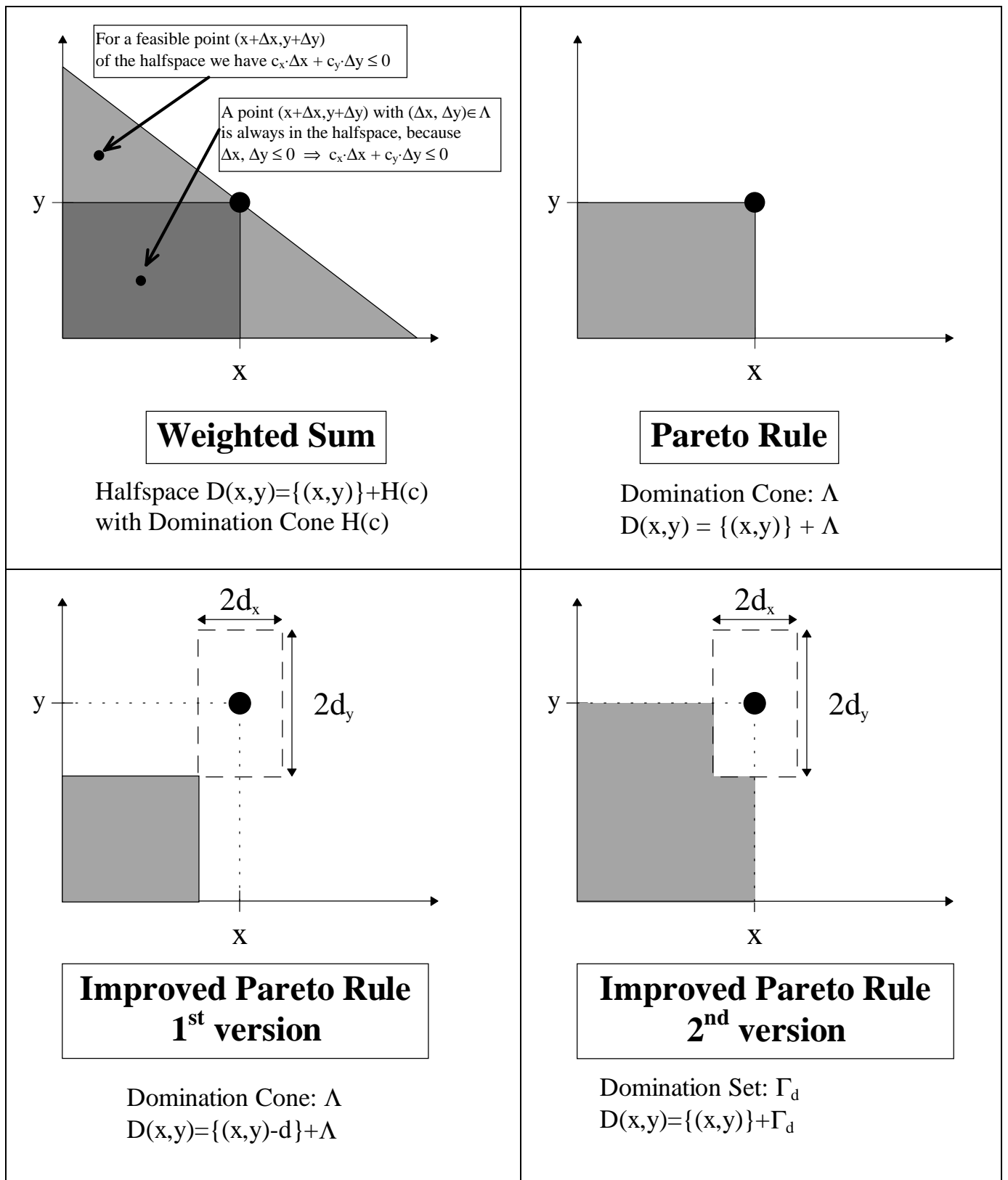


Figure 1 (part 1): Graphical presentation of different domination sets

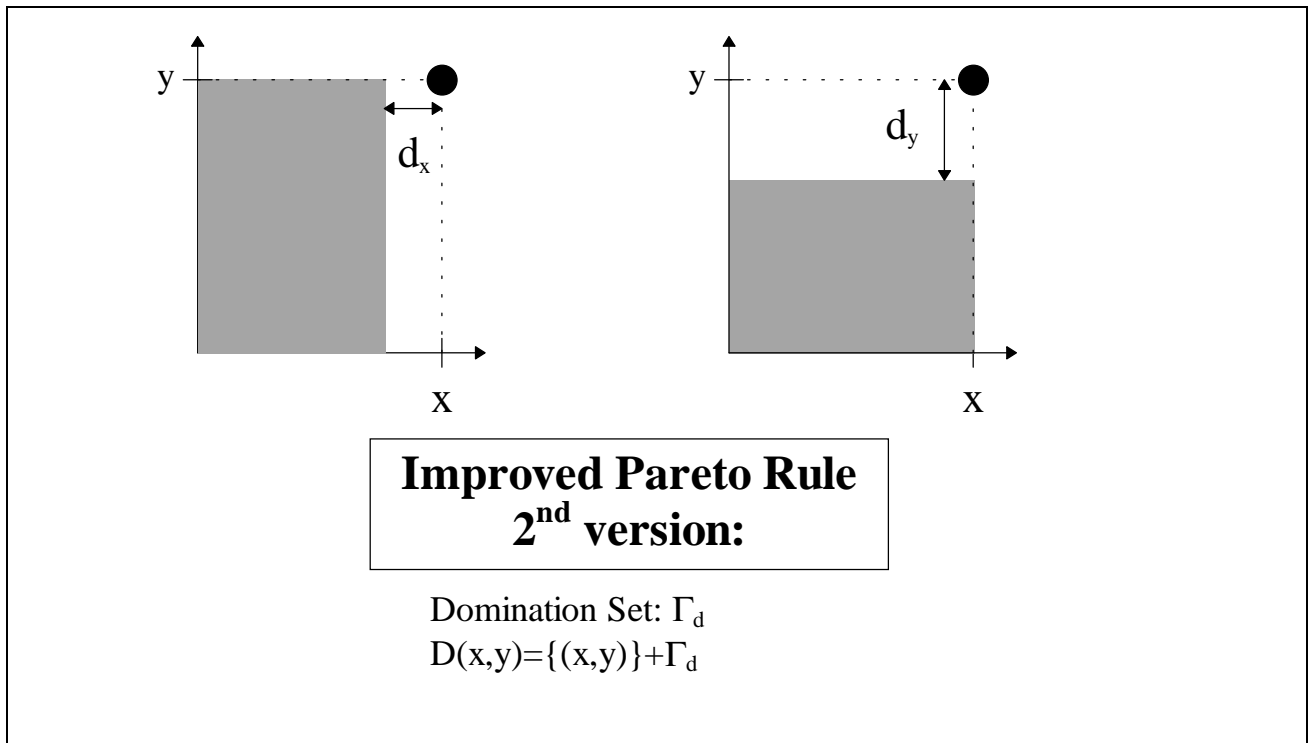


Figure 1 (conclusion): Further explanation of the 2nd version of the improved Pareto rule:
 Γ_d can be seen as "superposition" of n cones in the n -dimensional euclidian space (here: 2 cones)

3 Cluster Analysis

Separation of the clusters

The subject of cluster analysis can be described by the task to generate and **evaluate partitions of an object set**, i.e. derive properties of such partitions (this point of view is detailed discussed in H. H. Bock [4]). We define a partition as a family of pairwise disjoint subsets of an object set which covers this object set (their union equals the complete object set). Note that there are still other ways to define this term (sometimes the disjointment is given up, this leads to so-called "overlapping"). In our case the separation of the clusters is important, so we have to claim pairwise disjointment. The sets of a partition are called "**clusters**" and the process of generating a partition is called **partitioning, grouping** or **clustering**. Now the question of the goal of such grouping arises, which is given with the necessity that "a great set of unstructured objects (persons, objects, diseases or documents) has to be dispersed into smaller, homogenous classes, that are useful for practical tasks" (H. H. Bock [4]). A class is homogenous if its elements are "similar" to each other, and the **dissimilarity** is in most cases expressed by the euclidian distance of suitable representatives as we will see below.

In order to start in a formal correct way, we want to introduce some basic notions:

3.1 Cluster Analysis Ia: Partitions

3.1.1 Partitions.

Let O a finite set (of objects), and $A_1, \dots, A_m \subseteq O$ with

- $(\forall i, j \in \{1, \dots, m\}) \quad A_i \cap A_j = \emptyset$ (disjoint)
and
- $\bigcup_{i=1}^m A_i = O.$ (complete covering)

Then we call the system $\{A_1, \dots, A_m\}$ a **partition** of the object set O (of m clusters).

3.1.2 Unsharp Partitions

Let $O = \{o_1, \dots, o_N\}$ a finite set (of objects), and $U \in \mathbb{R}^{FCL \times N}$ a matrix with

- $(\forall i \in \{1, \dots, FCL\}) (\forall k \in \{1, \dots, N\}) \quad 0 \leq u_{ik} \leq 1$

and

- $(\forall k \in \{1, \dots, N\}) \quad \sum_{i=1}^{FCL} u_{ik} = 1.$

*U an unsharp partition
(or membership matrix)
of O*

Then we call U an **unsharp partition** (or **membership matrix**) of O (with FCL fuzzy clusters). (Each object is assigned to all clusters, "with different memberships u_{ik} ".) If further u_{ik} is 0/1-valued for all i, k we call U a sharp partition. Any sharp partition corresponds to a partition.

3.2 Cluster Analysis Ib: Discussion of Unsharp Partitions

3.2.1 Interpretation of unsharp Partitions.

For a sharp partition we can directly construct a unique partition

$$\{A_1, \dots, A_{FCL}\} \text{ with } u_{ik} = \begin{cases} 1 & \text{if } o_k \in A_i \\ 0 & \text{if } o_k \notin A_i \end{cases}.$$

But of course even for unsharp partitions the membership values give information about where an object belongs. We characterize objects according to their so called "purity" which is a measure how "pure" an object can be assigned to one cluster.

Purity

Purity. Let $U \in [0, 1]^{FCL \times N}$ be an unsharp partition (of a finite object set $O = \{o_1, \dots, o_N\}$), $o_k \in O$. Then we call $\text{pur}(o_k) := \max\{u_{ik} \mid i \in \{1, \dots, m\}\}$ the *purity* of the assignment of o_k .

Hybrids

Hybrids. Let $U \in [0, 1]^{FCL \times N}$ be an unsharp partition (of a finite object set $O = \{o_1, \dots, o_N\}$).

Then we call o_k a *relative hybrid* iff $\text{pur}(o_k) \in (0.5, 1)$ – we call it an *absolute hybrid* iff $\text{pur}(o_k) \in (0, 0.5]$. In general we call o_k a *TMF-hybrid* or a *hybrid* with corresponding threshold TMF ($\text{TMF} \in (0, 1]$) iff $\text{pur}(o_k) \leq \text{TMF}$. (We do not forbid the term *1-hybrid* for sharp assigned objects.) On the other hand: Iff o_k is no *TMF-hybrid* we consequently call o_k *TMF-pure*. Iff $\text{pur}(o_k) \in (0.5, 1]$ o_k is *relative pure*. After all $\text{pur}(o_k) = 1$ implies that o_k is *sharp assigned* or *pure* at all. To sum all this up we make a small decision tree (figure 2).

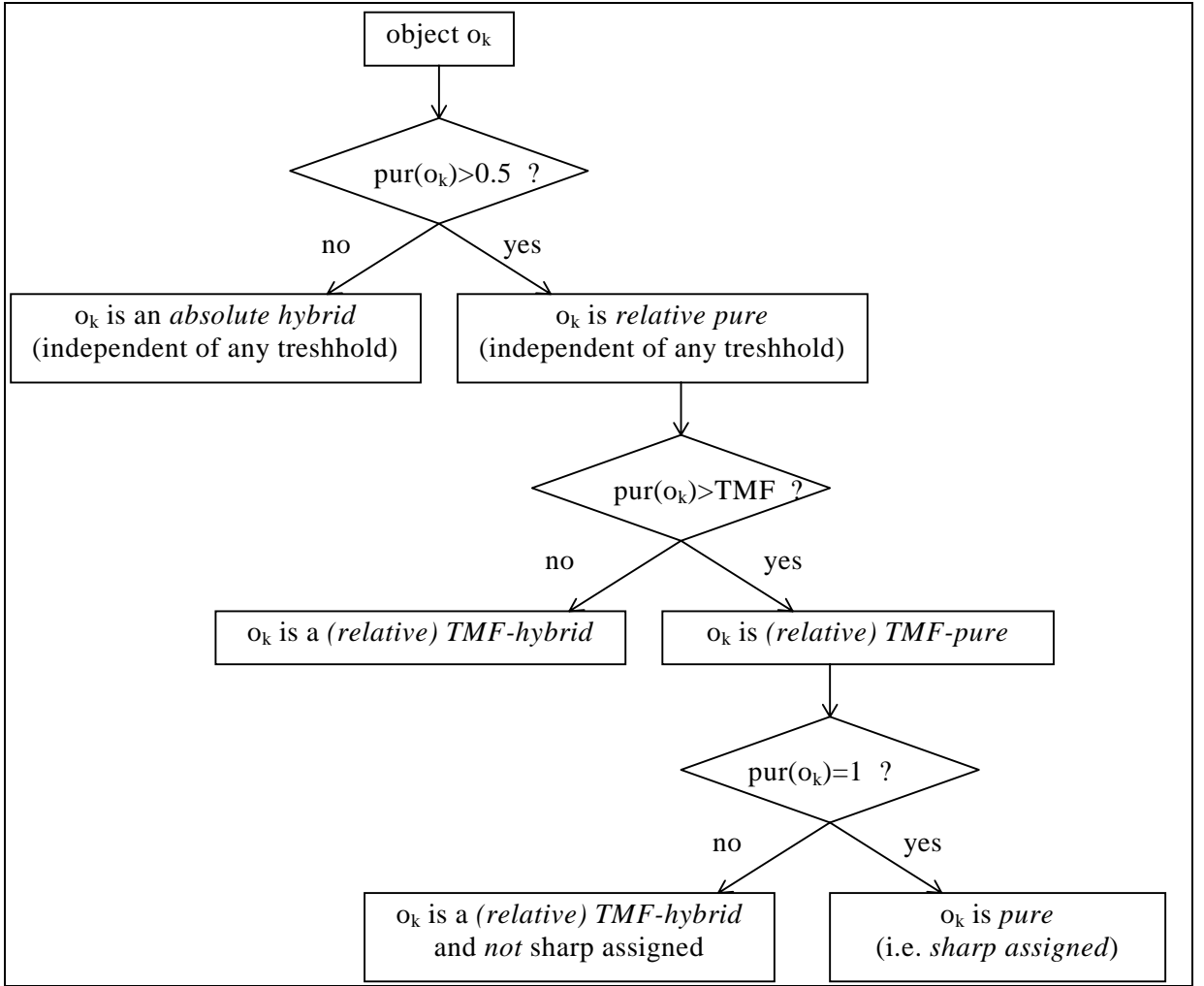


Figure 2: How to characterize objects in terms of their purity

It is easy to illustrate these terms geometrically. Assume that we want to get an unsharp partition for a set O with at most three clusters. The feasible set of possible "membership vectors" is then given by the 2-dimensional simplex (embedded in the \mathbb{R}^3) whose corners represent the three possible sharp assignments – the usual geometric representation of points on the 2-dimensional simplex is done by ternary diagrams (Fig. 3). We use this form of visualization because it shows our terms very well: The small triangle inside of the simplex forms the set of absolute hybrids – any object that is assigned to a point *out of this region* is only a *relative hybrid* because it can be considered to be a member of a *unique* cluster (corner) "according to its maximal membership $\text{pur}(o_k)$ ". In Fig. 3 we have drawn the unsharp

Ternary diagrams

$$\text{partition } U = \begin{pmatrix} \overbrace{O_1 \ O_2 \ O_3 \ O_4 \ O_5} \\ \frac{1}{3} \ \frac{1}{2} \ \frac{1}{10} \ \frac{2}{5} \ 1 \\ \frac{1}{3} \ \frac{1}{2} \ \frac{1}{10} \ \frac{3}{10} \ 0 \\ \frac{1}{3} \ 0 \ \frac{8}{10} \ \frac{3}{10} \ 0 \end{pmatrix} \begin{matrix} \text{1. cluster} \\ \text{2. cluster} \\ \text{3. cluster} \end{matrix}$$

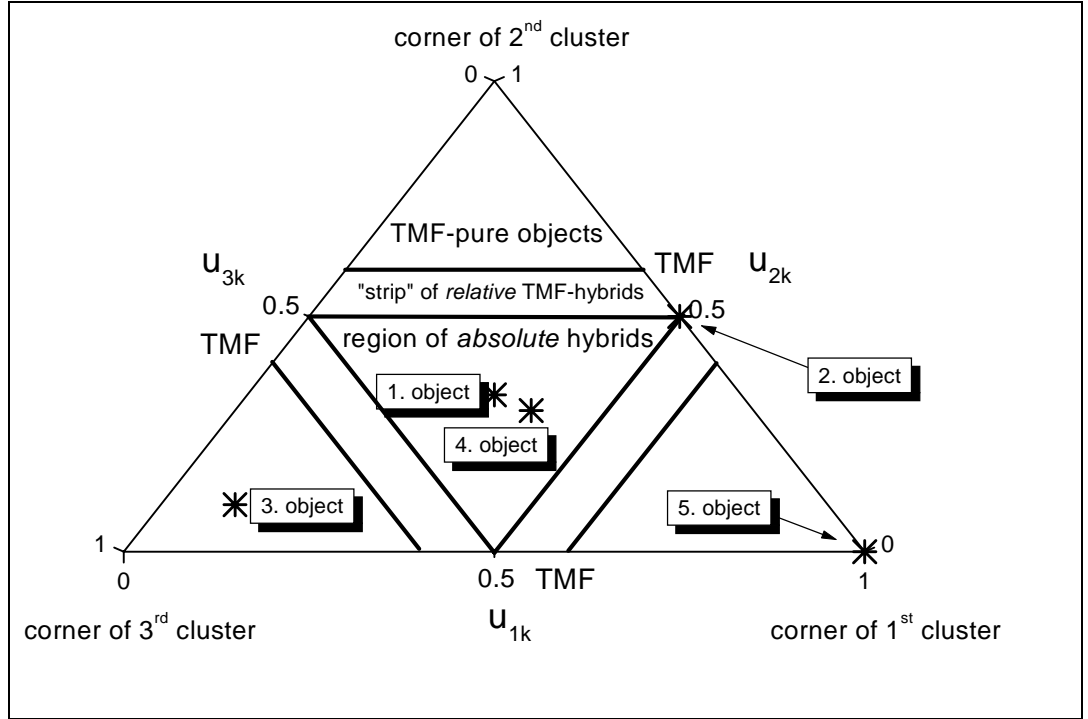


Figure 3: Good and bad regions on the "simplex of memberships"

3.3 Cluster Analysis II: Usual Assessment of Unsharp Partitions

We assume, as we said in section 0 that we've a matrix $Y \in \mathbb{R}^{N \times |B|}$ which row vectors represent the objects in the euclidian space, i.e. the dissimilarity between two objects can be expressed with the euclidian distance . So, if we define $Z \in \mathbb{R}^{FCL \times |B|}$ according to

$$\bullet \quad Z_{i.} := \frac{\sum_{k=1}^N u_{ik}^r \cdot y_k}{\sum_{k=1}^N u_{ik}^r} \quad (\forall i \in \{1, \dots, FCL\})$$

Center or r-weighted average of cluster i

where z_i , the i-th row of Z denotes the **center** or **r-weighted average** of cluster i and

$$\bullet \quad g(U) := \sum_{i=1}^{FCL} \sum_{k=1}^N u_{ik}^r \cdot \|y_k - z_i\|^2 \text{ as U's r-weighted sum of error squares}$$

we have a suitable measure for the quality of an unsharp partition refer to the dissimilarities between the objects. In case U is sharp, the formulas above turn into

$$\bullet \quad z_{i\cdot} := \frac{\sum_{o^{(k)} \in A_i} y_k}{|A_i|} \quad (\forall i \in \{1, \dots, FCL\})$$

and

$$\bullet \quad g(U) := \sum_{i=1}^{FCL} \sum_{o^{(k)} \in A_i} \|y_k - z_{i\cdot}\|^2$$

for the corresponding partition – this may show the intention "clearer". In *Luther et al. 1999* the quality functions are discussed in more detail. These will be a differentiation of the function g which makes evident the contribution of singletons versus those of nontrivial clusters.

4 Generalization: Three Strategies For Combining Order Theory And Cluster Analysis

We have done a generalization by replacing posets with presets (weakening the antisymmetry) in order to model equivalences of different objects. But a partition can be seen as equivalence relation too – we can consider the clusters as equivalence classes of the object set. And the reason why we have introduced preorders was really some kind of "lazy men's cluster analysis" for we have considered objects with the same datatuple (or however the same value of a weighted sum) as equivalent.

Three alternatives

Hence it is already clear, that the *results* of a clustered preset are expressed by a preset again, but *how* to do this is *not* clear. There are **three alternatives**.

4.1 Clustering → Preset

We can derive a partition of the object set in a 1st step, then we have to gain representatives of every cluster and construct the preorder for the representatives.

- Advantage: Since our objects are represented by vectors in $\mathbb{R}^{|\mathbb{B}|}$ we can easily derive representatives even for unsharp partitions as showed in 0. Cluster analysis is a well explored field, so there are many tools we can use to evaluate our results.
- Disadvantage: The clustering may destroy ordinal relations and generate new ones, so our resulting preset strongly depends on the cluster-

Clustering may destroy ordinal relations

ing which is much more arbitrary than (for example) the product order, because for example we either have to decide which partition we want to select (sharp partition) or we have to select appropriate values for FCL and TMF (rounding an unsharp partition). But it was a basic goal to avoid arbitrary!

4.2 Preset → Clustering

We can construct a preset and simplify it by clustering of a part of an antichain.

- Advantage: The restriction to antichains (or in general a restriction to a subset of objects) can help to reduce the distortion of the preset.
- Disadvantage: The choice of the antichain and the clustered objects may still be arbitrary.

4.3 Simultaneous method

We can derive a preorder directly from the datamatrix X under "suitable assumptions/ restrictions".

- Advantage: No more arbitrary, united model of structure and quality
- Disadvantage: We still know neither suitable assumptions/restrictions nor how to derive a good preset under *both* points of view.

There still is no "eternal" way how to combine order-theory and cluster analysis, (this would be an answer to "4.3"), but we will show some actual results that may illustrate the first two strategies and their meaning – they seem to be good heuristics.

5 Approach: Assessment of 59 Regions corresponding to their Pb-, Cd-, Zn- and S-Pollution

Pollution of several regions

An evaluation project performed for an environmental protection agency (LfU Baden-Württemberg) had to deal with the pollution of several regions. The task was to evaluate 59 regions according to the content of lead, cadmium, zinc and sulfur in different matrices. The matrices were: herb layer, the leaf layer (leaves of trees), moss and earth worms. Each of these matrices were considered to be indicators for different kinds of pollution patterns. Here the leaf layer will serve as an example.

5.1 Clustering → Preset

The simple product-order gives the "messy" system of lines drawn in Fig. 4.

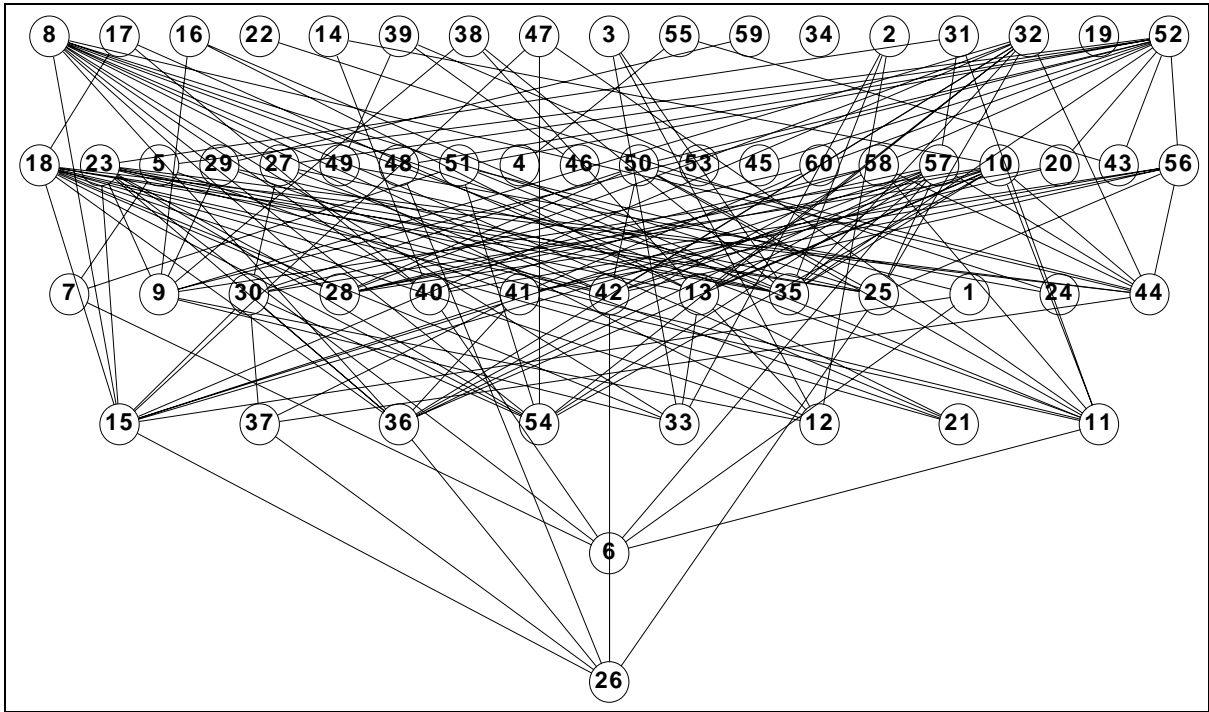


Figure 4: Hassediagram using original data.

Going the way "Clustering → Preset" we generate an unsharp partition with 6 clusters (Fig. 5).

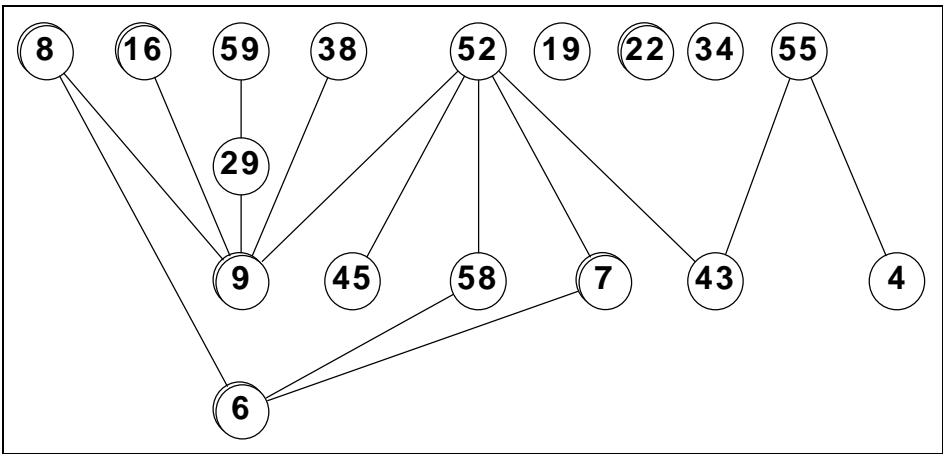


Figure 5: Hassediagram with clusters and hybrids.

0.6-hybrids (11)	{4, 19, 29, 34, 38, 43, 45, 52, 55, 58, 59}
nontrivial equivalent classes	{6, 15, 28, 40, 42, 37, 26, 36, 54, 25, 11, 44};
	{8, 17, 18, 23, 14, 57, 10, 31, 32};
(6 clusters, 49 assigned objects):	{7, 30, 5, 41, 27, 47, 48, 51, 50, 60, 1, 20, 56};
	{9, 13, 35, 33, 12, 21}; {16, 3, 2};
	{22, 39, 49, 46, 53, 24}

5.2 Preset \rightarrow Clustering

Using the "stronger" product order given by $D(x) = \{x\} + \Gamma_d$, we yield Fig. 6.

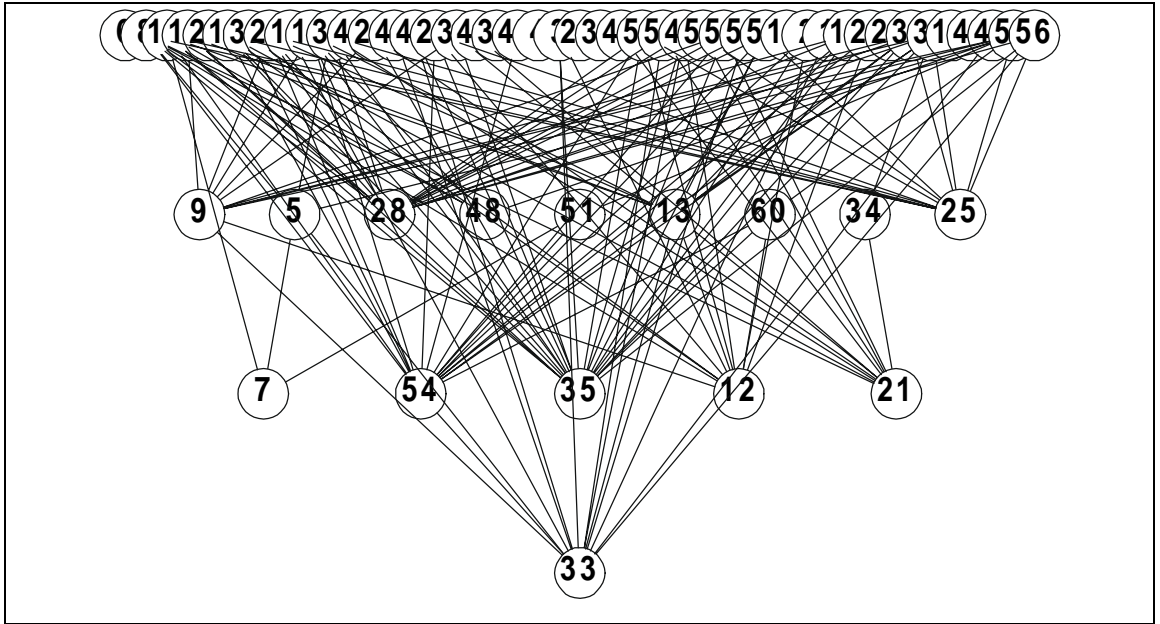


Figure 6: Hasse diagram derived of $D(x) = \{x\} + \Gamma_d$

We then cluster only the maximal elements (in this case we made a complete linkage clustering), and arrive at Fig. 7.

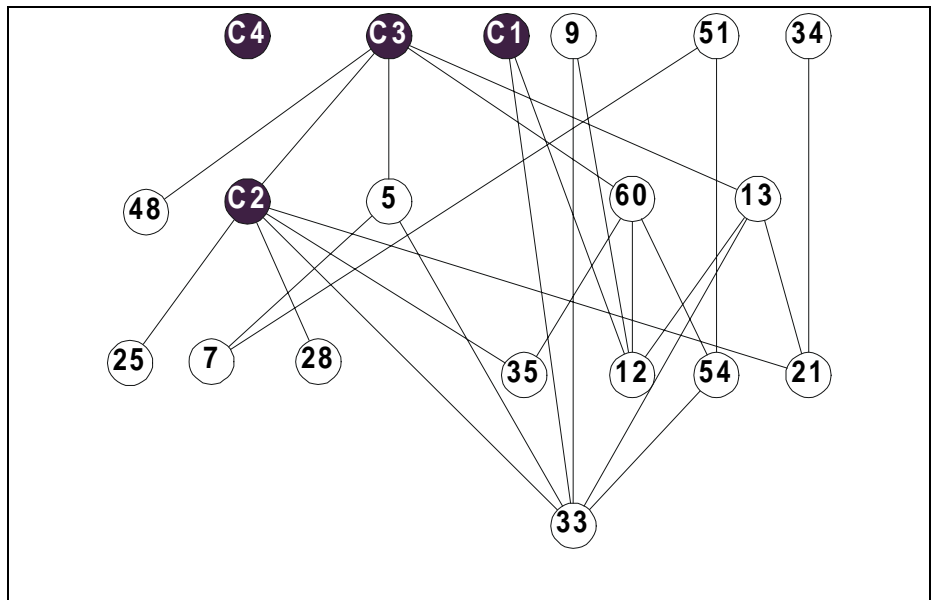


Figure 7: Hasse diagram of Fig. 6 after a clustering of the maximal elements (the clusters refer to the four black elements)

No preorder relation of objects is destroyed

Here it was a good idea first to *improve* the partial order (way $0 \rightarrow 2 \rightarrow 3$ on the schematically illustration in Fig. 8), so we could cluster only the maximal elements and stay order preserving in the sense, that no preorder relation of objects is destroyed.

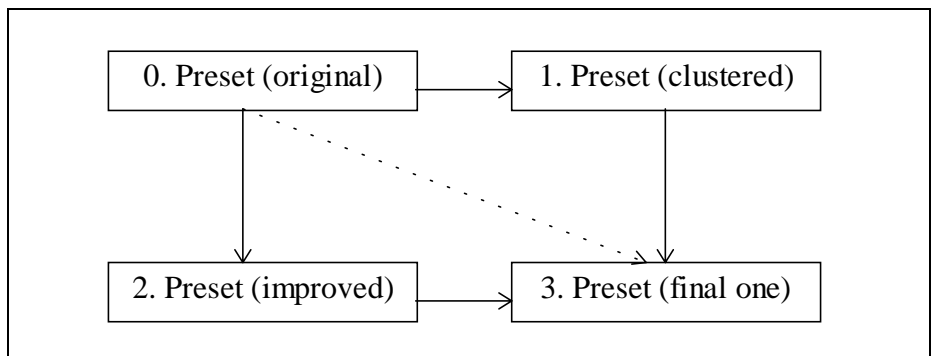


Figure 8: Schematically illustration of clustering and order improvement

6 Acknowledgement

We thank the Environmental Protection Agency of Baden-Württemberg for giving us access to their data.

7 References

Brüggemann, R. & Bartel, H.-G. (1999): A Theoretical Concept to Rank Environmentally Significant Chemicals. *J.Chem.Inf.Comp.Sc.* 39(2): 211-217.

Brüggemann, R., Bücherl, C., Pudenz, S., & Steinberg, C. (1999): Application of the concept of Partial Order on Comparative Evaluation of Environmental Chemicals. *Acta hydrochim.hydrobiol.* 27(3): 170-178.

Bock, H.H. (1974): *Automatische Klassifikation*, Göttingen: Vandenhoeck & Ruprecht, pp. 6-480.

Bock, H.H. *Klassifikation und Erkenntnis III* (1979): Numerische Klassifikation Vorträge und Diskussionen zur Numerischen Klassifikation bei der 3. Fachtagung der Gesellschaft für Klassifikation e.V. , Königsstein /Ts., 4.-6.4.1979, Frankfurt: Gesellschaft für Klassifikation, pp. 1-174.

Bock, H.H. (1979): Clusteranalyse mit unscharfen Partitionen. In: *Studien zur Klassifikation, Bd 6 Klassifikation und Erkenntnis III - Numerische Klassifikation*, edited by Bock, H.H.Frankfurt:Gesellschaft für Klassifikation, pp. 137-163.

Luther, B., Brüggemann, R*, & Pudenz, S. (1999): An Approach to Combine Cluster Analysis with Order Theoretical Tools in Problems of Environmental Pollution *Communications in Mathematical Chemistry*. (in press).

P.B. Sørensen, R. Brüggemann, B.B. Mogensen, L. Carlsen & S. Pudenz: Compound data analysis of monitoring data of pesticides in surface waters. Data interpretation and model development using partial order ranking theory (in preparation).

Yu, P.L. (1973): Introduction to Domination Structure in Multicriteria Decision Problems. *Multiple Criterion Decision Making* 1973:249-261.

New Tools in Hasse Diagram Technique

Example: Comparative Evaluation of Near-Shore Sediments by a Battery of Tests ¹

R. Brüggemann^{a)}, E. Halfon^{b)}, B. Luther^{a)}, S. Pudenz^{a)}, U. Simon^{a)},

^{a)} *Institute of Freshwater Ecology and Inland Fisheries
Müggelseedamm 310
D-12587 Berlin
Germany*

^{b)} *McMaster University
Hamilton, Ontario
Canada*

Abstract

When a ranking evaluation of some objects (geographical sites, chemicals, management strategies) by a multicriteria analysis is of concern, then the theory of partially ordered sets and its graphical representation (Hasse diagrams) are useful tools. From the point of application three new tools are discussed: (1) How can we get systematically information, obtained from the structure of the Hasse diagram: (2) How can we quantify the loss of information, when an aggregation of test results (for example by forming a weighted sum) is used. (3) Can we find a new set of attributes (latent variables) which generates the same (an isomorphic) Hasse diagram as that from the original attributes. This is done with the help of the concept of dimensions.

The three new tools are explained with the example of a test battery, consisting of five single tests (*developed by Dutka et al. 1986*) and applied for an evaluation of 55 samples of the Lake Ontario.

¹ Some parts of this manuscript, together with the theory of sensitivity analysis will also be published elsewhere.

1 Introduction

Hasse diagrams have been used to rank chemicals according to environmental hazard (see *Brüggemann (1998)* for a recent review). At the basis of the Hasse diagram technique (HDT) is the assumption (*Halfon and Reggiani, 1986*) that we can perform a ranking while avoiding the use of an ordering index. Mathematical background material can be found in *Bartel (1996)*, *Davey and Priestley (1990)* or *Lipschutz (1976)*. Hasse diagrams present information not only on the ranking but, most important, they show whether the criteria, characterizing the objects, lead to ambiguities in the ranking. For example, an object might be ranked higher according to one criterion but lower according to another. These two objects are not ordered because their data are "contradictory". This ambiguity is not evident when we use an index for ranking, i.e. if we aggregate the results of the battery of tests² to only one quantity (an index function). However, the ambiguity is immediately evident by the presence or absence of lines in a Hasse diagram. Therefore, Hasse diagrams and the techniques to draw information from them are very appropriate for applications on complex systems such as ecotoxicological batteries of tests. The reason is that batteries of tests deliver much simultaneous information which can be used to evaluate the tested objects. The HDT uses all ordinal information for a graphical representation of the test results. Other investigators have approached the problem of finding significance values (see *Sørensen et al., 1998a and 1998b*), to develop systematic instruments for data analysis (see *Brüggemann and Bartel, 1999 and Welzl et al., 1998*) and to find methods for relating attributes to latent order variables (*Grell and Brüggemann, 1998*). In this paper, the ecotoxicological test battery and its results will serve as an example to show some new tools of the Hasse diagram technique (HDT). These are:

- (1) Sets of samples, with specific test reactions, i.e. having a pattern of test results in common. We call this kind of results from Hasse diagrams "structural information".
- (2) The loss of information, which appears, when an aggregation of test results is used,
- (3) the dimension of Hasse diagrams with respect to redundancy in the battery of tests.

Some basics

Although some basics are already mentioned in the Introduction (see this Proceeding), it may be useful to repeat and to extend them.

Criteria

Criteria include both quantitative and qualitative properties. Often in environmental sciences there are protection goals, to which some criteria correspond. However, not all these criteria have a measurable quantity as counterpart. Therefore, subcriteria are to be defined, until quantities are available (by experiments, by modelling etc.) which match the (sub)criterion. Often a hierarchy of criteria arise, an example is shown elsewhere (*Brüggemann et al., 1999a*). Note that subcriteria on the same

² Sometimes we also write "test battery" or simply "battery"

hierarchical level may have some conflict potential, but all of them, joining to a common super-criterion are necessary to take the full account of the meaning of that super-criterion.

Attribute

An attribute is a numerical quantity logically related to a criterion, for example by a hierarchy mentioned above. We denote these attributes as q_1, q_2, \dots, q_m ; they are assigned to the objects, which are to be evaluated and they have to have the same orientation: If for example a high bioaccumulation is considered as hazardous, then any species toxicity, measured as LC-value should be multiplied with -1 or the reciprocal should be chosen, just to ensure that high numerical values have the same meaning. Properties, which are oriented and which are used in the ranking procedure are also called evaluative properties.

Data

Data are the numerical values corresponding to each criterion by which a given object is characterized.

Object

An object is the item of interest. Each object, x , is characterized by a tuple of data ($q(x) = (q_1(x), q_2(x), \dots, q_m(x))$). The set of n objects is called E . Objects are ranked graphically by Hasse diagrams, applying an order relation (see below). The operator "card" acts on finite sets and the result is the number of elements of sets. For example: When the object set E contains n objects, then $\text{card } E = n$.

Order relation

Order relation: Two objects $x, y \in E$ are characterized by the tuples $(q_1(x), q_2(x), \dots, q_m(x))$ and $(q_1(y), q_2(y), \dots, q_m(y))$. We say: x and y are comparable, if $q_i(x) \leq q_i(y)$ or $q_i(y) \leq q_i(x)$, **for all** $i \in I = \{1, 2, \dots, m\}$ (I : an index set). If $q_i(x) < q_i(y)$ for **all** i , or if for $i \in I' \subseteq I$ $q_i(x) = q_i(y)$ and for the residual index set $i \in I'' \neq \emptyset$ and $I' \oplus I'' = I$, $q_i(x) < q_i(y)$, then: $q_i(x) \leq q_i(y)$ **for all** $i \in I$ and $x \leq y$. If $q_i(x) \leq q_i(y)$ **not** for all i , i.e. if there exists at least one i^* with $q_{i^*}(x) > q_{i^*}(y)$ and one i^{**} with $q_{i^{**}}(x) < q_{i^{**}}(y)$, then the two objects x, y are incomparable (with respect to the considered set of attributes). In that case we write: $x \parallel y$. The demand "for all" to set up an order relation we call the generality principle³.

Equivalence relation

Equivalence relation: If two different objects $x, y \in E$ have the same tuple, i.e. if $q(x) = q(y)$ then these two objects are considered as equivalent. The equality can be seen as another example of binary relation among objects. This relation is called an equivalence relation and will be denoted by \mathfrak{R} . An equivalence relation leads to equivalence classes of equivalent objects. The set of equivalence classes which arise from a given \mathfrak{R} is called a quotient set and is denoted by E/\mathfrak{R} . (More details and examples: see "Introduction to the General Principles of the Partial Order Ranking Theory", this issue and Brüggemann et al., 1997.)

Aggregation

An aggregation is a method to assign to a vectorial quantity a scalar: The tuple $(q_1, q_2, \dots, q_m) \in \mathbb{R}^m$ (m -dimensional space of attributes) is mapped onto a scalar $\Gamma \in \mathbb{R}$ (a one dimensional space). In evaluation studies this map often is realized by weighted sums, i.e.

$$\Gamma = \sum g_i \cdot q_i \quad (i=1, \dots, m), \quad g_i \geq 0 : \text{weights.} \quad (1a)$$

³ In economic sciences also known as Pareto rule.

Generally an aggregation may be formulated as a monotonous differentiable function f of q_1, q_2, \dots, q_m :

$$\Gamma = f(q_1, q_2, \dots, q_m); \partial f / \partial q_i \neq 0 \text{ for all values of } q_i \text{ and } i=1, \dots, m \quad (1b)$$

(see Sørensen et al., 1999).

Sometimes bilinear aggregation models are used, which can be formulated as follows:

$$\Gamma = (\sum g_i q_i) \cdot (\sum g_j q_j) \quad (2)$$

Note, that other aggregation tools are possible, which are not based on numerical transformations like those shown in Eq.1 (Kemeny, 1959; Roberts, 1989). A rather important method may be the tournament theory, based on directed acyclic graphs. It can be shown that under some assumptions a unique total order can be derived (see Clark et al., 1994, Barthelemy, 1989, Bondy and Murty, 1976). A first attempt is shown by the contribution of U. Simon, this issue.

However the aggregation by Eq.1 is mostly used in risk assessment procedures (as multi-attributive utility theory (MAUT), Schneeweiß, C., 1991), therefore we concentrate ourselves on that point.

Product order

There are many other possibilities to define order relations (see "Introduction"), the specific one, shown above has an own name, namely product order or (syn.) componentwise order. Sets, equipped with an order relation are called partially ordered sets (posets). A total order is a set, whose order relation leads to complete comparability, i.e. each object is comparable with each other. Not all objects are mutually comparable in posets, because of the generality principle.

Hasse diagrams

The so called Hasse diagram technique is explained in several publications (see for example Halfon et al., 1986, Brüggemann and Halfon, 1997). Especially Klein et al. (1995) demonstrated how Hasse diagrams are constructed. For the sake of convenience of the reader and because this diagram will serve for further examples, we briefly repeat some facts. The Hasse diagram (Figure 1) is part of the comparative evaluation of the results of a battery of tests, which is fully discussed later.

In Table 1 some objects and five attributes are presented and Figure 1 shows the visualization of the poset after demonstrating the generality principle:

Table 1: Data for an example (details of the test battery and its attributes are explained later)

<i>objects\attributes</i>	q_1	q_2	q_3	q_4	q_5^4
a (sample 1)	2	0	0	4	0
b (sample 17)	3	0	0	6	0
c (sample 7)	2	0	0	8	0
d (sample 5)	3	3	2	0	0
e (sample 95)	3	5	2	6	0

Obviously: $a \leq b$, because $q_1(a) < q_1(b)$ and $q_4(a) < q_4(b)$, $q_2(a) = q_2(b) = q_5(a)$. However, $a \parallel d$. With respect to (q_1, q_2, q_3) $a < d$; $q_5(a) = q_5(b)$, the incomparability arises from q_4 : because $q_4(d) < q_4(a)$. Therefore the relation $q_i(a) \leq q_i(d)$ does **not** hold for all attributes; the generality principle is not fulfilled. The graphical presentation by an Hasse diagram is shown in Figure 1:

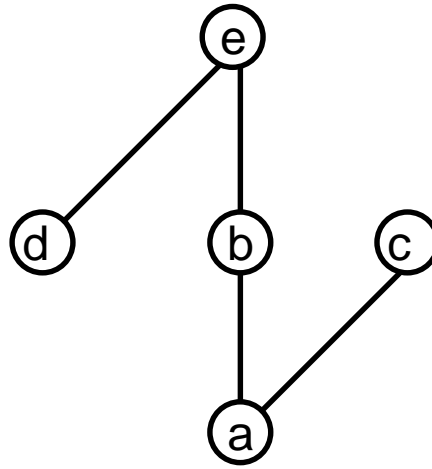


Figure 1: Hasse diagram based on the data of Table 1. This graphic serves as example. Maximal elements are e and c. Minimal elements are d and a. There is neither a greatest nor a least element.

Key elements and successor sets

Substructures within a Hasse diagram, i.e., relations among objects as well as the importance of criteria in ranking are investigated with the help of key elements and successor sets. Any object of the poset (E, IB) can be chosen as a starting point to begin the analysis⁵. We call this object, a "key element". For convenience, all chosen key elements form a set K , a subset of E .

Analysis of the successors of a key element implies a search of all objects located lower than, or equivalent to, that of the key element and connected to it by a path, being a sequence of connecting lines. The set of all successors of key element k is denoted as $G(k)$ ⁶.

⁴ Note that real data are used, by chance all values of q_5 were zero for the objects under consideration.

⁵ Note that for this kind of analysis we refer to E not to E/\mathcal{R} . The reason is that we aim to study the effect of different attributes. When different cases are examined, then different quotient sets would arise. We want to avoid cumbersome notations.

⁶ Note the similar concept of "down-sets" and order ideals generated by some elements in Davey and Priestley (1990). We also write " $G(k)$ is generated by object k ." $G(k) \cup \{k\}$ is a principal order ideal.

Relations between elements of posets

A new tool to analyze relations among elements of posets is that of "local" contradictions of an object $x \in E/\mathfrak{R}$: $U(x)$. That is the number of objects $y \in E/\mathfrak{R}$, $y \neq x$, $y \parallel x$:

$$E'(x) = \{y \in E/\mathfrak{R}: y \neq x, y \parallel x\} \quad (3a)$$

$$U(x) = [\text{card } E'(x)] \quad (3b)$$

Note that $U(x)$ has an upper bound, namely $\text{card } E - 1$, therefore the normalized quantity $u(x) = U(x)/(\text{card } E - 1)$ may be useful too.

Here a remark may be necessary. In the "Introduction" a quantity U (without parentheses and an object identifier) was mentioned. U is the count of all incomparabilities within a poset i.e., $U = \sum U(x) \ x \in E$

Therefore U is two times the count of each pair $x \parallel y \quad x, y \in E$.

If similarly all comparabilities are counted: $V(x) = \{(x, y) \in (E/\mathfrak{R}) \times (E/\mathfrak{R}), x \neq y, x \leq y\}$ then clearly

$$U(x) + V(x) = \text{card } (E/\mathfrak{R}) - 1 \quad (4)$$

Contradictions are crucial for each aggregation procedure which maps several test results into an index. This fact will be discussed in greater detail in the Results section.

By the estimation (with the assumption of uncorrelated attributes), shown in *Sørensen et al. (1998b)* $\langle V(x) \rangle = (\text{card } E/\mathfrak{R} - 1) * 0.5^m$

(Brackets like $\langle \rangle$ denote average values for large data sets) we arrive at:

$$\langle U(x) \rangle = (Z - 1) * (1 - 0.5^m) \quad (5)$$

with $Z = \text{card } E/\mathfrak{R}$.

The quantity $u(x) := U(x)/\langle U(x) \rangle$ helps to evaluate $U(x)$ and -as will be shown later- to assess the uncertainty in ranking, if an aggregation procedure is applied. A schematical representation may be useful:

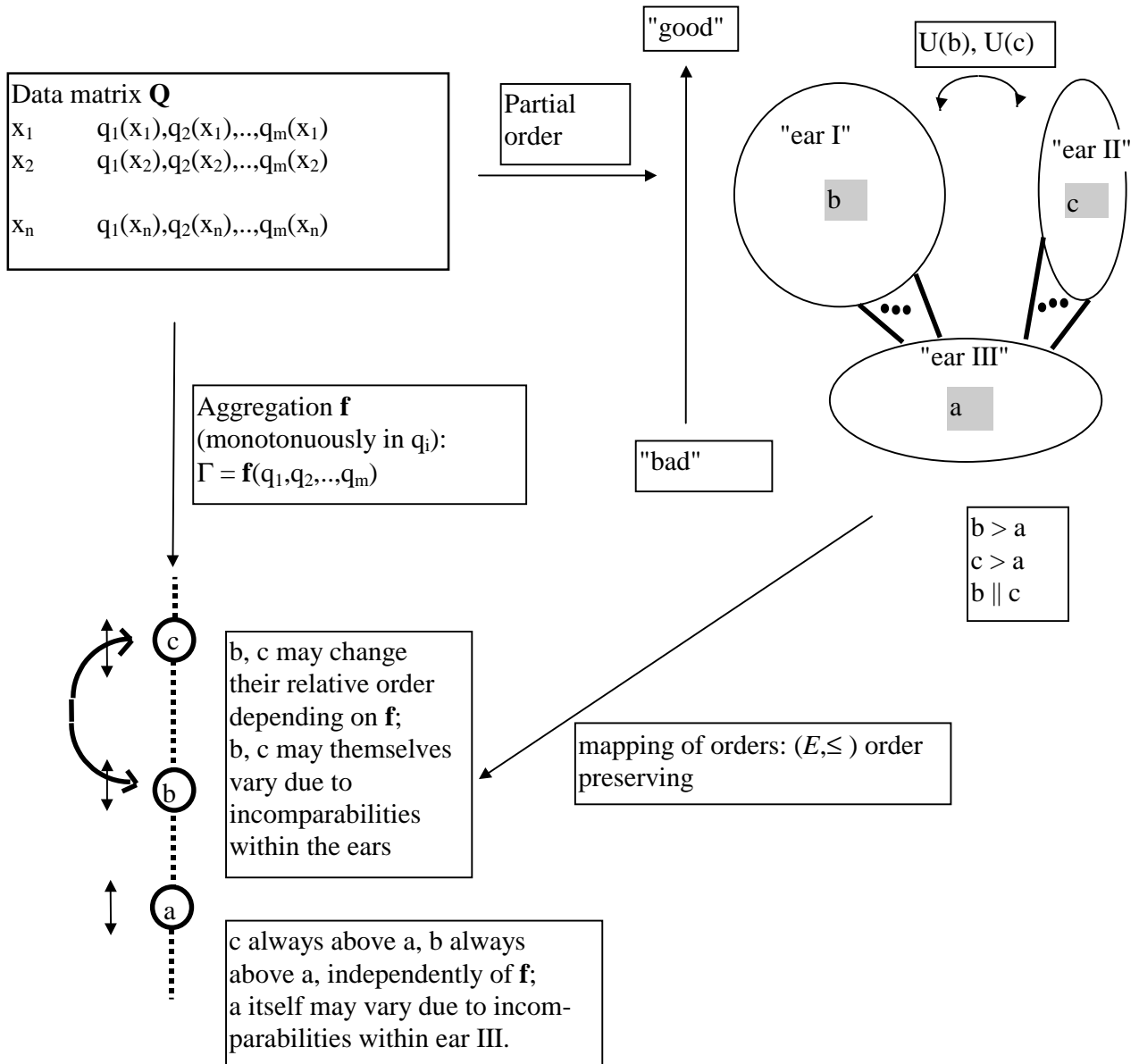


Figure 2: Variabilities after aggregation procedures due to incomparabilities. The dotted line representing the total order generated by Γ should demonstrate that there are other objects above c and below a and between the three objects a, b and c .

2 Results

Structural information (task 1)

Some successor sets of maximal elements are of specific interest: If maximal elements are selected, which have as test results the lowest possible numerical value, then **all** successors must also have this value, according to the order relation. We formalize this observation: Let Z_k be the set of attributes of object k exhibiting the lowest possible numerical value:

$$Z_k := \{ q_i \in IB: q_i(k) = \min \{ q_i(k') \mid k' \in E \} \} \subseteq IB \quad (6)$$

If -for example- zero is the lowest possible value as it is the case in the example of the battery of tests, then Z_k consists of those attributes $q_i(k)$ of object k which have the value 0. In the application part we will use the concept of templates, i.e. a tuple is written as a combination of the signs "*" and "0". The sign "*" indicates values greater or equal the lowest possible value, whereas "0" indicates that the lowest value is actually present. Then:

$$\text{For all } x \in G(k) \Rightarrow Z_k \subseteq Z_x \quad (7)$$

that means that the lowest values (gathered in Z_k) can be found for all objects of $G(k)$, starting from object k and proceeding downwards in the Hasse diagram and

$$\text{for all } x \in (G(k) \cap G(k')) \Rightarrow Z_k \cup Z_{k'} \subseteq Z_x \quad (\text{generality principle}) \quad (8)$$

that means that all zeros of k and k' resp. can be found for any object $x \in (G(k) \cap G(k'))$.

The evaluation of Eqs. (6)-(8) is very simple, if q_i are discrete variables. In some cases a Boolean arithmetic can be established (*Brüggemann et al., 1999b*).

Note that the numerical values of "0's" at different positions in the template may differ. Equations (6)-(8) can be somewhat simpler formulated as:

For all $x \in G(k)$ the number of "0's" of their templates can only increase proceeding downwards within $G(k)$ and for all $x \in (G(k) \cap G(k'))$ all "0" of k and k' resp. are present in the template of x

With the help of the concepts of templates a partitioning of the set E can be performed. Begin with the maximal elements, k . Select that key element which has the largest number of "0's" in its template and gather all successors. Then the key element together with its successors form the first subset. Continue this procedure till the set E is exhausted. Some care is needed, when there are several key elements with the same number of "0's". If the template is the same, then unify the subsets; if this is not the case then an arbitrary assignment is only necessary for representatives, which are common to both successor sets. This fact should be documented.

Aggregation procedure (task 2)

(2a) The role of a contradictory pair:

Let be q_1, q_2, \dots, q_m the m results of the battery containing m tests. Often an overall aggregation is performed by calculating a weighted sum (Eq. 1). If there is a contradiction between two tested objects, for example two sediment sites a and b , then $\Gamma(a) > \Gamma(b)$ or $\Gamma(b) \geq \Gamma(a)$ depending on the weights, whereas for comparable pairs of sites from $a \geq b$ always $\Gamma(a) \geq \Gamma(b)$ follows. The advantage of HDT is based on the fact that there is no need to find an aggregation procedure to perform a ranking.

(2b) Expectation values for rankings in dependence of local incomparabilities:

Objects with large local incomparability values are very sensitive with respect to the selection of weights in forming an ordering index like equation (1a) or with respect to the particular form of the function f (1b).

Let Γ be ordered like $\Gamma(a) < \Gamma(b) < \dots < \Gamma(x)$, then we say object a has (with respect to Γ) the rank 1, object b the rank 2 etc. The rank of an object x is denoted as $Rk(x)$.

Consider different weights, g_i , for example motivated by different environmental protection goals, then the ranks (of a, b, c given by Γ) changes in dependence of g_i . An object x may get its maximum rank Rk_{\max} for a specific selection of g_i -values and its minimum rank Rk_{\min} for another one.

Firstly we define:

$$\text{var}(x) := Rk_{\max}(x) - Rk_{\min}(x) \quad (9)$$

Each ranking index via Eq's. 1 defines a total order. The relation between the poset and the total order is described as an order preserving map: Each comparability $x \leq y, x, y \in E/\mathfrak{R}$ of $(E/\mathfrak{R}, IB)$ will be reproduced in the total order generated by the weighted sum Eq. 1a or functions of type Eq. 1b. Therefore it suffices to look for the set of all total orders, which can be obtained from the poset by order preserving maps. We call this set *LE* (after linear extensions, see Trotter, 1991). For example, the linear extensions of the Hasse diagram shown in Figure 1 are shown in Table 2: Within *LE* it can be shown that

$$\text{var}(x) = U(x) \quad (10)$$

holds. The set of total orders obtainable by Eq. 1a is a subset of *LE*, therefore $\text{var}(x)$ must be equal or less than $U(x)$, i.e. the inequality (11) holds.

$$\text{var}(x) \leq U(x) \quad (11)$$

The inequality (11) and the Eq. 10 demonstrates the loss of qualitative insights into the set of objects, if -for example- a battery of tests is used, but an ordering index like that calculated by Eq. 1 is applied. With the help of the quantity $u(x)$ a decision can be made whether an object is unusually (in comparison with $\langle U(x) \rangle$) variable if aggregation procedures are applied.

This will be the case, when $u(x) > 1$. A study about rankings in dependence of different aggregation functions was recently published by *Sørensen et. al., 1999*

If functions of the general form (1b) are allowed, then equality holds, Eq. 10. However for linear functions (1a) there can be found counterexamples, thus only \leq holds.

A systematic way to generate linear orders "by hand" is shown in the publication of *Atkinson, 1989*. From the computational point of view, *Trotter, 1991* published an algorithm.

Table 2: Linear extensions of the Hasse diagram shown in Figure 1

No1	No2	No3	No4	No5	No6	No7	No8	No9	No10	No11
e	e	e	e	e	e	e	e	c	c	c
d	b	b	b	d	c	c	c	e	e	e
b	d	c	c	c	d	b	b	d	b	b
c	c	d	a	b	b	d	a	b	d	a
a	a	a	d	a	a	a	d	a	a	d

It can be easily seen that each linear extension reproduces the order relations, shown in Figure 1. The manifoldness of linear extensions arises from the incomparabilities. For example object d must be below object e, due to the order preserving map, but the ranks of incomparable pairs can be mutually reversed. The equality (Eq. 10) can be demonstrated too:

$U(b)=2$, $U(c) = U(d)=3$. For example: $Rk_{\max}(c) = 4$, $Rk_{\min}(c)=1$, thus $var(c)=3$. Indeed $var(b)=2$ and $var(c)=var(d)=3$ as it must be.

The set of linear extensions can be seen as probability space (*Winkler, 1982, Graham, 1982*). Thus, given a function of the type of Eq. 1b) the probability can be calculated that -for example- $d > b$: There are four linear extensions out of the total of 11, therefore the probability is $4/11^7$. I.e. the probability that by any function (1b) the sample 5 is ranked higher than sample 17 is 0.36; the probability that sample 5 is ranked higher than sample 7, i.e. that $d > c$ is $3/11$. Both pairs of samples are incomparable, but it is more unlikely to find an aggregation function (Eq. 1b) which ranks d higher than c.

A series of important conclusions can be drawn from the fact that LE is a probability space, but this is not on the focus of this paper.

Dimension analysis (task 3)

As we have seen, from posets $(E/\mathfrak{R}, IB)$ total orders can be derived, which are not in contradiction with the order relations of $(E/\mathfrak{R}, IB)$ (so called linear extensions of the poset $(E/\mathfrak{R}, IB)$). The construction of ordering indices, like Γ (Eq. 1) leads to one specific total order, which is the theoretical reason for the preference of such aggregation procedures.

Linear extensions

⁷ The calculation of the total number of linear extensions $e(E/\mathfrak{R}, IB)$ is difficult. If the set E can be partitioned into $E_1, E_2 \dots$ then Stanley, 1986 gives a formula: $Z_i = \text{card } E_i / \mathfrak{R}$ $e(E/\mathfrak{R}, IB) = B(Z_1, Z_2, \dots) * e(E_1/\mathfrak{R}, IB) * e(E_2/\mathfrak{R}, IB) * \dots$ with $B(Z_1, \dots) = (Z_1 + Z_2 + \dots)! / (Z_1! * Z_2! * \dots)$

Somewhat simplified (*details, see Trotter (1991)*) the dimension of a poset, $\dim(E/\mathcal{R}, IB)$ is the minimum number of total orders, which together represent the original poset. The wording "together represent" means, that

1. a subset LE' of LE is to be found
2. assign to each total order of LE' the set of ordered pairs
3. derive the intersection of all sets of ordered pairs of LE' ,
4. if this set is a representation of the original Hasse diagram then the task is finished to find a set of total orders representing the Hasse diagram. If a minimal set LE' is found, then this set is called a realizer.

An example might be helpful.

Consider the linear extensions no. 1 and no. 2:

The set of ordered pairs of no. 1 and no. 2, each of them will contain $5 \cdot 4/2$ elements:

No. 1.: ac, ab, ad, ae, cb, cd, ce, bd, be, de

No. 2.: ac, ab, ad, ae, cb, cd, ce, db, be, de

There is only an inversion of the ranking between d and b, thus the Hasse diagram is almost a linear (total) order (Figure 3):

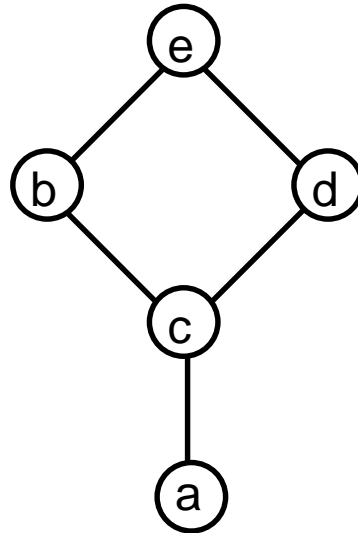


Figure 3: Hasse diagram, resulting from the intersection of the ordered pairs of no. 1 and no. 2.

Obviously, the Hasse diagram differs from that of Figure 1, thus $LE' = \{1. \text{ and } 2. \text{ linear extensions}\}$ is not a realizer.

If, however no. 1 and no. 11 (found after checking the whole list of linear extensions) are analyzed, then:

No. 1.: ac, ab, ad, ae, cb, cd, ce, bd, be, de

No. 11: ac, ab, da, ae, bc, dc, ec, db, be, de

The common pairs are: $\{ac, ab, ae, be, de\}$. If these pairs are graphically displayed, then a graph like that of Figure 1 is obtained. Therefore no. 1 and no. 11 form a realizer and we know that the dimension of the poset of Figure 1 is 2.

In general the recipe, given above is cumbersome and only computationally tractable. In the monography of *Trotter, 1991* there are many theorems reported, concerning the relation between the dimension and the set objects. Here, some examples of theorems found by Hiraguchi are shown, which have some practical importance:

Let (E, IB) be a poset with $\text{card } E = n$, then: $\dim(E, IB) \leq n/2$, if $n \geq 4$;
 $\dim(E, IB) \leq 2$, if $n \leq 5$; $\dim(E, IB) \leq 3$, if $n \leq 7$.

The three statements above may be useful to find an upper limit. A little bit more tricky theorem, again found by Hiraguchi is: Let (E, IB) be a poset and let $C \subseteq E$ be a chain with $E-C$ nonempty. Then: $\dim(E, IB) \leq 2 + \dim(E-C, IB)$.

Thus, just from one of the theorems of Hiraguchi we know -without the troublesome procedure of looking explicitly for linear extensions- that the dimension of the poset, shown in Figure 1 is 2. There is an important consequence: Often one would like to reduce the number of objects, for example by cluster analysis, then there is a "built-in" redundancy if $\text{card } IB$ is large enough.

Here another more restricted definition is given, which is more handsome and which helps to find latent variables: If a Hasse diagram (eventually fictitiously supplied by a greatest and least element) can be drawn in a plane without any crossing of the lines, then $\dim(E/\mathcal{R}, IB) \leq 2$ and the Hasse diagram can be embedded into a two-dimensional grid (*Trotter, 1991*). For example, the Hasse diagram of Figure 1 can be supplied by a greatest and a least element and can be drawn without crossing of lines. Therefore the order dimension of the poset, visualized in Figure 1, is 2.⁸

If, for example the information base IB contains five elements, as it is the case in the battery of tests example, but the dimension is 2, then there must be a considerable redundancy with respect to the comparative evaluation.

If the dimension of a poset is found ($d = \dim(E/\mathcal{R}, IB)$), then corresponding many new latent ordering variables $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_d$ can be used to form the same Hasse diagrams as found by the original attributes. From a practical point of view this is only of interest, if $\dim(E/\mathcal{R}, IB) < \text{card } IB$. Because then the same ranking must be possible by a lower number of latent ordering variables. However there is still an open point: The numerical relation between the original attributes and the latent ordering variables may be very difficult, and if found, then hardly to interpret (as it is often the case in principal component analysis too).

Example

Battery of tests and results

A battery of tests developed by *Dutka et al. (1986)* to test the sediments of near-shore sites of Lake Ontario (Canadian part) is used to exemplify the definitions and some results of HDT. Experimental details and an extensive explanation, how to use HDT can be found elsewhere (*Brüggemann and Halfon, 1997*). In Lake Ontario 55 sediment samples corresponding to 50 different sites (some samples: same site, but different depths) were tested,

⁸ Note that the reverse conclusion is not true. Crossing of lines does not necessarily imply that the dimension is greater 2.

many sites were selected in front of Port Harbour (no. 9 and no's 91-95), of Toronto (no's 26-32), of Hamilton (no. 43) and near the mouth of the Niagara River (no's 47-55). Thus, the object set E contains 55 samples. *Dutka et al.* classified their results and used discrete scores instead of the measured (raw) data. For our analysis we took their classification. Therefore instead of measured data, discrete values describe the degradation of near shore sediments. To clarify the use of scores, we use, instead of the symbol q_i for the i -th attribute, the symbol s_i for the score of the i th test of the battery. The following tests are combined to form a battery:

Fecal Coliforms „FC“. FC is an indicator designed to control the health state of the sediments

Coprostanol „CP“ and Cholesterol „CH“. Both are indicators of loadings by fecals.

Microtox tests „MT“ and Genotoxicity tests „GT“. They describe some kind of acute toxicity and the potential for cancerogenicity, resp.

Scoring of sedimenttest results

Dutka et al. created many equivalence classes by scoring their data. It is convenient to refer only to these classes by specifying a representative for each class, i.e. in fact besides the sensitivity study we apply the concept of quotient set. With the specific equivalence relation \mathfrak{R}_5 meaning equality in all **five** scores s_{FC} , s_{CP} , s_{CH} , s_{MT} and s_{GT} , the following sediment samples are equivalent: (Table 3)

Table 3: Equivalence classes and their battery of tests pattern. No. of sites in bold letters are later used as representatives for the whole equivalence class. Note that now numbers instead of literal strings are used as identifier of objects.

<i>Equivalence class</i>	<i>FC</i>	<i>CP</i>	<i>CH</i>	<i>MT</i>	<i>GT</i>	<i>remark</i>
{ 2 ,8}	1	0	0	2	0	-
{ 4 ,6,10,13,19,21,22,29,30,48,94}	3	0	0	0	0	-
{ 11 ,16,40,41,42,43,44,45}	1	0	0	0	0	the best class
{15, 92 }	3	0	0	4	0	
{ 17 ,35}	3	0	0	6	0	
{20,24,26,28,34,37,39,49,50,51, 91 ,93}	2	0	0	0	0	
{ 23 ,60}	1	0	0	0	4	
{ 27 ,33,46,47}	5	0	0	0	0	one of the priority classes

The quotient set is denoted as E/\mathfrak{R}_5

As mentioned above equivalence classes can be perform by a cluster analysis. A very successful cluster analysis of sediment samples, for example, is published by *Ahlf et al. (1998)*. A combination of their results with HDT might be very promising.

Partial ordering

The information base of the battery of tests is : $IB := \{s_{FC}, s_{CP}, s_{CH}, s_{MT}, s_{GT}\}$. The partial ordering of the samples arises from the product order as explained in former sections. The attributes are now discrete variables s_i , $i = FC, CP, CH, MT, GT$. The visualisation of the poset $(E/\mathfrak{R}_5, IB)$ as is usually generated by the program WHASSE is shown in Figure 4, whereas the same poset, visualized with help of dimension theory is shown in Figure 5.

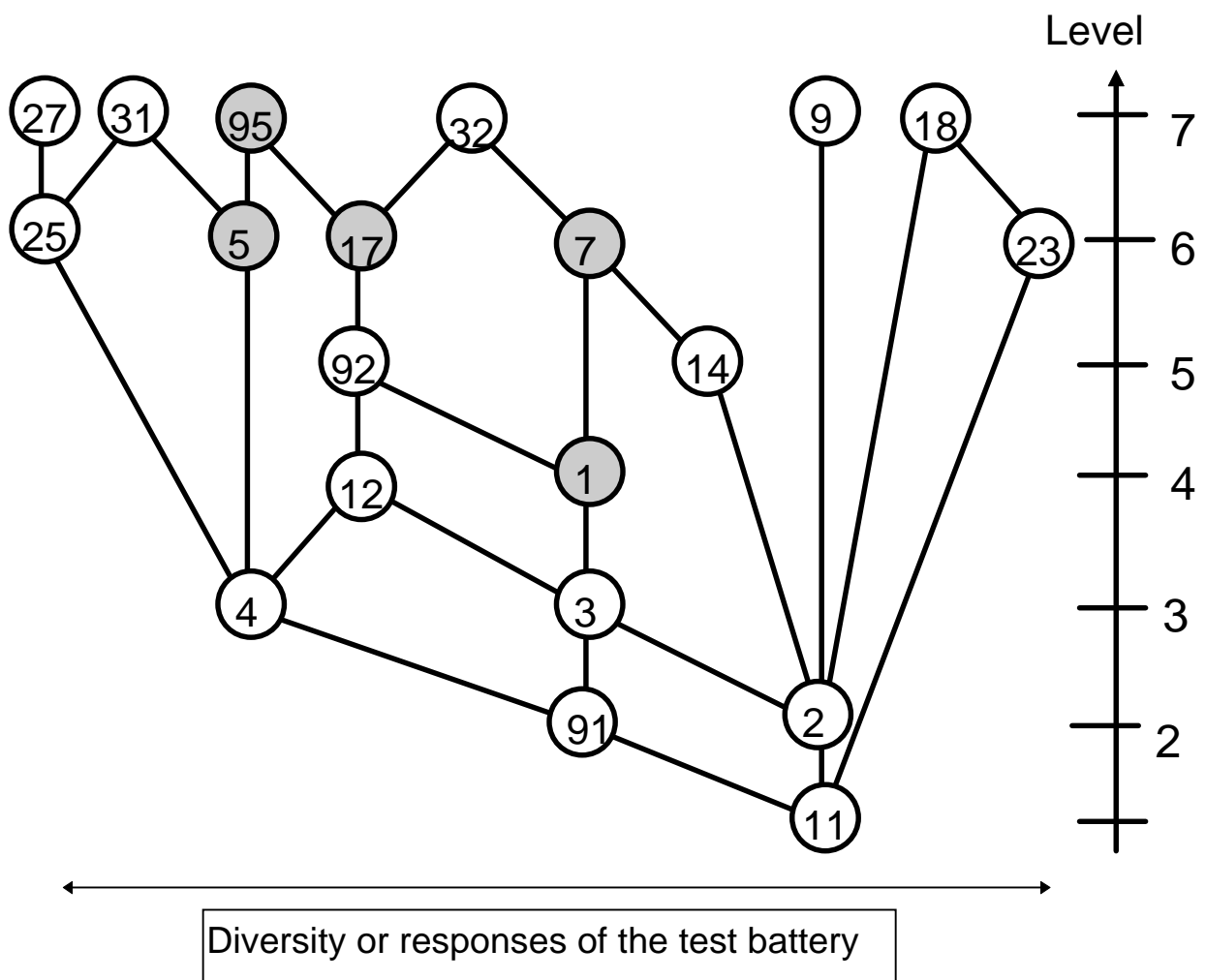


Figure 4: The comparative evaluation of samples of the Lake Ontario, as generated by the EDP-program WHASSE. Note, that the quotient set is partially ordered, i.e. that only the representatives (i.e. objects belonging to an equivalence class and representing all others of that class) are shown. Hatched circles correspond to the subset, shown in Figure 1.

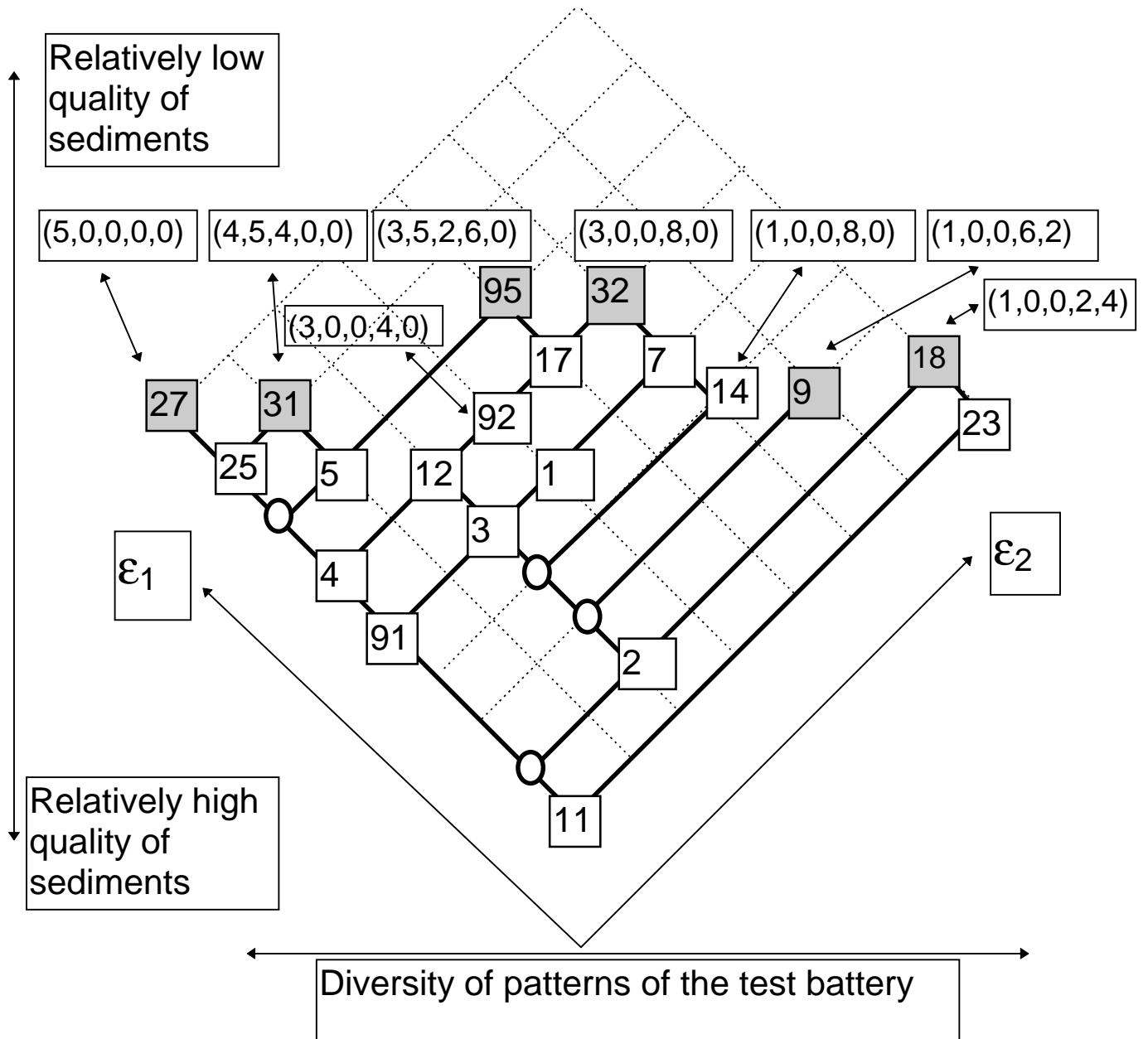


Figure 5: Hasse diagram of equivalence classes (see text) based on the scores of a battery of tests. The maximal elements are hatched {27,31,95,32,9,18}. There is only one minimal element (representative 11); it is therefore the least element of the poset $(E/\mathcal{R}, IB)$.

Compared to other graphical representations of Hasse diagrams (as for example in Figure 4), new graphical elements are used here:

- Rectangles to represent equivalence classes (some may only consist of only one object)
- Empty circles represent bifurcations points within the regular grid. For example, object 2 is comparable with 9. Instead of drawing a line directly, we follow the edges of the grid, introduce an empty circle and proceed upwards to object 9.
- Bold lines denote comparability. They are introduced in order to facilitate the comparison with the usually drawn Hasse diagram of Figure 4. The order relations among the objects are maintained if one proceeds on paths where either $\Delta\epsilon_i \geq 0$ or (exclusively) $\Delta\epsilon_i \leq 0$.

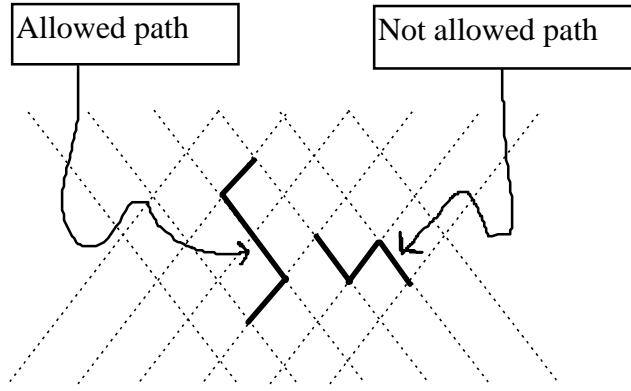


Figure 6: Allowed and not allowed pathes within the grid.

- Broken lines denote the regular grid by which the Hasse diagram is embedded.
- Some objects and connecting lines could be embedded into the grid on alternative ways. However, the order theoretical information, namely the comparabilities are maintained.
- Some patterns of the scores of the test battery are additionally shown. From them the values of the scores of other objects can be estimated or exactly calculated. For example, for class 17, FC must have the value 3, because the lower object 92 and the higher object 95 have $s_{FC} = 3$. The value of CP must be 0 because $s_{CP}(32) = 0$, which is the lowest value. Similarly $s_{CH}(17) = 0$ and $s_{GT}(17) = 0$, whereas for $s_{MT}(17)$ only the interval $4 \leq s_{MT}(17) \leq 6$ can be predicted from the knowledge of the neighbours in the Hasse diagram.

The grid can be thought of as being a coordinate system, with one axis of a latent order variable ϵ_1 and another by ϵ_2 . This is another expression of the fact that the dimension of the poset $(E/\mathcal{R}_5, IB)$ equals 2. By these two latent ordering variables, each class $\in E/\mathcal{R}_5$ can be characterized by a pair. Note that this characterization represents correctly the order relations (important for ranking) but is clearly not unique with respect to a numerical representation. Table 4 gives some examples.

Table 4: Order theoretical classification of representatives found in Figure 3.

<i>Representative</i>	ϵ_1	ϵ_2	<i>remark</i>
11	0	0	least element
91	4	0	
3	4	2	
17	5	4	
9	2	6	maximal element

Obviously the latent variable ϵ_1 describes the first three tests, whereas the latent variable ϵ_2 expresses the test results of MT or GT.

Summarizing: The dimension analysis shows that with respect to ranking there is a considerable redundancy, furthermore it helps to find new variables which simplify the interpretation of the ranking by partial orders: Some objects are ranked high, because at least one of the first three tests is high, some other objects are ranked high, because MT or GT have large values.

Structural information (task 1)

1) Subsets of specific patterns:

Application of Eq. 4 lead to the subsequent classification: Selecting a maximal element with zeros in their scores. For example the key element 32 may be considered as represented by the template $(*,0,0,*,0)$. Now all successors of 32 must have at least the same pattern of zeros as object 32. The other tests may have decreased values, including zero. Let us now consider the intersection of $G(32)$ and $G(31)$. Their templates are: $(*,0,0,*,0)$ and $(*,*,*,0,0)$. Then the template for all objects of $G(32) \cap G(31)$ must have the form: $(*,0,0,0,0)$. For all common successors of the objects 32 and 31 s_{FC} may have values greater 0, whereas we know that s_i ($i=CP,CH,MT,GT$) must be 0. The representatives of $G(32) \cap G(31)$ are 4, 91 and 11. For these samples we therefore know their templates and thus their qualitative loading pattern without using (lengthy and boring) tables.

Partitioning

By applying Eq. 4 and the procedure explained above a partitioning of the set of samples is suggested (Table 5).

Table 5: Partitioning of the set of samples

Name of the subset	Representatives of the subset	Common template	Generating key elements	Number of "0's" ^a
F	4,11,25,27,91	$(*,0,0,0,0)$	27	4
FM	1,2,3,7,12,14,17,32,92	$(*,0,0,*,0)$	32	3
FMG	9,18,23	$(*,0,0,*,*)$	9,18	2
FCC	5,31	$(*,*,*,0,0)$	31	2
"95"	95	$(*,*,*,*,0)$	95	1

^{a)} The number of "0's" defines the order, by which the subsets are formed. In case of equality **and** conflicting assignments an arbitrary choice is to be made and as such documented.

According to table 5 the set E/\mathfrak{R} can be partitioned into:

$$E = F \oplus FMG \oplus FM \oplus FCC \oplus \{95\}^9$$

Therefore from the battery of tests the following hierarchy (Figure 7) can be derived:

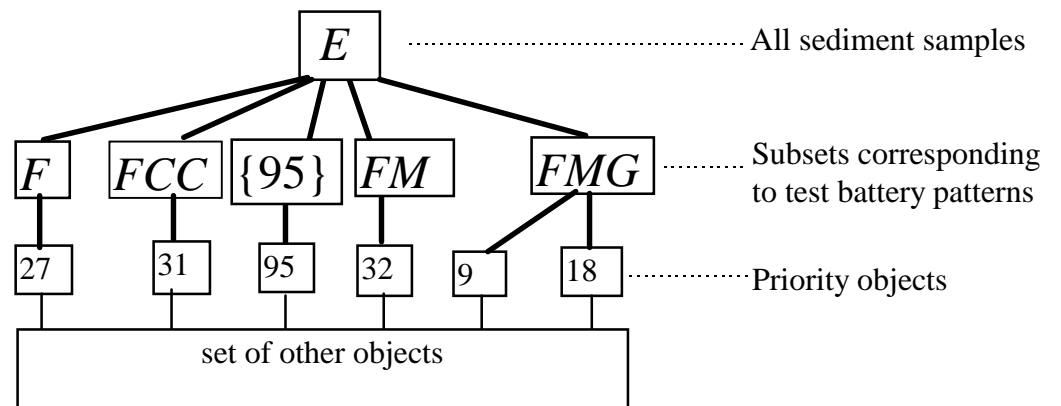


Figure 7: Hierarchy of sediment samples of Lake Ontario

⁹ As representatives. The symbol \oplus means a union of disjunctive sets.

Aggregation and further structural information (task 2)

Sample 9 cannot be compared with many other equivalence classes. $U(9)=17$, whereas $U(32)=8$. Indeed for sample 9 the pattern: High score for MT and a medium value for GT is a singular property, which leads to only few objects comparable with sample 9. Analogously from the numbers $U(18)=16$ and $U(23)=17$ resp. one finds that the group 18 and 23 is comparable only with 2 and 11 resp. Especially the local incomparabilities indicate a high potential for arbitrariness in rankings based on aggregation procedures, like Γ of Eq. 1. Again the pattern of results of the battery for sample 18 is rather singular, since there are high GT-values. To be more specific: If there is a consensus, that GT is not as important as the other tests, then its score will get a low weight in Eq. 1a. In consequence the sites 18, 23 would be ranked very low, according to Eq.1a. If however, GT would have a high weight, then the sites 18, 23 may be located on the top of any total order.

In summary: Aggregation procedures would mask such singularities in objects or subsets of objects.

To estimate the peculiarity of objects of high degree of local incomparabilities we calculate $\langle U(x) \rangle$ by Eq. 5. $m=5$, $n=20$ $\langle U(x) \rangle = 19 \cdot 0.5^5 \approx 18$

However, the assumption of uncorrelated attributes, is important. As shown in *Brüggemann et al. (1999c)* two attributes, namely CP and CH are highly correlated. Furthermore the dimension analysis (see below) shows that only two latent variables are necessary to generate an isomorphic Hasse diagram. Therefore a more realistic value for $\langle U(x) \rangle$ would be:

$$\langle U(x) \rangle = 19 \cdot 0.5^2 \approx 14$$

$U(9) = 17$, therefore there is in comparison to the average some specific uncertainty to rank object (class) 9.

Dimension analysis (task 3)

The dimension of $(E/\mathcal{R}_5, IB)$ (Figure 3 and Figure 4 resp.) is 2, but the battery contains 5 tests, therefore there seems to be a high degree of redundancy within the ranking using the scores of the battery of tests. An order theoretical characterization of the samples can be performed by only two (latent) variables. Beyond this there is a qualitative assignment: FC, CP, CH corresponds to ε_1 and MT, GT to ε_2 . As mentioned above, a numerical relationship is hardly available.

The analysis of the dimension, should be seen within the following background:

- It is a result of a specific set of samples
- It depends on the data representation (here given by the classification, performed by Dutka et al.) and
- there may be causal explanations for high rank correlations.

3 Discussion

The battery of tests approach helps to evaluate objects using different criteria simultaneously: The decision of which sites are "good" or "bad" is more difficult the larger the number of samples and especially the larger the number of tests, since there is more information that can be used to differentiate among the tested objects. This, in turn, leads to difficulties for ranking, because the complexity of a well designed battery is being lost, if in order to compare the tested objects, an ordering index is constructed (e.g. Eq. 1).

Here an alternative method of data-driven evaluation is shown. The notation "data-driven" has its origin from the fact that an evaluation procedure is done by the help of the data themselves. Once the information base of evaluation, *IB*, is selected, then there is no subjectivity, derived, for example, from the selection of protection goals.

The presentation by a Hasse diagram must not be performed in isolation. The use of cluster analysis and principal component analysis may be helpful to obtain a statistical relevant data representation and to avoid insignificant numerical differences of the attributes. A very promising approach is started by *Sørensen et al., 1998a* and *b* to overcome this problem, another strategy is to introduce the concept of robustness, which is discussed in *Brüggemann and Bartel (1999)*). Furthermore a data analysis by multivariate statistics may support the interpretation of the partial ordering. The data, used here from the publication of *Dutka et al.*, serve as demonstration. An example of data analysis by multivariate statistics is not in the scope of this publication.

4 References

Ahlf W. & Gratzner H. (1998): Erarbeitung von Kriterien zur Ableitung von Qualitätszielen für Sedimente und Schwebstoffe - Entwicklung methodischer Ansätze. Report of the German Environmental Protection Agency, UBA-FB 98-119

Atkinson, M.D. (1989): The complexity of Orders. In: Algorithms and Order NATO ASI series, Series C: Mathematical and Physical Sciences, Vol. 255 ISBN 0-7923-0007-6, pp. 195-230. Editor: I. Rival. Kluwer Academic Publishers, Dordrecht.

Bartel, H.-G. (1996): Mathematische Methoden in der Chemie, Oxford: Spektrum Akademischer Verlag

Barthelemy, J.P. (1989): Social Welfare and Aggregation Procedures: Combinatorial and Algorithmic Aspects. In: Applications of Combinatorics and Graph Theory to the Biological and Social Sciences, pp. 39-73. Editor: F. Roberts. Springer-Verlag, New York.

Bondy, J.A., & Murty, U.S.R. (1976): Ranking the participants in a tournament. In *Graph theory with applications*. (pp. 185-190). New York: Elsevier/Mc Millan.

Brüggemann, R. (1998): Order Theoretical Tools in Environmental Sciences. Introduction. In: *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences. Berichte des IGB, Heft 6, Sonderheft I*, 1-9

Brüggemann, R. & Halfon, E. (1997): Comparative Analysis of Nearshore Contaminated Sites in Lake Ontario: Ranking for Environmental Hazard. *J Environ Sci Health*, **A32**(1), 277-292.

Brüggemann, R., Pudenz, S., Voigt, K., Kaune, A. & Kreimes, K. (1999b): An algebraic/graphical tool to compare ecosystems with respect to their pollution. IV: Comparative regional analysis by Boolean arithmetics. *Chemosphere*, 38, 2263-2279.

Brüggemann, R. & Bartel, H.-G. (1999): A Theoretical Concept to Rank Environmentally Significant Chemicals. *J. Chem. Inf. Comp. Sc.* 39, 211-217.

Brüggemann, R., Bücherl, C., Pudenz, S. & Steinberg, C. (1999a): Application of the concept of Partial Order on Comparative Evaluation of Environmental Chemicals. *Acta hydrochim. hydrobiol.* 27, 170-178.

Brüggemann, R., Halfon, E. & Steinberg, C. (1999c): (submitted)

Clark, J. & Holton, D.A. (1994): Graphentheorie Grundlagen und Anwendungen, pp. 1-363. Spektrum Akademischer Verlag, Heidelberg.

Davey B. A. & Priestley H. A. (1990): *Introduction to Lattices and Order*. Cambridge University Press, Cambridge.

Dutka, B. J., Walsh, K., Kwan, K. K., El Shaarawi A., Liu, D. L. & Thompson, K. (1986): Priority site selection from degraded areas based on microbial and toxicant screening tests. *Water Poll Res J Canada*, **21**(2), 267-282.

Ganter, B. & Wille, R. (1996): *Formale Begriffsanalyse Mathematische Grundlagen*, Berlin:Springer-Verlag

Graham, R.L. (1982): Linear extensions of partial orders and FKG Inequality. In: *Ordered Sets NATO Advanced Study Institutes Series C: Mathematical and Physical Sciences C83*, pp. 213-236. Editor: I. Rival. D.Reidel Publishing Company, Dordrecht.

Grell, J. & Brüggemann, R. (1998): On Dimension Analysis in Comparative Regional Evaluation. In: *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences. Berichte des IGB, Heft 6, Sonderheft I*, 85-93.

Halfon E. & Reggiani M.G. (1986): On ranking chemicals for environmental hazard. *Environ. Science Tech* **23**, 1173-1179.

Kemeny J.G. (1959): Mathematics Without Numbers. *Daedalus* **88**, 577-591.

Klein, D.J. & Babic, D. (1997): Partial orderings in Chemistry. *J.Chem.Inf.Comp.Sc.* 37:656-671.

Klein, J., Brüggemann, R., Voigt, K. & Steinberg, C. (1995): Advances in Comparative Analysis of Adverse Effects in Aquatic Ecosystems with Emphasis on Studies with Humic Substances and on Progress in Mathematical Analysis Technique. *Water Research*, 29, 2261-2268.

Lipschutz, S. (1976): Discrete Mathematics. New York: Mc Graw-Hill Inc. (Schaum's Outline Series).

Lundgren, J.R. (1989): Food webs , Competition graphs, Competition-common enemy graphs, and niche graphs. In: *Applications of Combinatorics and Graph theory to the Biological and Social Sciences*, pp. 221-243. Editor: F. Roberts. Springer-Verlag, New York.

Roberts, F.S. (1989): Applications of Combinatorics and Graph Theory to the Biological and Social Sciences : seven Fundamental Ideas. In: *Applications of Combinatorics and Graph Theory to the Biological and Social Sciences*, pp. 1-37. Editor: F. Roberts. Springer-Verlag, New York.

Schneeweiss, C. (1991): *Planung 1 - Systemanalytische und entscheidungstheoretische Grundlagen*. Berlin: Springer-Verlag.

Sørensen, P.B., Mogensen B.B., Carlsen L., & Thomsen M. (1998a): The role of uncertainty in Hasse diagram ranking. In *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences .Berichte des IGB, Heft 6, Sonderheft I*, 71-84.

Sørensen, P.B., Mogensen, B.B., Gyldenkaerne, S., & Rasmussen, A.G. (1998b): Pesticide Leaching Assessment Method for Ranking both Single substances and Scenarios of Multiple Substance Use. *Chemosphere* **36**(10), 2251-2276.

Sørensen, P.B., Carlsen, L. & Mogensen, B.B. (1999): "Partial order ranking (Hasse diagrams): How to make interpretation of the results.", 9th Annual Meeting of SETAC-Europe Leipzig - Germany, 25-29 May 1999.

Stanley, R.P. (1986): Enumerative Combinatorics Volume I, Monterey:Wadsworth&Brooks/cole.

Trotter, W.T. (1991): Combinatorics and Partially Ordered Sets Dimension Theory; John Hopkins Series in the Mathematical Science, The J. Hopkins University Press, Baltimore.

Welzl, G., Voigt, K. & Rediske, G. (1998): Visualisation of Environmental Pollution - Hasse diagram technique and Explorative Statistical Methods. In *Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences. Berichte des IGB Heft 6, Sonderheft I*, 101-110.

Winkler, P. (1982): Average height in a partially ordered set. *Discr. Math.* 39, 337-341.

Comparative Evaluation of Materials in Car Production

Stefan Pudenz¹⁾, Thomas Bittner²⁾, Rainer Brüggemann¹⁾

*¹⁾ Institute of Freshwater Ecology and Inland Fisheries
Dep. Ecohydrology
Rudower Chaussee 30
D-12489 Berlin
Germany*

*²⁾ Volkswagen AG
Dep. Environmental Planning
PO Box 1897
D-38436 Wolfsburg
Germany*

Abstract

For industrial/practical application of comparative evaluation of objects by the method of partial order and the technique of Hasse diagrams (HDT) respectively, several characterizing parameter are developed. In cooperation with the Volkswagen AG examples for the evaluation of different plastic foils for packaging are worked out. Therefore several attributes are selected which are used as criteria for the evaluation of materials by the Departments of Environmental Planning and Operational Safety. Evaluations were done by experts by means of aggregation of attribute values into one quantity on the one hand and by the objective method of HDT on the other hand. It is seen that HDT leads to a more distinct evaluation of foils with respect to low hazard potential. Instead of 7 foils resulting from evaluation by aggregation of attributes, HDT identifies only 2 foils which have actually lower values in all attributes compared to all other foils. Furthermore a parameter that facilitates the selection of substitutes of materials is introduced and applied on the comparative evaluation of foils. With two other quantities, namely selectivity and diversity, a measure is given of how far the ranking result is similar to a total order and antagonism of attribute values leads to incomparabilities between foils/objects respectively.

1 Introduction

In factories of car production a wide range of materials are used. Ranging from a simple washing powder for dirty hands of workers to high-tech materials for engine production sometimes a selection among different materials serving the same purpose is necessary.

Table 1: Data matrix for the evaluation of plastic foils (VCI foils) by the department of Operational Safety. Binary values express the existence or non-existence, i.e. 0 or 1, of an attribute and a chemical respectively. The resulting hazard classification by experts is shown in the last column.

No.	Name	Attributes						HC Resulting hazard classification
		NI Natrium- nitrite	NA Natrium- nitrate	DI Dicyclo- hexylamin- compound	AA Other Amin	A+ Combi- nation Amin + Chemical X	PE Necessity of personal protecting equipment (yes/no)	
1	A1	1	0	0	1	1	0	3
2	A2	1	0	0	1	1	0	3
3	B1	0	0	1	0	0	0	2
4	B2	0	0	0	0	0	0	1
5	B3	1	0	1	0	1	0	2
6	B4	0	0	0	1	0	0	1
7	O	0	0	0	0	0	1	1
8	C	0	0	1	0	0	0	2
9	E1	1	0	0	0	0	0	1
10	E2	1	0	0	0	0	0	1
11	E3	1	0	0	0	0	0	1
12	C1	0	0	-	-	-	-	3
13	C2	0	0	-	-	-	-	3
14	C3	0	0	-	-	-	-	3
15	F1	-	-	-	-	-	1	3
16	F2	-	-	-	-	-	1	3
17	F3	-	-	-	-	-	1	3
18	F4	-	-	-	-	-	1	3
19	F5	-	-	-	-	-	1	3
20	M1	1	0	1	0	0	0	3
21	M2	0	0	1	0	0	0	2
22	M3	0	0	0	0	0	0	1
23	N	0	0	1	0	0	0	2

Dependent on the economic, technical aim and/or protection aims like environment and human/employees health different compilations of criteria are used for a selection and a comparative evaluation of materials respectively. This means that in any case the comparative evaluation of materials will be performed by means of several criteria. At the Volkswagen AG comparative evaluations of materials are often done by aggregating several criteria/attributes, where an attribute is the quantitative expression of an criteria. For example the comparative evaluation of several plastic foils for packaging (abbrev. VCI-foils) by the Department of Operational Safety at Volkswagen uses the attributes in Table 1. Aggregating of all attribute values for each foil leads to the hazard classification in the last column, where foils/objects with gaps in

their attributes get the highest value with regard to the whole hazard classification (conservative replacement). Thus each foil is characterized by only one quantity, namely the hazard classification, and a ranking of all VCI foils by means of a total order is practicable.

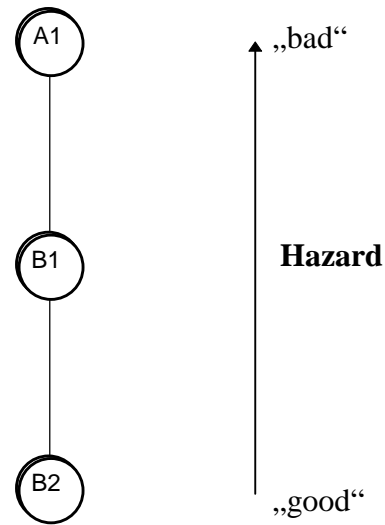


Figure 1: Hasse diagram for the ranking of VCI foils only by hazard classification (HC). Circles with a second line indicate the presence of an equivalence class, i.e. foils/objects with equivalent data values. Namely: {A1, A2, C1, C2, C3, F1, F2, F4, F5, M1}, {B1, B3, C, F3, M2, N}, {B2, B4, O, E1, E2, E3, M3}.

The graphical visualization of the ranking by HC shows Figure 1. Obviously A1 and all its equivalents are most hazardous compared to the other equivalence classes. However as was already shown in several publications (e.g. Brüggemann et al. 1999, Brüggemann & Bartel, 1999) this kind of ranking, i.e. aggregating of attributes to one quantity, may cover information and aggregates attributes respectively, which originally may lead to antagonism and therefore to incomparabilities between foils/objects. For example, the Hasse diagram in Figure 2 which is the result of ranking by means of all attributes, where all data gaps are filled by maximum values (i.e. 1), shows, that originally A1 is not comparable with B1. Moreover it is no more a maximum element, i.e. an element that has no upper neighbour but at least one lower neighbour. Instead of A1 and its equivalents, now F1, F2, F3, F4 and F5 are maximum elements and therefore most hazardous compared to all other foils (with respect to all criteria). It is striking too that the minimal elements, i.e. elements which have no neighbour below but at least one upper neighbour, are diminished as well ({B2, M3} instead of {B2, B4, O, E1, E2, E3, M3}). However, instead of discussing the differences between both rankings in Figure 1 and 2, here we will focus our attention on the practical use of Hasse diagram technique (HDT) in car production, especially the task of finding substitution materials which fulfil certain criteria and the interpretation of such Hasse diagrams with the objective to get holistic information about the ranking result. Therefore we will introduce information parameter concerning the task of substitution option of materials and so-called form indices which may facilitate the interpretation of Hasse diagrams and the ranking result respectively.

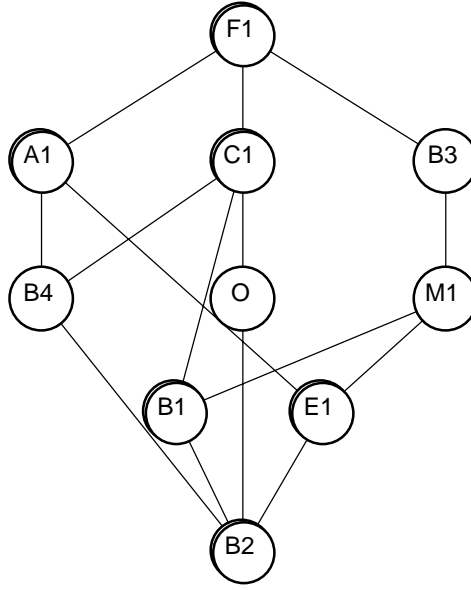


Figure 2: Hasse diagram for the ranking of VCI foils by means of all attributes in Table 1, except hazard classification (HC). Equivalence classes: {A1, A2}, {B1, C, M2, N}, {B2, M3}, {E1, E2, E3}, {C1, C2, C3}, {F1, F2, F3, F4, F5}.

2 Information parameter for industrial application of HDT

2.1 Substitution option for materials

Almost all materials used in production process are subject of a substitution check. Using environmental (chemical and toxicological information) and/or operational safety criteria this work is done by experts without any computational evaluation algorithm/system so far. The result of a comparative evaluation by means of criteria which are relevant for operational safety is already shown above (Table 1, Figure 1 and Figure 2). For example based on the ranking result by HC substitutes for A1 are all foils that are below (i.e. B1, B3, C, F3, M2, N, B2, B4, O, E1, E2, E3, M3) because they all have a lower hazard classification than A1. However using the result of the comparative evaluation by means of all attributes, all foils which are below and connected with A1 are potential substitutes, i.e. E1, E2, E3, B4, B2 and M3. Obviously these are by far less foils than based on the hazard classification because only this foils are potential substitutes which have lower values in all attributes. Formalizing the task, how many objects are "better" than an arbitrary object x we will introduce the degree of freedom of x :

$$FG(x) := \text{card } O(x), \quad \text{Eq. 1}$$

Degree of freedom of x

where $O(x)$ is the order ideal of object/foil x (or its equivalents; see Introduction), which includes all objects that can be reached downwards from x without turn in direction and without counting the equivalents of x . The symbol $card$ is the cardinality of a set. For example $FG(C1) = 8$, where $O(C1) = \{O, B4, B1, C, M2, N, B2, M3\}$ and $FG(M1) = 9$ with $O(M1) = \{E1, E2, E3, B1, C, M2, N, B2, M3\}$.

Certain degree of freedom

It may be also of interest to know how many objects/foils exist which have a certain distance d to an object x . Therefore we define a certain degree of freedom:

$$SFG_{d_0} := card\{y \in O(x) \text{ with } d(x, y) > d_0\}, \quad \text{Eq. 2}$$

where $d(x, y)$ is the minimum number of edges on a path from x to y such as to pass through the edges without any turn in direction. For example:

$$SFG_2(B_3) = 2 \text{ and } O_{d_0=2}(B_3) = \{B2, M3\}.$$

Selection criterion

Furthermore sometimes substitutes have to fulfil a definite criterion. For example foil F1 has to be substitute by a foil whose treatment needs no personal protecting equipment and that has a distance greater than d_0 . We will define this definite criterion as a "selection criterion" (Sek) because it only selects foils/objects fulfilling a certain criterion. In other words, Sek is a criterion by which a subset of the original object/foil set P is composed:

$$Sek(P) := \{x \in P : x \text{ fulfils criterion } Sek\}. \quad \text{Eq. 3}$$

Then $x(a, Sek)$ is a substitute for a at a distance $> d_0$ and under the selection criterion Sek . For example, selecting $d_0 = 1$ and $Sek =$ no personal protecting equipment (NPE) it follows $SFG_1(F1) = 12$ and

$$O(F1) = \{B4, M1, O, E1, E2, E3, B1, C, M2, N, B2, M3\} \text{ with } d(F1, x) > 1.$$

Application of Sek leads to substitutes

$$x(F1, NPE) = (B4, M1, E1, E2, E3, B1, C, M2, N, B2, M3),$$

where compared to $O(F1)$ in P and with $d(F1, x) > 1$ only foil O does not fulfil Sek , i.e. NPE (see also Table 1).

2.2 Form indices

In order to ensure the flow of information within the company and to hold information about comparative evaluations of materials at the disposal of decision maker it is of advantage to store the ranking results in a data bank. As ranking result it may be obvious to store the complete Hasse diagram. However in case of comparing the results of for example several evaluations using different compilations of attributes, i.e. at least more than one Hasse diagram, it may be useful to use indices which

facilitate the interpretation of different ranking results and the comparisons among each other.

Here we will introduce two so-called form indices, which characterize

- I. the selectivity of ranking and
- II. the diversity of ranking,

that is selectivity and diversity respectively, between objects ranked as "good" and those ranked as "bad".

For example the Hasse diagram from ranking of VCI foils by means of only one attribute represents a total order (see Figure 1). Because here all elements are arranged one upon an other and connected by a line, i.e. they form a "chain", and therefore indicate a unique orientation onto "good" and "bad", a maximum of selectivity is reached. The opposite, i.e. a minimum of selectivity represents a so-called antichain. That means, all elements are incomparable among each other and therefore arranged side by side without any connecting lines (see also Figure 3).

Selectivity

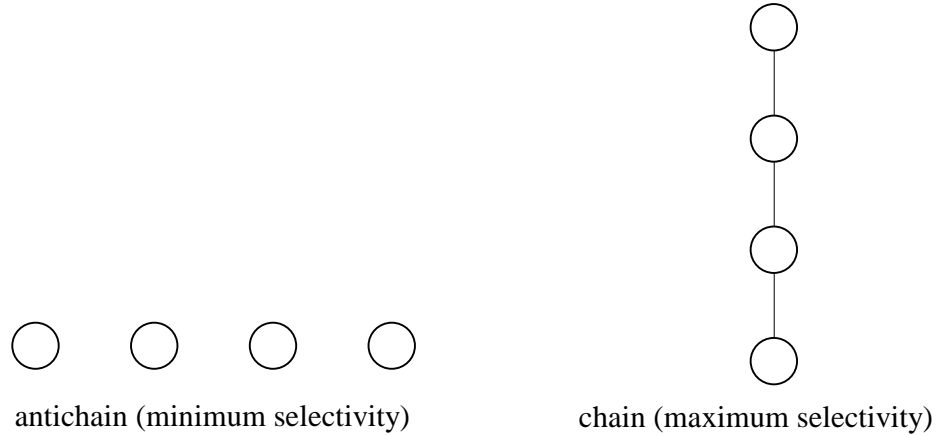


Figure 3: Maximum and minimum selectivity of ranking. Here the chain denotes a total order whereas the antichain represents a maximum number of incomparabilities in a partial ordered set.

Now, to give a formalism defining this "good-bad selectivity" seems to be very easy: denoting the number of elements in the chain as number of layer, NL , and the number of elements of an object set P as N (where $N > 1$), the selectivity standardized on a 0..1 scale is

$$t = \frac{NL - 1}{N - 1}. \quad \text{Eq. 4}$$

However, if we apply Eq. 4 on the Hasse diagram in Figure 2 difficulties arise because more than one chain can be obtained. Following the lines upwards from the lowest element, B2, to the element at the top of the Hasse diagram, i.e. F2, we find 7 chains and total orders respectively (see Figure 4). The largest unique orientation onto "good" and "bad", that is the largest chain from the bottom to the top of the diagram, are chains no. IV and V. Finally, the number of level NL is defined as the number of elements in the largest chain, without counting the equivalents, i.e. in our case $NL = 5$. That means,

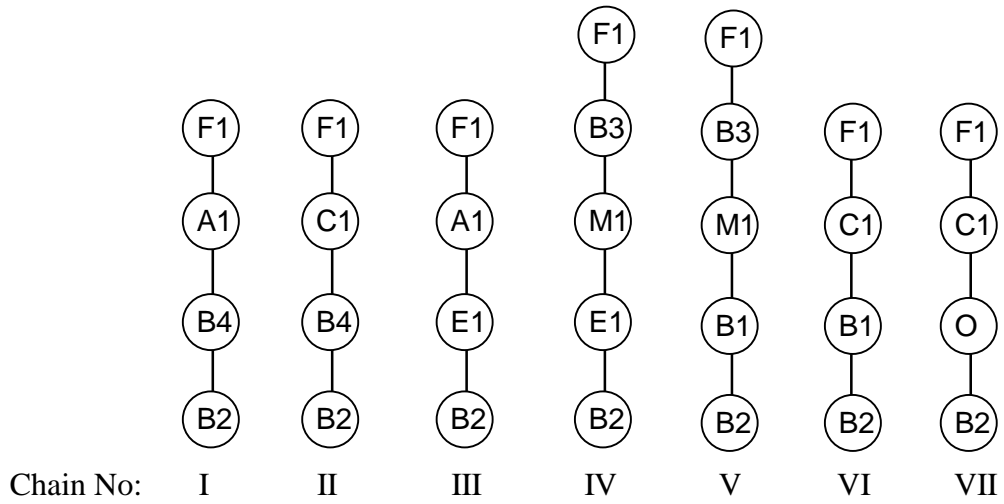


Figure 4 : Chains and total orders respectively, of the Hasse diagram in Figure 2.

NL is related to the quotient set P/\mathfrak{R} and not to the object set P , where \mathfrak{R} is the equivalence relation between two objects in the partial ordered set, for example $A1 \sim A2$ in Figure 2. Furthermore, for selectivity we also have to differentiate between object set and quotient set. For consider all elements of the Hasse diagram in Figure 2, i.e. related to the object set P , we calculate a selectivity of

$$t(P) = \frac{5-1}{23-1} = 0.18,$$

whereas related to the quotient set

$$t(P/\mathfrak{R}) = \frac{5-1}{10-1} = 0.44.$$

In order to have a better differentiation between the relations to object set and quotient set, we will introduce Z as the number of equivalence classes. Therefore in case of $t(P/\mathfrak{R})$ N has to be replaced by Z in Eq. 4:

$$t(P/\mathfrak{R}) = \frac{NL-1}{Z-1}, \text{ with } Z > 1. \quad \text{Eq. 5}$$

It is seen, that the calculated selectivity of 0.44 may reflect a visual interpretation of the Hasse diagram. Only for demonstration we will show another Hasse diagram which is the result of ranking of the same foils, but by means of ecotoxicological attributes (see Table 2) selected by the Dep. of Environmental Planing (Figure 5). Obviously the ranking result is very close to a total order which is also well reflected by a selectivity of 0.83 ($NL = 6, Z = 7$).

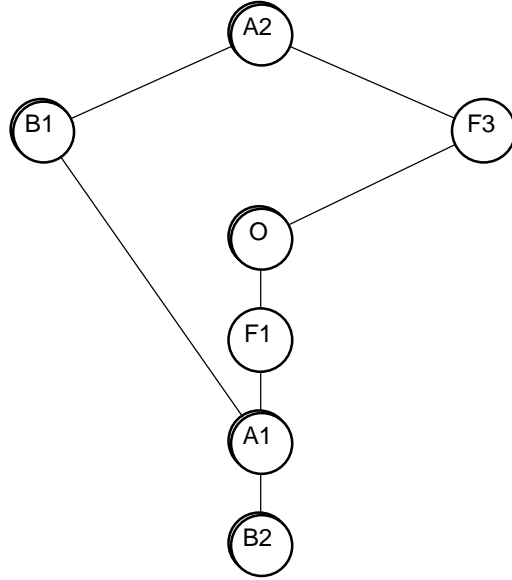


Figure 5: Hasse diagram from ranking of VCI foils by means of ecotoxicological attributes. Equivalence classes: {A1, F5, N}, {A2, C1, C2, C3, F4}, {B1, M1, M2, M3}, {B2, B3, B4, E1, E2, E3}, {O, C, F2}.

Diversity

Analogously to selectivity it may be also useful to introduce a measure of diversity of ranking. A ranking that results in many incomparabilities between elements, indicates that the object set used is characterized by a high diversity in the peculiarities of attributes.

Therefore that antichain, which contains the most elements represents the maximum of diversity. Levels are antichains. Although even the target level may not be the largest antichain, because of the conservatism in drawing Hasse diagrams levels will be the basis of the calculation. The "good-bad" diversity is expressed by the ratio of the number of elements in the largest level and the number of elements in the Hasse diagram, where it has to be differentiated too between the relation to object set P and the quotient set P/\mathfrak{R} . Again standardized on a 0..1 scale the diversity related to the object set P is

$$d(P) = \frac{NEL(P) - 1}{N - 1} \quad \text{Eq. 6}$$

and related to the quotient set P/\mathfrak{R}

$$d(P/\mathfrak{R}) = \frac{NEL(P/\mathfrak{R}) - 1}{Z - 1}, \quad \text{Eq. 7}$$

where NEL is the number of elements in the largest level.

Applying Eq. 7 on both Hasse diagrams in Figure 2 and 5, results in diversities of

$$d(P/\mathfrak{R}) = \frac{3 - 1}{10 - 1} = 0.22$$

and

$$d(P/\mathfrak{R}) = \frac{2 - 1}{7 - 1} = 0.17$$

respectively.

3 Conclusions

Applying the technique of Hasse diagrams on the comparative evaluation of materials leads to several advantages in comparison to the usual aggregation method. Following the latter one, the aggregation of attributes, i.e. the existence or non-existence of chemical ingredients of VCI foils, into the quantity hazard classification, 7 foils are obtained as "good" and less hazardous respectively, compared to all other foils. Here, the comparative evaluation with HDT by means of all attributes yields at least 2 foils (B2 and M3), where all their attributes have values better than all other foils. Therefore evaluation with HDT is more objective than aggregation into one quantity as often done by experts.

Introducing the degree of freedom FG and a certain degree of freedom SFG respectively, facilitates the selection of substitutes for materials.

Table 2: Data matrix for the evaluation of plastic foils (VCI foils) by the Department of Environmental Planning. Data gaps are replaced conservatively, i.e. with maximum values.

No.	Name	Attributes						
		AL	AE	SV	TV	Dep	WGK	Gef
		Classified by national waste law	Duty of disclosure for disposal	Materail utilization (yes = 0, no = 1)	Possibility of waste incineration with thermal dissipation (yes = 0, no=1)	Ability of deposit	Water hazard class	Hazardous material
1	A1	0	0	0	0	0	1	0
2	A2	-	-	-	-	-	-	-
3	B1	0	0	0	0	0	-	0
4	B2	0	0	0	0	0	0	0
5	B3	0	0	0	0	0	0	0
6	B4	0	0	0	0	0	0	0
7	O	1	1	1	1	1	1	0
8	C	1	1	1	1	1	1	0
9	E1	0	0	0	0	0	0	0
10	E2	0	0	0	0	0	0	0
11	E3	0	0	0	0	0	0	0
12	C1	-	-	-	-	-	-	-
13	C2	-	-	-	-	-	-	-
14	C3	-	-	-	-	-	-	-
15	F1	1	1	0	1	1	1	0
16	F2	1	1	-	1	1	1	0
17	F3	1	1	-	1	1	2	1
18	F4	1	1	-	1	1	3	1
19	F5	0	0	0	0	0	1	0
20	M1	0	0	0	0	0	-	0
21	M2	0	0	0	0	0	-	0
22	M3	0	0	0	0	0	-	0
23	N	0	0	0	0	0	1	0

Moreover foils can be identified which fulfil a certain criterion and which are less hazardous than a foil subject of investigation/substitution.

The quantities selectivity and diversity can be used as parameter characterizing both the ranking result with respect to the shape of the Hasse diagram, for example approaching a total order and therefore a high potential of substitutes, and the effect of attributes selected on ranking. Moreover a high diversity also implies many of antagonism among the foils/objects with respect to their attribute values, therefore showing there is a high degree of freedom to select a safe substitute with additional constraints.

Further characterizing parameters with respect to industrial/practical application of comparative evaluation by HDT were under investigation. A summarized presentation of HDT and its practical use including a formulary is still in preparation.

4 References

Brüggemann, R., Bücherl, C., Pudenz, S. & C.E.W. Steinberg (1999): Application of the Concept of Partial Order on Comparative Evaluation of Environmental Chemicals. *Acta Hydrochim. Hydrobiol.*, 27 (3), 170-178.

Brüggemann, R. & H-G. Bartel (1999): A Theoretical Concept to Rank Environmentally Significant Chemicals. *J. Chem. Inf. Comput. Sci.*, 39, 211-217.

Estimation of Octanol-Water Distribution Coefficients

Using Partial Order Technique

Lars Carlsen*, Peter B. Sørensen and Marianne Thomsen

*National Environmental Research Institute
Department of Environmental Chemistry
DK-4000 Roskilde, Denmark*

Abstract

Partial order ranking appears as an attractive alternative to conventional QSAR methods, the latter typically relying on the application of statistical methods. In the present study it is demonstrated that QSAR models based on a partial order ranking approach satisfactorily can be used to predict octanol-water distribution coefficients for a range of organic compounds, using well established LSER descriptors for the validation of the approach. The precision and the uncertainties of the method is discussed in terms of the number of descriptors involved and the number of possible comparison between the investigated compounds. The requirements to - and the limitations of - the method are discussed.

Introduction

100.000 chemical substances

Today more than 100.000 chemical substances are in use and constitute a potential risk to the environment [EEA, 1998]. It is obvious that it is not practically possible experimentally to generate all necessary input for the risk assessment of these compounds. Thus it appears necessary that information concerning these substances fate and effect in the environment must be obtained through modelling, e.g., by comparison with structurally related, well investigated compounds.

QSAR

QSAR - Quantitative Structure Activity Relationships - in general terms denotes models which, based on the variation in structural and/or electronic features in series of selected, molecules, describe variation in a given end-point of these molecules. These end-points may be, e.g., biological effects or physical-chemical parameters, which experimentally can be verified. Based on the developed QSAR model end-points of new, structurally related compounds, hitherto not being experimentally studied, may be predicted.

Since the variation in, e.g., biological effects or physical-chemical parameters typically can not be described by one single descriptor, QSAR modelling relies heavily on statistical methods. Further, since QSAR modelling may often involve seeking unknown relations between several descriptors and a given end-point, traditional statistical approaches such as simple multilinear regression (MLR) may not be the ideal choice although widely used. Thus, development of QSAR models are often successfully based on multivariate projection methods, such as principal component analysis (PCA) followed by MLR using the principal components as descriptors or, more common, partial least square (PLS) projection, as the modelling in many cases can be described by linearization of complex unknown relations. Further the data material often may include uncertainties and/or strong co-linearities. [Thomsen, *et al.*, 2000, and references therein]

Partial Order Ranking

Partial Order Ranking [Brüggemann *et al.*, 1995], which from a mathematical point of view constitute extremely simple, appears as an attractive and operationally simple alternative to the above rather demanding statistical method.

The partial order ranking method allows ranking of series of well investigated compounds, e.g., based on octanol-water distribution coefficients using selected descriptors characterizing the structural and/or electronic nature of the compounds. The mutual ranking of the compounds can then be compared to the ranking based on the experimentally derived values for octanol-water distribution coefficients. If the ranking model resembles the experimental ranking of the parameters under investigation, the model is validated and other compounds not being experimentally investigated, can be assigned a rank in the model and hereby obtain an identity based on the known compounds.

Octanol-water distribution coefficients

Linear Solvation Energy Relationship

The present study focus on the possible applicability of partial order ranking as a tool for QSAR modelling. Thus, we have selected a series of non-hydrogen bond donor molecules, which have previously been studied using statistically based QSAR's in order to verify the applicability of the partial order ranking method to a well-known system. Thus, octanol-water distribution coefficients [Kamlet *et al.* 1988] was retrieved for a group of approx. 40 compounds exhibiting rather different structural and electronic characteristics. The experimental data was closely mimicked through a Linear Solvation Energy Relationship (LSER) approach [Carlson, 1999; Kamlet *et al.*, 1998], the corresponding statistical approach being MLR. The present study applies the same molecular descriptors as the LSER studies, i.e., the volume ($V_i/100$), the polarity (π^*) and the hydrogen bond basicity (β) [Kamlet *et al.*, 1998].

2 Method

Partial order ranking is a mathematically simple technique allowing a series of elements, in the present case compounds, to be mutually ranked using more than one ranking parameter (descriptor) simultaneously (Brüggemann, GSF 20/95). Thus, two compounds are comparable if all descriptors of one compound have equal or higher values than the corresponding descriptor of the other compound. From this follows that for two comparable compounds A and B, A will be ranked higher than B if one or more descriptors of A are higher than the corresponding descriptors of B. If all descriptors of A equal the corresponding descriptors of B the two compounds are ranked equal.

The ranking of the compounds is graphically displayed in the so-called Hasse diagrams [GetSynapsed].

Degree of Agreement

To elucidate the goodness of the ranking model we have introduced a "goodness of fit" measurement disclosing the percentage of rankings in the model which can be refound in the experimental data. Thus, the Degree of Agreement, DA , is given by

$$DA = N_a / (N_a + RN_d) \quad (1)$$

where N_a is the number of agreements, i.e., the same rank is identified in the model and in the experimental data and RN_d is the number of disagreements, i.e., a contradiction exists between the rank of the two compounds in the model and the experimental data, respectively [Sørensen *et al.*, *subm-b*].

In the present study we have ranked the molecules under investigation applying molecular volume ($V_i/100$), polarity (π^*) and hydrogen bond basicity (β) as descriptors [Kamlet *et al.*, 1998]. Subsequently the model rankings were compared to the ranking of the single molecules based on

experimentally derived values for octanol-water distribution coefficients, the “goodnes of fit” being estimated according to eqn. 1.

Predicted values

For the compounds placed within the ranking net as displayed by the Hasse diagrams, i.e., the compounds not being located in the top (maximals) and bottom (minimals) layers, respectively, a further verification of the model applicability can be obtained by using the model to predict octanol-water distribution coefficients, respectively. The predicted values for a given compound X (*ValueX*) were obtained by simple arithmetic means between the lowest value of the comparable compounds ranked above X (*minAbove*) and the highest value of the comparable compounds ranked below X (*maxBelow*).

$$ValueX = (minAbove + maxBelow)/2 \quad (2)$$

The predicted values are compared to the corresponding experimentally derived values.

The QSAR modelling was carried out using the program POR/QSAR [Sørensen & Carlsen, 1999] whereas the Hasse diagrams were constructed using the program Hasse [GetSynapsed].

3 Data

The experimental data for octanol-water distribution coefficients as well as the LSER descriptors molecular volume ($V_i/100$), polarity (π^*) and hydrogen bond basicity (β) for a series of non-hydrogen bond donor molecules were retrieved from the work of Kamlet et al. [Kamlet et al., 1998]. In Table 1 the applied data are given.

Table 1. Experimental data for octanol-water distribution coefficients for selected molecules together with QSAR descriptors [Kamlet et al., 1998].

Number	Compound	Log Kow	Vi/100	π^*	β
1	n-C ₅ H ₁₂	3.39	0.553	-0.08	0
2	n-C ₆ H ₁₄	3.9	0.648	-0.04	0
3	n-C ₇ H ₁₆	4.66	0.745	-0.02	0
4	n-C ₈ H ₁₈	5.18	0.842	0.01	0
5	c-C ₅ H ₁₀	3	0.5	-0.01	0
6	c-C ₆ H ₁₂	3.44	0.598	0	0
7	CH ₃ CH ₂ CH=CH ₂	2.4	0.428	0.08	0.07
8	CH ₂ Cl ₂	1.15	0.336	0.82	0.1
9	CHCl ₃	1.94	0.427	0.58	0.1
10	CCl ₄	2.63	0.514	0.28	0.1
11	CFCl ₃	2.53	0.455	0.22	0.1
12	t-CHCl=CHCl	2.09	0.406	0.44	0.05
13	CCl ₂ =CCl ₂	2.88	0.578	0.28	0.05
14	CHCl=CCl ₂	2.35	0.492	0.53	0.05
15	CH ₃ CCl ₃	2.49	0.519	0.49	0.1
16	CH ₂ ClCH ₂ Cl	1.48	0.442	0.81	0.1
17	CHCl ₂ CHCl ₂	2.39	0.617	0.95	0.1
18	CCl ₃ CHCl ₂	3.05	0.7	0.62	0.1
19	C ₃ H ₇ Cl	2.04	0.45	0.39	0.1
20	C ₄ H ₉ Cl	2.64	0.548	0.39	0.1
21	(CH ₃) ₃ N	0.22	0.433	0.16	0.65
22	(C ₂ H ₅) ₃ N	1.36	0.704	0.14	0.71
23	(n-C ₃ H ₇) ₃ N	2.79	0.985	0.14	0.69
24	(C ₂ H ₅) ₂ O	0.89	0.505	0.27	0.47
25	(n-C ₃ H ₇) ₂ O	2.03	0.699	0.27	0.46
26	(i-C ₃ H ₇) ₂ O	2.03	0.699	0.27	0.47
27	CH ₃ COOCH ₃	0.18	0.424	0.6	0.42
28	CH ₃ COOC ₂ H ₅	0.73	0.521	0.55	0.45
29	CH ₃ COOC ₃ H ₇ -n	1.24	0.622	0.53	0.45
30	CH ₃ COOC ₄ H ₉ -n	1.82	0.716	0.51	0.45
31	CH ₃ CH ₂ COOC ₂ H ₅	1.21	0.622	0.53	0.45
32	HCOOC ₃ H ₇ -n	0.83	0.521	0.6	0.38
33	CH ₃ CN	-0.34	0.271	0.75	0.31
34	CH ₃ CH ₂ CN	0.1	0.369	0.7	0.31
35	CH ₃ -CO-CH ₃	-0.24	0.38	0.71	0.48
36	C ₂ H ₅ -CO-CH ₃	0.29	0.477	0.67	0.48
37	n-C ₃ H ₇ -CO-CH ₃	0.91	0.574	0.65	0.48
38	n-C ₄ H ₉ -CO-CH ₃	1.38	0.67	0.63	0.48
39	n-C ₅ H ₁₁ -CO-CH ₃	1.98	0.767	0.61	0.48
40	cyclohexanone	0.81	0.619	0.76	0.53
41	n-C ₃ H ₇ CH=O	0.88	0.48	0.63	0.41
42	n-C ₅ H ₁₁ CH=O	1.78	0.674	0.63	0.41
43	CH ₃ -SO-CH ₃	-1.35	0.466	1	0.76
44	CH ₃ -CO-N(CH ₃) ₂	-0.77	0.543	0.88	0.76
45	H-CO-N(CH ₃) ₂	-1.01	0.444	0.88	0.69
46	CH ₃ -CO-N(C ₂ H ₅) ₂	0.34	0.737	0.84	0.78

4 Results

In Fig. 1 the model ranking of the 46 compounds with respect to octanol-water distribution coefficients is displayed in the Hasse diagram format. The decision rules for the 3 descriptors are summarized in Table 2.

Table 2. Decision rules used in the ranking of octanol-water distribution coefficients.

<i>Descriptor</i>	<i>High Value</i>	<i>Low Value</i>
Volume ($V_i/100$)	high rank	low rank
Polarizability (π^*)	low rank	high rank
Hydrogen bond basicity (β)	low rank	high rank

“Goodness of fit”

Comparison of the above model ranking of the compounds to the ranking based on the experimental octanol-water distribution coefficients (cf. Table 1) disclosed that out of a total of 408 comparisons in the Hasse diagram, 407 and 1 was found to in agreement and in disagreement, respectively, compared to the experimental ranking, which gives a “goodness of fit”, DA , equal to 0.998.

In Fig. 2 the comparison between the experimentally derived octanol-water distribution coefficients and the corresponding values predicted by the model for the compounds located within the net (cf. Method) is depicted.

In certain cases it can be noted that significant differences between experimental and predicted octanol-water distribution coefficients prevail. In Fig. 3 the deviation of the predicted values for the octanol-water distribution coefficients from the corresponding experimentally derived values are displayed.

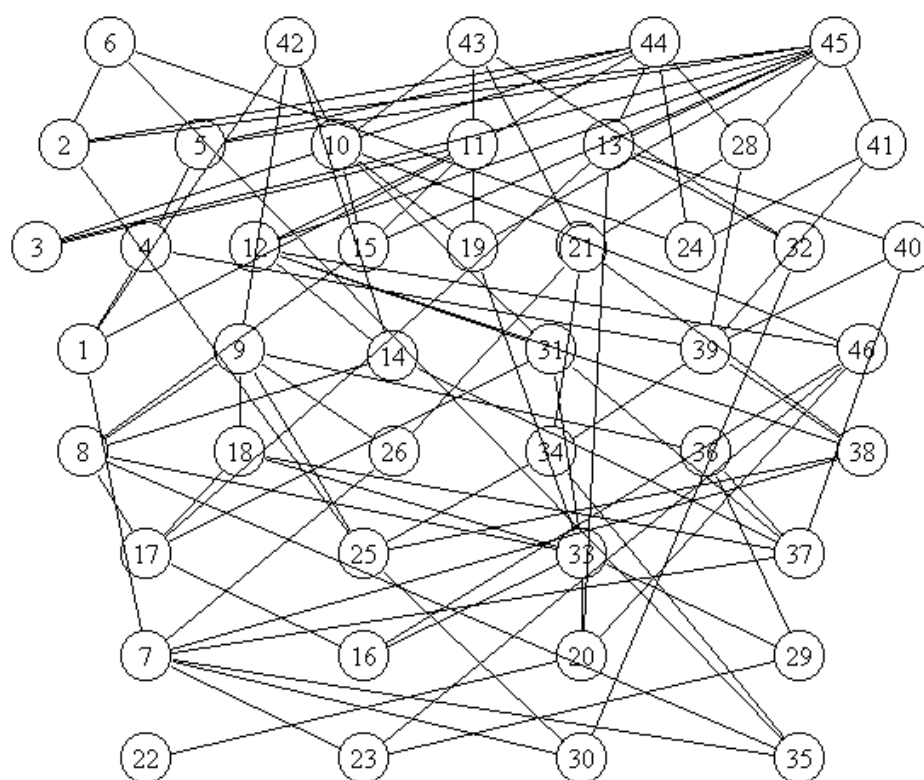


Figure 1. Hasse diagram displaying the model ranking of octanol-water distribution coefficients

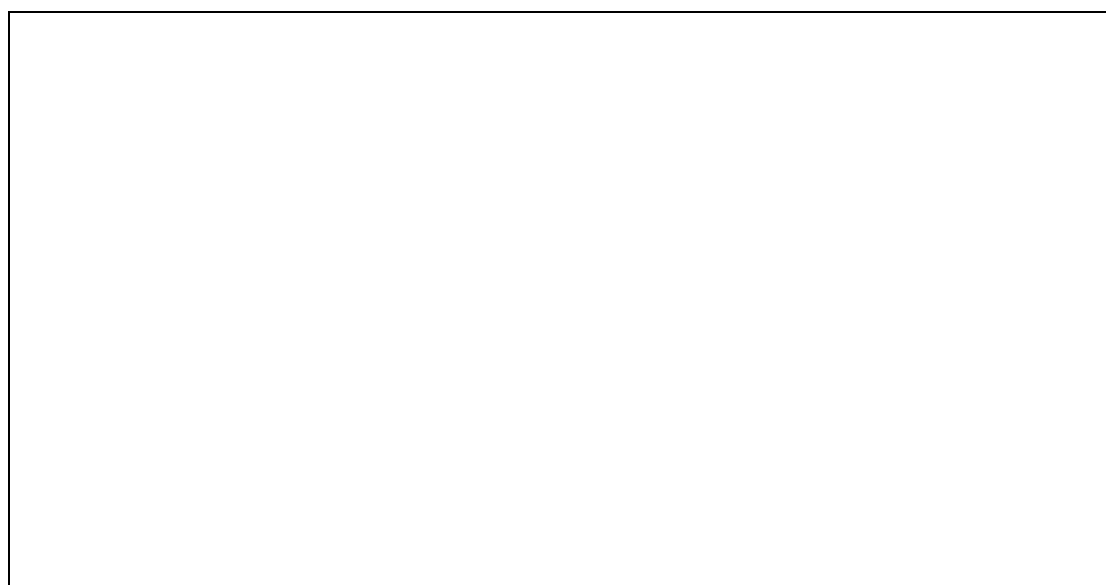


Figure 2. Experimental vs. predicted octanol-water distribution coefficients for the compounds located within the net.

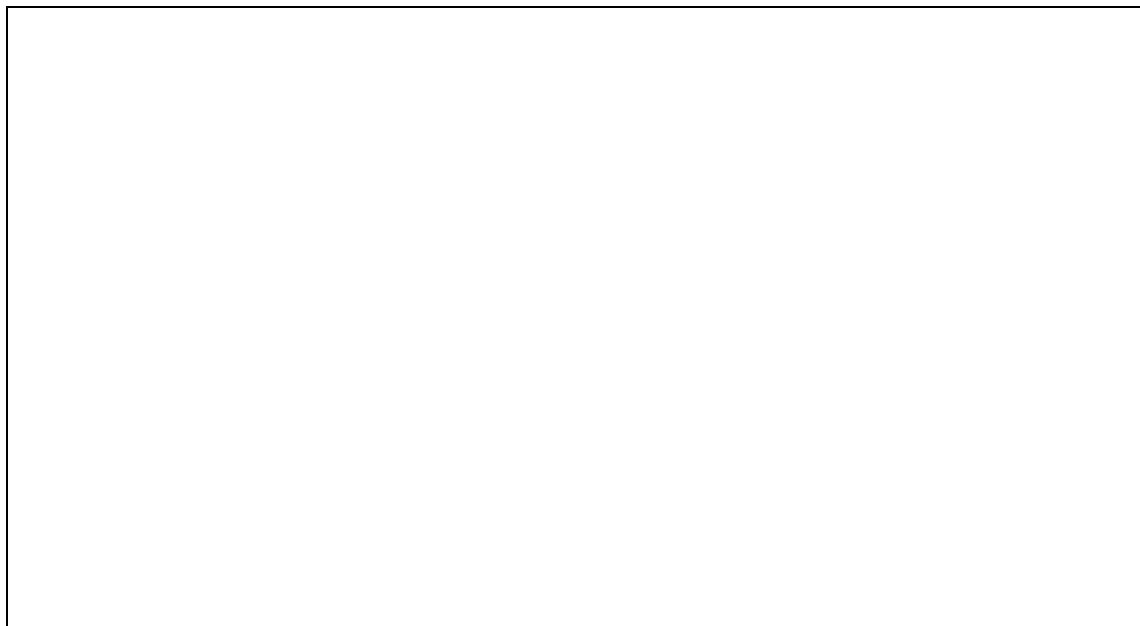


Figure 3. The deviation of the predicted octanol-water distribution coefficients from the values experimentally derived as function of the distance between the values of *minAbove* and *maxBelow*.

5 Discussion

It is immediately noted (Fig. 2) that in the partial order ranking based models octanol-water distribution coefficients reasonably well reproduce the experimentally derived values. However, it is equally clear that the actual distance between the *minAbove* and *maxBelow* elements is crucial. Thus, from Fig. 3 it can be deduced that as long as this distance is lower than approx. 1.5 log-units octanol-water distribution coefficients can be predicted within ± 0.4 log-unit, which for risk assessment often will be satisfactory.

In this connection it should be emphasized that predictions by the models are made as simple arithmetic means between the values of *minAbove* and *maxBelow* (cf. eqn. 2). Thus, the larger the distance between these two values the larger the potential uncertainty in the prediction (cf. Fig. 3).

Precision

To increase the precision of the model the *minAbove* - *maxBelow* distances have to be reduced, which obviously could be achieved by using increased basis sets of compounds. However, it is emphasized that a further crucial factor in this connection is the actual number of comparisons between the single compounds in the ranking model.

Requirements

Obviously two requirements to the model with regard to precision prevail. First of all the model must be able to rank the single compounds in the basis set correctly compared to the experimental data, which in the present study was achieved perfectly as disclosed by *DA* value (cf. eqn. 1) of 0.998. Secondly, the model should be based on a basis set of com-

pounds large enough to secure a satisfactorily fine-meshed net taking the number of descriptors into account.

In this context we have to deal with two demands apparently operating in opposite direction in order to develop a fine-meshed net. Thus, in order to model, e.g., physico-chemical parameters octanol-water distribution coefficients, it is often necessary to include several descriptors. In the present study we included three. On the other hand, an increase in the number of descriptors significantly influence the actual number of comparisons between the single compounds [Sørensen *et al.*, *subm.-a*]. This can to a certain extent be remedied by increasing the number of compounds included in the ranking model basis set [Sørensen *et al.*, *subm.-a*]

Limitation

A significant limitation of the model approach is that it is not possible to make predictions for compounds which turn out to maximals or minimals, i.e., they are ranked either in the top or the bottom of the ranking net as visualized by the Hasse diagrams. In these cases it is apparently only possible to state the octanol-water distribution coefficients to be above or below, respectively, of a certain value. Thus, in line with the above discussion it can be concluded that when developing QSAR models based on partial order ranking it should be assured that the basis set of compounds stretch a net that covers a range sufficiently large to accommodate unknown compounds to be studied within the net.

5.1 Selection of descriptors

Choice of descriptors

As is general the case, the choice of descriptors are crucial in order to set up a reasonable model. This may in the case of partial order ranking be even more crucial, since this method a priori allows a ranking of a set of compounds based on various descriptors independently of the actual nature of the latter. Thus, the descriptors obviously should be chosen to reflect, chemical and physical, the features to be modelled. Furthermore, the decision rules set up for the single descriptors should be explicable in chemical/physico-chemical terms.

In the present study the descriptors applied to rank the selection of compounds according to their octanol-water distribution coefficients were molecular (van der Waal) volume, $V_i/100$, polarizability of the molecules, \bullet^* , and the hydrogen bond basicity of the molecules, β , respectively [Kamlet *et al.*, 1998]. These descriptors all reflect important features in relation to the solubilization and the partitioning process. The volume is a descriptor for the size of the cavity in the solvent necessary to host the molecule, and the polarizability and hydrogen bond basicity terms are descriptors for the reorganization of the solvent around the solute molecule [Carlsen, 1999 and references therein].

Decision rules

Looking at the decision rules for the model ranking of the octanol-water distribution coefficients (Table 2), it appears that log Kow will increase with an increasing molecular volume, but decrease with increasing polarizability and hydrogen bond basicity. This in perfect agreement with an increased affinity for organic solvents simultaneous to a decreased affinity for water as reflected through the generally accepted linear free en-

ergy relationship between octanol-water distribution coefficients and water-solubility [Schwarzenbach *et al.*, 1993].

6 Conclusions and Outlook

Attractive alternative

It has been demonstrated that partial order ranking provides an attractive alternative to conventional QSAR modelling tools. The method appears, from a mathematical point of view, easy to handle as it does not require any application of statistical methods. It is further worthwhile to remember that in contrast to other QSAR modelling approaches, development of QSAR models based on the partial order ranking technique combines model development and validation in one step.

Predicting ability

The predicting ability of the model has been elucidated and it appears that within certain limits the precision of predicted octanol-water distribution coefficients are well acceptable for risk assessment purposes. However, the paper further suggests possibilities for a further improvement of the precision of the models.

It is possible using partial order ranking techniques to accommodate otherwise non-comparable descriptors and it is thus suggested that partial order ranking has a general potential in the area of risk assessment of environmentally hazardous chemicals. However, further analyses of the method appears appropriate to fully elucidate the potential of partial order ranking for QSAR modelling.

7 References

Brüggemann, R., Halfon, E., & Bücherl, C. , (1995): Theoretical base of the program “Hasse”, GSF-Bericht 20/95, Neuherberg.

Carlsen, L., (1999): Linear Solvation Energy Relationships (LSER). A method to assess the solubility of inorganic species, Research Notes from NERI, No. 107, National Environmental Research Institute, Roskilde, 38 pages.

EEA, (1998): Chemicals in the European environment: Low doses, high stakes?, European Environment Agency, Copenhagen, 33 pages.

Kamlet, M.J., Doherty. R.M., Abraham, M.H., Marcus, Y., & Taft, R.W., (1988): An improved equation for correlation and prediction of octanol/water partition coefficients of organic nonelectrolytes (including strong hydrogen bond donor solutes), J.Phys.Chem. **92**, 5244-5255.

Schwarzenbach, R.P., Gschwend, P.M., & Imboden, D.M., (1993): Environmental Organic Chemistry, Wiley, New York

Sørensen, P.B., Mogensen, B.B., Carlsen, L., & Thomsen, M. (subm-a): The influence on partial order ranking from input parameter uncertainty. Definition of a robustness parameter, submitted for publication.

Sørensen, P.B., Brüggemann, R., Mogensen, B.B., Carlsen, & L., Pudenz, S. (subm-b): Compound data analysis of monitoring data of pesticides in surface waters. Data interpretation and model development using partial order ranking theory, Submitted for publication.

Sørensen, P.B., & Carlsen, L., (1999): POR/QSAR - an Excel based software package for QSAR modelling. Currently available only as a non-verified β -version.

Thomsen, M., Hvidt, S., & Carlsen, L., (2000): Solubility of phthalates revisited. Environmental implications, in: ‘QSARs for Endocrine Disruption, Pollution Prevention, Toxicity Screening and Persistence Predictions’, J.D. Walker, ed., SETAC, in press

GetSynapsed, Hasse für Windows 1.0, Benutzerhandbuch, Gesellschaft für Software entlichlung und Beratung, München

Assessment of Water Management Strategies by Hasse Diagram Technique

U. Simon, R. Brüggemann

*Institute of Freshwater Ecology and Inland Fisheries,
12489 Berlin - Germany.*

Abstract

In the present paper first results of a multidisciplinary research project are introduced, which aims to develop a method to assess the sustainability of water management strategies. To practise the method of assessment, 40 different management strategies in the small catchment area of river Wuhle in Berlin Germany had been evaluated. in the small catchment area of river Wuhle in Berlin Germany. Each of the assessed strategies consists of one management element of the management groups: 1. wastewater treatment (five different technical advice), 2. rainwater treatment (four different ways to handle urban stormwater events) and 3. rivers morphology (two different shapes of the river bed). Every possible combination of the management elements - one element of every group - had been evaluated by a set of 15 indicators. Thus a data matrix of $5 \times 4 \times 2 = 40$ different strategies and 15 indicators arises. The comparative evaluation has been done by Hasse diagram technique (HDT). As compared with all strategies under investigation, 8 strategies were identified as favourable. Strategies generated by the random combination of management elements which does not make sense from a logical point of view had been identified by the method of assessment. To show the results of the developed method three new concepts are introduced: (1) some kind of algebraic combination of partially ordered sets, (2) the term of antagonistic indicators and (3) first attempts of application of tournament theory. The results should be helpful to enable especially decision makers, stakeholders, politicians etc. to come to competent and comprehensible decisions which further development can be declared as sustainable under the previous circumstances.

Key words: Hasse Diagram Technique, Sustainable Development, Tournament Theory

1 Introduction

*Multidisciplinary re-
search project*

Especially in urban areas the economical needs of water management causes serious problems such as the availability of fresh water and the nutrification and toxification of aquatic ecosystems. For the economic activities in cities sustainability is getting more and more demanded (*Umweltbundesamt 1997*). Up to now there is no uniform definition, what sustainability exactly means and how to assess, if different possibilities of development will be sustainable or not. For that reason a multidisciplinary research project was founded. The project is financially supported by the DBU (German Foundation of Environment) for a period of three years. In June of 1999 the results of the first years work could be presented, which has to be seen as an example and a test, whether the chosen method of assessment works in principle.

*Development of a
method to assess the
sustainability of water
management strategies*

The team working together in the project consists of members of the Institute of Freshwater Ecology and Inland Fisheries, members of the Technical University of Berlin and engineers of a commercial Office for Environmental Planning. For the sake of brevity in this manuscript the project group is called PSW (project sustainable water management). The aim of the project is to develop a method to assess the sustainability of different water management strategies. This method should be general enough to be transformable to other urban zones. So the method has to be understood as a guide, how to check for sustainable development. The main part of the comparative evaluation is done by HDT.

2 Area under Investigation

*Catchment area of the
small lowland river Wuhle*

*Wastewater treatment plant
"Falkenberg"*

*Sewage content in the river
causes serious ecological
problems*

To get some experience with the method of assessment the catchment area of the small lowland river Wuhle in the north-east of Berlin was chosen (Fig.1). In the area there is mostly urban housing with a view allotment gardens and public parks close to the river. The sewage and wastewater of about 540.000 inhabitants is disposed to the wastewater treatment plant (wwtp) "Falkenberg". The release of the plant is 96.000m³/d in average (1997) (*Müller et al. 1998*). The purified water flows into the river "Wuhle" which itself is a tributary of the river Spree. The amount of purified water from the plant is about ten times higher (1,1m³/s)¹ than the water flow of the so called "Alte Wuhle" (0,1 m³/s)¹ upstream the point, where the canal flows into the river. Down the river there are a couple of outlets of drain channels coming from the streets. The sewage content in the river causes serious ecological problems, as the river Wuhle is polluted with nutrients and toxicants (*Körner, 1995*).

¹ Average of the year 1997

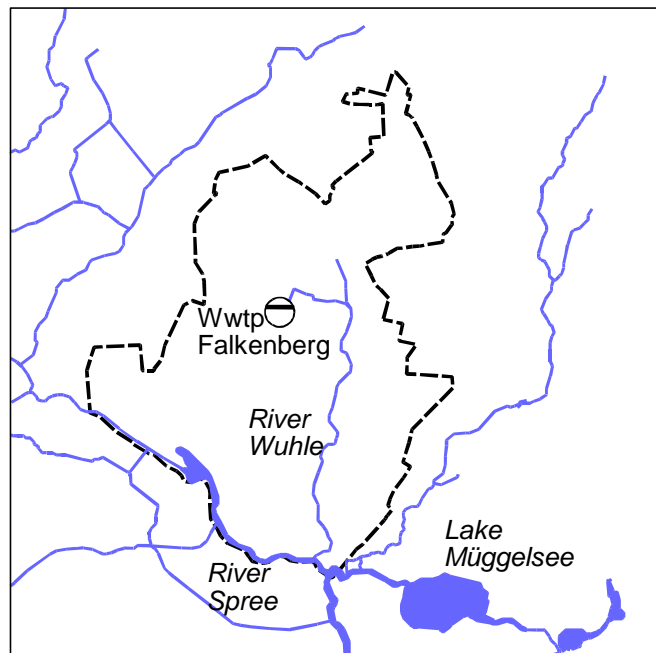


Figure 1: Area under investigation

It is planned to close up the wwtp Falkenberg. In this case the sewage and wastewater would be pumped to other purification plants. From an ecological point of view this would be good on one side, because a decrease of the concentration of nutrients and toxicants in the river Wuhle could be expected. On the other side a new problem will arise. During long periods without rain, there will a high probability that there will be no water left in the river. The little water coming from the Alte Wuhle during the summer would not be enough to keep the river flowing constantly. Therefore there is a conflict in the protection aims, which arises typically if sustainability has to be considered.

To solve these problems 40 different management strategies had been checked for their sustainability. These management strategies and the method of assessment will be presented in the following.

3 Management Strategies

3.1 Description

Each management strategy (MS) which has been assessed is consisting of three management elements (ME), each selected from one of three different management groups (MG), concerning:

1. The sewage and wastewater purification: five elements (= **T**-Group, for technical advice)
2. The rainwater treatment: four elements (= **R**-Group)

Conflict in the protection aims

Management groups

3. River morphology (from an ecological point of view): two elements (= **M**-Group).

Management elements

Each of the three groups contains some so called "management elements" (ME). These are different (technical) advice or possibilities, which had been determined by experts of the field in question and which are discussed in more detail as follows:

Management Elements for Sewage and Wastewater Purification (Technical elements)

- T1** The present state of the art: The wwtp Falkenberg. The wastewater is getting purified mechanical and biological. Phosphorus is getting precipitated simultaneously. Nitrification occurs not all of the time. The sludge is used to produce sewage gas. Because of the old technical equipment of the plant, there is still a high amount of inorganic nitrogen released to the outflow of the wwtp.
- T2** Close up of the wwtp Falkenberg. The disposed wastewater would be pumped to the wwtp's Schönerlinde and Waßmannsdorf. Both plants provide new technical equipment.
- T3** Rebuilding of the wwtp Falkenberg. New mechanical, biological and chemical purification technique. The sludge is used to produce sewage gas for heating.
- T4** Separation of urine and faeces in 50% of all households. Faeces and biological waste can be used to get gas. Afterwards the well rotted compost can be used as fertilizer. Urine can be purified by macrophytes. The purified water is disposed to the drain system. The residual wastewater is getting pumped to the wwtp's Schönerlinde and Waßmannsdorf.
- T5** Separation only of urine in 50% of all households. The urine is collected in tanks. The consisting nutrients can be used as fertilizer. The residual sewage is disposed to the wwtp Falkenberg.

Management Elements for Rainwater Treatment (Rainwater treatment elements)

- R1** The present state of the art: There are a couple of drain water outlets along the river. The water is polluted with nutrients, heavy metals and organic chemicals.
- R2** Rainwater is collected in a drainage system, consisting of hollows and drain pipes.
- R3** A cheaper version of the drainage system as described before.
- R4** Construction of more overflow basins along the river.

Management Elements for the River Morphology (Morphology elements)

- M1** The present state: The river Wuhle shaped as a technical drain channel.
- M2** Restoration of the river Wuhle as far as possible.

3.2 Algebraic Combination

Creation of different management strategies

To create different management strategies, always one element of each of the three management groups had been mixed together randomly. As a formula one can write:

$$MS = (ME \in T) + (ME \in R) + (ME \in M) \quad (1)$$

So every possible combination of management elements had been created. Let be T_i one of the five technical elements, R_j one of the four rainwater treatment elements and M_k one of the two morphology elements, then a specific management strategy MS_r would be written as: $MS_r = (T_i, R_j, M_k)$.

I.e. the "+" sign in formula (1) means simply a concatenation of elements of management elements of different groups.

Altogether there had been five elements of sewage purification multiplied by four elements of rainwater treatment multiplied by two elements of river morphology equals 40 management strategies to assess.

Advantages coming up by the random combination of management elements

There are some advantages coming up by the random combination of all management elements:

- The combination can be done automatically. This saves time, especially if there a large data sets.
- The random combination of all management elements can lead to new strategies.
- Because of the random combination of all management elements there will be some strategies, which does not make sense from a scientific or technical point of view. These "senseless" strategies can be used to test the functioning of the method of assessment.
- It is not necessary to decide in advance which management strategy will be assessed or not. This too saves a lot of work, especially if there a large numbers of strategies.
- If not all of the possible strategies will be assessed, one had to explain why some of the strategies are out of question. This again would be a lot of work.
- If some strategies are selected by men, there is always the danger, that the decision is made because of political or economical reasons. These subjective selections would casts doubt on the method of assessment.

Nonsense results Trivial results

Summarising: We take the full power of computers to generate management strategies, even at the cost that there are some nonsense or trivial results. Nonsense results however have to be automatically filtered out by the ranking method. Trivial results are considered as a confirmation that the method does not generate unexpected results; if however the "+"-sign in formula (1) is the expression of concatenation under additional constraints (correct order of actions, correct selection of sites, etc.) then the resulting management strategies are no more easy to be grasped.

4 Method of Assessment

4.1 Indicators

Set of 15 indicators to assess sustainability

In literature an enormous number of indicators is getting discussed (*OECD 1994, Walz 1997*) but as far as we know, there does not exist a generally valid set of indicators to assess sustainability. In the project up to now we used a set of 15 indicators to assess sustainability concerning:

1. the protection of the rivers ecosystem
2. the ground water supply and
3. social and economic aspects.

Table 1 shows the indicators used to assess the management strategies. The table shows as well the orientation of these indicators. This orientation is naturally according to the goal of protection. It has to be said, that these set has to be developed while progressing the project.

Table 1: Used Indicators and their orientation

Indicators concerning the protection of the ecosystem of the river Wuhle		
Short cut	Indicator	Orientation of evaluation (toward good)
Qf	vary between low and high water	small variations between high and low water
BOD	biological oxygen demand (BOD)	low BOD
P	entire concentration of phosphorus	low phosphorus content in the water
N	entire inorganic nitrogen (concentrations)	low inorganic nitrogen content in the water
Tox	contamination by inorganic and organic toxicants (concentration)	no contamination of the river by toxicants
Temp	water temperature	no heat up of the river Wuhle
Flo	lowest environmental necessary flow	constant flow during the hole year
Mor	river morphology	natural morphology of the river bed
FL	use of flood land	immediate flood in case of high water
Indicators concerning the protection of the supply of groundwater		
Short cut	Indicator	Orientation for a good evaluation
Gw	danger of groundwater contamination by nutrients and toxicants	no contamination of ground water by nutrients and toxicants
Fw	consumption of fresh water	cut down the use of fresh water
Indicators concerning other goals of protection		
E	consumption of energy	reduction of the consumption of energy
DF	damage caused by floods	no damage on buildings or farmland
DGw	damage caused by ground water	no damage on buildings or farmland
P	transfer of problems	here we took into account whether the problems are getting solved within the area under investigation or not. If some of the problems are getting shifted into another area, one had to assess this area as well to make a decision, whether the strategy is bad or good. So this indicator gives a warning, that there are further investigations necessary.

4.2 Evaluation of the management strategies

Classification of three values to evaluate management elements

At first each of the 15 indicators had been assessed for every management element separately. A classification of three values to evaluate management elements had been used ²:

- 0 = small values, with the orientation in mind, directed toward a good evaluation.
- 1 = middle
- 2 = high values, with the orientation in mind, directed toward a bad evaluation.

Scoring expresses a relative measure

It has to be said that the (scoring just expresses a relative measure). Table 2 gives an example: in the group of sewage and wastewater purification the element with the lowest inorganic nitrogen emission was given the 0 (good) and the element with the highest emission was given the 2 (bad). The residual elements had been classified as 1 (middle).

Table 2: Example of the evaluation of the management strategies.

Management element	T1	T2	T3	T4	T5
Inorganic nitrogen, calculated concentration [mg/l]	25,00	-	15,00	6,25	14,4
Value for evaluation	2	0	1	0	1

In the case of rainwater treatment, by now there were no data available. Therefore the evaluation had to be done by the knowledge of experts. Because of the rough classification by only three scores, the procedure does not cause serious problems. For the aim of the project - the development of an assessment method of generally use - this is a very important point. In most cases one can not expect to get real (measured or model supported) data. As the HDT is based on an ordinal concept of data, there is enough flexibility to integrate such qualitative statements.

165 management elements evaluated by knowledge of experts or by computed models

For the entire procedure 5 T-elements multiplied by 15 indicators plus 4 R-elements multiplied by 15 indicators plus 2 M-elements multiplied by 15 indicators = 165 management elements had to be evaluated by the knowledge of experts or, if available, by computed models. The combination of the management elements as described above leads to a so called start matrix, which contents 40 management strategies (MS) multiplied by 3 elements multiplied by 15 indicators for each of the three management groups, leading to 1800 entries. In this start matrix all values are independent from each other (Table 3).

Start matrix where all values are independent from each other

² For water quality indicators this does not necessarily mean that the status of the river obeys some normative or legislative constraints.

Table 3: Structure of the start matrix.

<i>Indicators 1-15/ MS 1-40</i>	I_1			I_2		
MS ₁	I ₁ (ME _T)	I ₁ (ME _R)	I ₁ (ME _M)	I ₂ (ME _T)	I ₂ (ME _R)	I ₂ (ME _M)	...
...

Mathematical rules to generate a resulting indicator tuple of the management strategies

The single elements within a management strategy may influence each other – there could be synergetic, redundant or antagonistic effects. Therefore one has to introduce some mathematical rules to, generate a resulting indicator tuple of the management strategies, defined by formula (1). Taking into account the qualitative character of the indicators the rules have to be accordingly simple:

1. Indicators related to concentrations are calculated according to the mixing of the flows.
2. Other indicators were determined following an optimum principle (for details see the interim-report of the project group (*Wasserforschungs e.V. 1998*)).

Evaluation matrix

By that procedure the start matrix has been transformed to an evaluation matrix (Table 4). Any further processing was done by the Hasse software WHASSE. For further information about HDT see for example *Brüggemann (1998)* or *Brüggemann et al. (1999)*.

Table 4: Structure of the evaluation matrix. f_{1-15} represents the rules to combine the indicators I_{1-15} of each ME to one indicator I'_{1-15} of the MS; for other explanation: see table 3.

<i>Indicators 1-15/ MS 1-40</i>	I_1	I_2
MS1	$f_1(I_1(ME_T), I_1(ME_R), I_1(ME_M))$	$f_2(I_2(ME_T) I_2(ME_R) I_2(ME_M))$...
...

4.3 Results by HDT

Resulting Hasse diagram

Figure 2 shows the resulting Hasse diagram. The first number in a circle stands for management elements of sewage purification (T₁-T₅), the second number represents the management elements of rainwater treatment (R₁-R₄) and the last number stands for the rivers morphology (M₁-M₂). The letters T, R and M are not shown in the diagram.

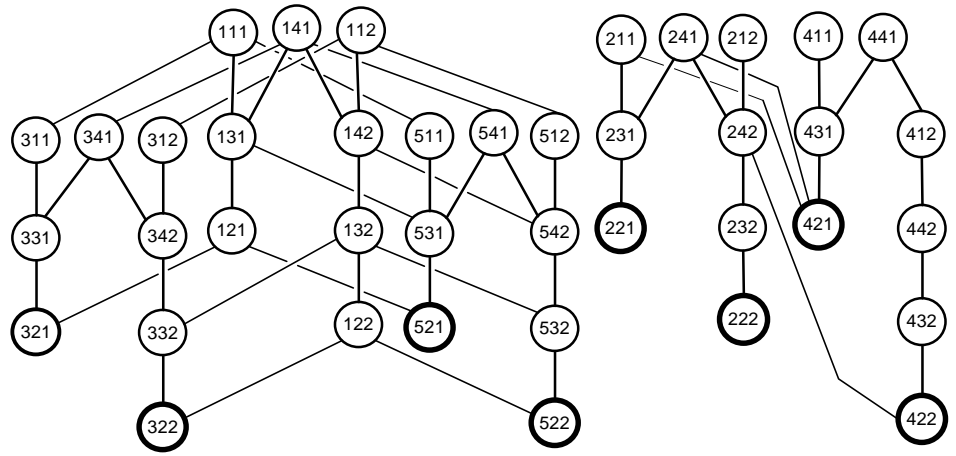
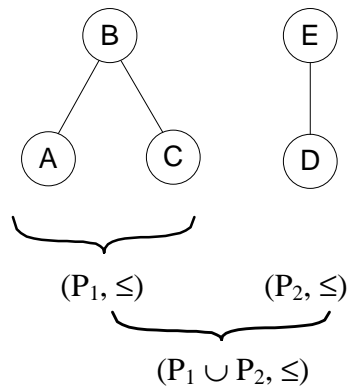


Figure 2: Hasse diagram of the 40 assessed management strategies.

Hasse diagram formulated as an algebra

The Hasse diagram and the poset resp. can be formulated as an algebra: Formally we can use \oplus as addition and \otimes as multiplication. For example the HD



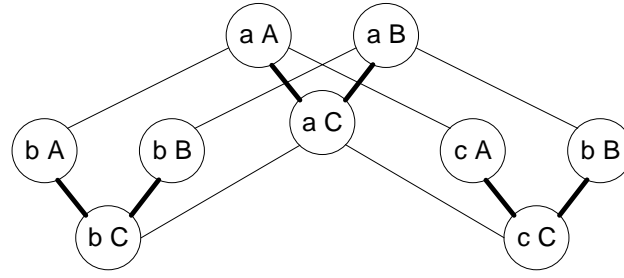
can be written as:

$$(P_1 \cup P_2, \leq) = (P_1, \leq) \oplus (P_2, \leq)$$

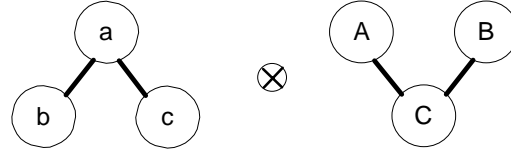
By the sign \oplus the fact can be expressed, that several hierarchies are present within a Hasse diagram. Because posets are represented by Hasse diagrams we write (rather suggestively):

$$HD_1 \oplus HD_2.$$

Similarly the HD



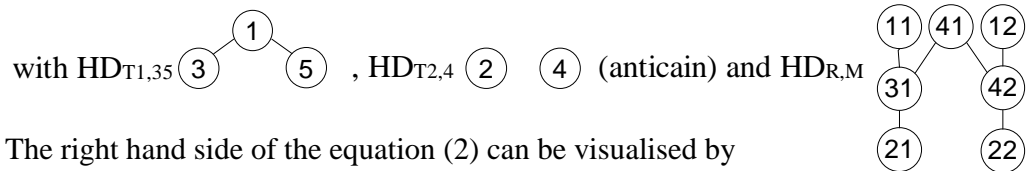
can be thought of as built up by



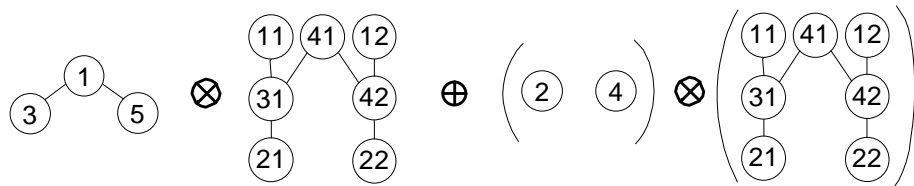
(see Davey and Priestley, 1990).

The Hasse-Diagram of Fig.2 expresses that the objects were factorially combined. In that sense the HD of the 40 management strategies can be thought of as being generated by substructures as shown in Fig.2. The corresponding formula is:

$$HD_{\text{real}} \approx (HD_{T1,3,5} \otimes HD_{R,M}) \oplus (HD_{T2,4} \otimes HD_{R,M}) \quad (2)$$



The right hand side of the equation (2) can be visualised by



A deviation is only found for the part $HD_{T2,4} \otimes HD_{R,M}$ which shows the increasing amount of interactions among the different ME, if the wwtp is closed.

Such formalism is of great help:

1. to find a structure in the "jungle" of lines
2. to constitute new theses, for example: The term in the first brackets of equation (2) expresses the combination of different T-elements with always the same pattern of R and M-elements. This shows, that the T-, R- and M-elements does not influence each other. The reason is the dominance of the release of the wwtp. In case of the term in the sec-

ond brackets the T- and R-,M-elements do influence each other, as the substructures are not identical. Here the reason is the very low release of purified water from the wwtp, which is nearly as high as the runoff of the Alte Wuhle. Therefore the mixing rule becomes important.

A further partitioning of the structure would be possible but is not shown here.

Eight minimal elements

Eight of the 40 management strategies under investigation are minimal elements. That means, compared to all the other strategies these eight are to be favoured. Two of the minimal elements are comparable to the present state (T_1, R_1, M_1). These are (T_3, R_2, M_1) and (T_5, R_2, M_1). In every of the 15 indicators, these two management strategies are evaluated better than the present state. However the technical realisation will be expensive because either a reconstruction or the utopian urine separation is implied.

4.4 Antagonistic Indicators

Term of antagonistic indicators to explain the incomparability in the Hasse diagram

The Hasse diagram (Fig. 2) shows two main parts, which are not comparable with each other. As said above, the difference is caused by the management elements of the wastewater treatment. These are the strategies with the wwtp Falkenberg (T_1, T_3, T_5) and without it (T_2, T_4). To explain the differences between the two main groups of the diagram the term of antagonistic indicators is introduced. In the current case it is on one hand the quality of the water and on the other hand it is the warning that problems are getting transferred to other areas:

In the left group there will be still the wwtp under operation. And for that reason the pollution of the river with nutrients and toxicants is always expected to be higher than without a wwtp releases its water into the Wuhle. In case of closing the purification plant and pumping the water somewhere else (group on the right hand side) one has to examine the affected area in question to decide, whether the strategy is better or worse than the current situation. For these reasons, from a scientific point of view, the two big groups of the Hasse-Diagram are not comparable to each other.

Structure of the diagram

Paying attention to the structure of the diagram one can see, that:

1. The management groups rainwater treatment (R) and river morphology (M) are responsible for the ranking within one substructure, independent of the kind of wastewater purification.
2. Looking at the substructures of the diagram one can see, that each of it consists of two branches. The separation of the two branches is caused by the river morphology. On the left branch there are always the management strategies with the present state of affairs (M_1). The right branch is built by strategies with river restoration (M_2)
3. The kind of wastewater treatment causes the origin of different substructures within the Hasse diagram.
4. The different kind of management elements does not influence each other very much.

Sensitive indicators

The most sensitive indicators (see the Introduction, this issue) are those which describe:

- The variations between low and high water (Qf),
- the consumption of fresh water (Fw) and
- damage caused by flood (DF).

Eight favourable strategies

Eight of the 40 assessed strategies are minimal objects, and therefore favourable strategies. These strategies consist of the different elements of wastewater purification T_2 , T_3 , T_4 and T_5 , always combined with the rainwater treatment element R_2 and the present shape of the river bed (M_1) or the restoration of the Wuhle (M_2). So the question of river restoration or not doubles the number of minimal elements. This is caused by the pair of the antagonistic indicators river morphology (Mor) / danger of flood (DF). The reason is easy to understand. From a ecological point of view a river restoration is demanded. But without any further precautions, that would cause a high possibility of floods and the danger, that peoples properties will get damaged. Thus, in the frame of sustainable development, where social-economies is one of the dimensions of discussion, the antagonism becomes evident.

In the main group with the wwtp on full or reduced operation, the rebuilding of the plant (T_3) and separation of urine in 50% of all households (T_5) are favourable strategies. Both solutions are not comparable because of the indicators consumption of fresh water (Fw) and consumption of energy (E). In case of rebuilding the plant the consumption of energy would be cut down because of the new technical equipment. In case of separation of urine the consumption of freshwater would be cut down because of the safe of water to flush the toilet.

In the second main group without the wwtp Falkenberg the two remaining favourable strategies (T_2 and T_4) are incomparable because of the indicators consumption of energy (E) and freshwater (Fw) and release of phosphorus (P). In case of closing up the plant Falkenberg and pumping the sewage to other plants there will be a higher consumption of energy and fresh water than in case of separation of urine and faeces and pumping only the residual sewage to other plants.

Two management strategies are comparable with the present state

Only two management strategies, $T_3R_2M_1$ and $T_5R_2M_1$, (T_3 = rebuilding of the wwpt, T_5 = separation of urine, R_2 = draining of rainwater in hollows and drain pipes, M_1 = present shape of the river bed) are comparable with the present state (T_1 , R_1 and M_1). All indicators of these management strategies are evaluated as better.

Summarising: There is – up to now and taking in mind the crude estimation of the indicators – no best strategy but several optimal ones. This fact makes evident, that advantages in some aspects are at the costs of disadvantages in other aspects.

Because of the random combination of the management elements there had been some strategies, which does not make sense from a logical point of view. This is the T_1 , R_4 , M_2 strategy for example. In case of keeping the present plant running plus building more overflow basins for

Useless strategies are identified by the method of assessment

rainwater and additionally restore the riverbed it would cause floods every time it rains. This useless strategy was identified by the method of assessment and labelled as a nearly maximal element, which means, this is one of the worse strategies.

5 Implication of Tournament Theory

5.1 Method

Tournament Theory to obtain a further ranking

Sometimes decision makers might feel uncomfortable about having a couple of favourable but incomparable management strategies as a result, which is typical for HDT. To solve this problem, one could apply the elegant approach of *Sørensen et al. (1998)*. This approach is based on Monte Carlo simulations and calculating a probability of ranks based on a given confidence level. Taking into account, however, the data quality (roughly estimated qualitative data combined with rather sharp data calculated for water quality parameters), we felt that such an approach is only feasible in later stages of the project. Therefore, we select Tournament Theory (TT) to obtain a further ranking (see *Clark, J., 1994 and Bartel, 1996*). *Bondy et al. (1976)* describes TT from the algorithmic point of view as the following: It solves the problem, how the participants in a tennis tournament can be ranked, when we have a number of players each playing one another.

Ranking by computing the scores and compare them

The ranking is obtained by computing the scores, which are the numbers of games won by one player and compare them. If the first level score (s_1) does not give a clear ranking (there might be players with the same number of wins) one have to look at the second-level score and so on (s_2, s_3, \dots, s_n), until there is a definite ranking. For further information about TT see *Bondy et al. (1976)*.

Decision of win and loose by the evaluation of the indicators

In our case the 8 favourable management strategies can be seen as the players in a tournament. The decision which wins and which looses is made by the evaluation of the indicators. For example: Comparing two strategies, the one with the more indicators evaluated as good, wins (Table. 5).

Table 5: Competition of strategies by the evaluation of the indicators.

Strategy	→ better	equal	worse	→ Strategy
Strategy	← worse	equal	better	← Strategy
221	3	8	4	222 (winer)
(winner) 221	7	4	4	321
221	6	3	6	422 (draw)
...				...

Tournament depicted as a graph

The tournament can be depicted as a graph (Fig. 3). The arrows always leads from the winner to the loser strategy (smaller number of positively evaluated indicators). Games ending in a draw are depicted with a broken line.

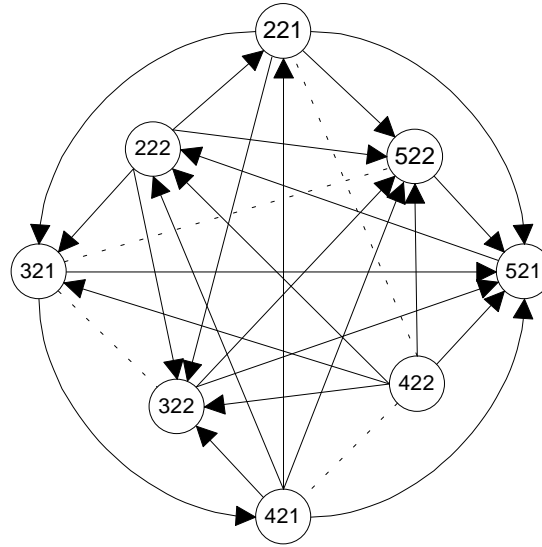


Figure 3: Graph of the 8 favourable strategies

Tournaments ending in a draw

An important demand of TT is that there always has to be a winner and a loser. For that reason in case of the four tournaments ending in a draw, every possible combination of win and loose has been created. By that procedure one obtains 16 tournaments ($= 4^2$, according to 4 draws) with a random distribution of winner-looser combinations.

Another important demand of TT is that the graph has to be connected in a way that it is possible to reach every node of the graph from every other node. Looking at the 16 tournaments one notice, that in all cases strategy 422 is set as the winner of the drawn games, it does not obey the rule mentioned above. In these cases, one has to eliminate 422 and set it first (because it wins all tournament anyway). For the 7 remaining strategies the score ranking has to be computed by an algorithm, as described in the following.

Ranking of the score-levels

Each tournament can be written as a matrix (so called adjacency matrix of the directed graph A, (Fig.3)), where 1 stands for a win and 0 for a loose. To compute the ranking of the score-levels until a definite ranking appears, the following recursive formula holds:

$$\vec{S}(t) = A\vec{S}(t-1) \quad (3)$$

where $\vec{S}(t); \vec{S}(t-1)$ = scores after iteration step t and t-1 resp. and A = adjacency-matrix of the tournament.

Therefore calculating a definite ranking without ties can be done by repeating the recursive scheme (3). This however is equivalent to calculate the eigenvalue of the matrix A.

5.2 Results

Rankings of the tournaments Table 6 shows the rankings of the 16 tournaments when calculated as described above.

Table 6: Rankings of the 16 tournaments.

T1:	422 > 421 > 222 > 221 > 321 > 322 > 522 > 521
T2:	422 > 421 > 221 > 222 > 321 > 322 > 521 > 522
T3:	422 > 421 > 222 > 221 > 321 > 522 > 521 > 322
T4:	422 > 421 > 222 > 221 > 321 > 322 > 521 > 522
T5:	421 > 422 > 222 > 221 > 321 > 322 > 522 > 521
T6:	422 > 421 > 221 > 222 > 321 > 522 > 521 > 322
T7:	422 > 421 > 222 > 321 > 221 > 521 > 322 > 522
T8:	421 > 422 > 222 > 321 > 221 > 322 > 521 > 522
T9:	422 > 221 > 421 > 222 > 321 > 521 > 322 > 522
T10:	422 > 221 > 421 > 222 > 321 > 521 > 322 > 522
T11:	422 > 421 > 221 > 222 > 321 > 322 > 521 > 522
T12:	421 > 422 > 321 > 222 > 221 > 521 > 322 > 522
T13:	422 > 421 > 221 > 222 > 321 > 522 > 521 > 322
T14:	422 > 421 > 221 > 222 > 321 > 322 > 521 > 522
T15:	422 > 421 > 221 > 222 > 321 > 322 > 521 > 522
T16:	421 > 422 > 222 > 321 > 221 > 522 > 521 > 322

Probability of ranking

By setting the 16 tournaments as 100%, one can calculate very easily the probability of ranking on which place every management strategy is to be found (Fig. 4).

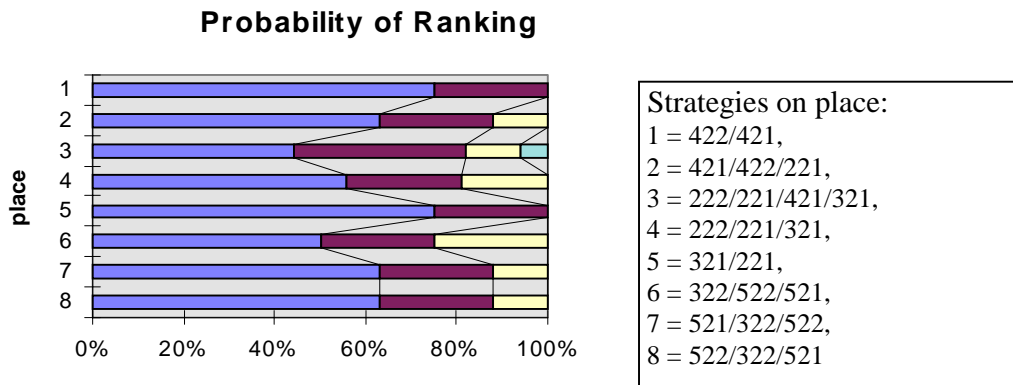


Figure 4: Probability of Ranking

Hasse diagram depicting the results of Tournament Theory

Ordering the strategies corresponding to the evaluation of their indicators

To depict the results of the 16 tournaments, one can use HDT once again. Figure 5 shows the resulting Hasse diagram depicting the results of Tournament Theory, when the ranking yielded by the 16 tournaments are set as attributes and the 8 strategies as elements. The two strategies (422 and 421) which are always at first place are shown as minimal elements. All other strategies are ranked on two lower levels. By using HDT a second time, one can clearly see, which are the strategies with the most positive evaluated indicators (minimal elements) and which are the ones with a less number of positive evaluated indicators (maximal elements).

So the application of TT leads to a further ranking just by ordering the strategies corresponding to the evaluation of their indicators. Again, no subjective evaluation or aggregation has to be done. For that reason TT might be a powerful tool to come to a further ranking, when HDT is used as the method of assessment in the first place.

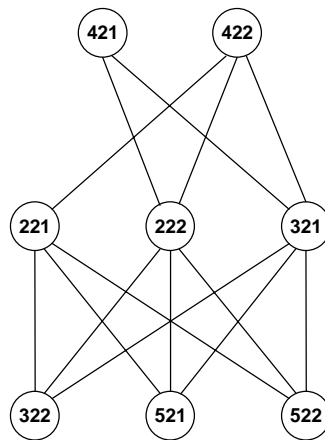


Figure 5: HD of the favourable strategies, ranked by Tournament Theory.

6 Conclusion

Method of assessment works in principle

The results of the assessment of the 40 management strategies shows, that the chosen method of assessment works in principle. The method offers different favourable management strategies and labels the antagonistic indicators, which lead to incomparabilities. The Tournament Theory can be additionally used to come to a further ranking of the 8 favourable strategies obtained by HD, without the need of numerical aggregation of data. So both mathematical methods, the HD in combination with the TT could be powerful and helpful tools to come up to decisions about sustainable development.

7 References

Bartel, H.-G. (1996): Mathematische Methoden in der Chemie, pp: 1-363. Spektrum Akademischer Verlag, Oxford.

Bondy, J.A. and Murty, U.S.R. (1976): Graph Theory With Applications. 185-186.

Brüggemann, R. (1998): Workshop on: Order Theoretical Tools in Environmental Sciences. Introduction. In Group Pragmatic Theoretical Ecology (Ed.), Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences. Berichte des IGB ;Heft 6, Sonderheft I, 1998. (pp. 1-9). Berlin: IGB.

Brüggemann, R., C. Bücherl, S. Pudenz, C. Steinberg. (1999): Application of the concept of Partial Order on Comparative Evaluation of Environmental Chemicals. Acta hydrochim hydrobiol, 27, 170-178.

Clark, J. and Holton, D.A. (1994): Graphentheorie, Grundlagen und Anwendungen, pp. 1-363. Spektrum Akademischer Verlag, Heidelberg.

Davey, B.A. and Priestley, H.A. (1990): Introductions to Lattices and Order, pp. 1-248. Cambridge University Press, Cambridge.

Körner, S. (1995): Selbstreinigungsprozesse im Klärwerksableiter Wuhle unter besonderer Berücksichtigung der submersen Makrophyten. Dissertation, Humboldt-Universität, Berlin.

Müller et al. (1998): Abschätzung der Umwelterheblichkeit der Einstellung der Klärwerksableitung auf das Seelgraben-Wuhlesystem. 1-112, 6 Beilagen, Berlin (unveröffentlicht).

OECD (1994): Environmental indicators - core set, Paris

Sørensen, P.B., Mogensen, B.B., Gyldenkaerne, S. and Rasmussen, A.G., (1998): Pesticide Leaching Assessment Method for Ranking both Single Substances and Scenarios of Multiple Substance Use. Chemosphere, 36, 2251-2276.

Sørensen, P.B., Mogensen, B.B., Carlsen, L. and Thomsen, M. (1998): The role of uncertainty in Hasse diagram ranking. In: Proceedings of the Workshop on Order Theoretical Tools in Environmental Science; Berichte des IGB; Heft 6, Sonderheft I, 1998, pp. 71-84. Editor: Group Pragmatic Theoretical Ecology. IGB, Berlin.

Umweltbundesamt (1997): Nachhaltiges Deutschland - Wege zu einer dauerhaften umweltgerechten Entwicklung. Erich Schmidt Verlag, Berlin.

Walz, R. (1997): Grundlagen für ein nationales Indikatorensystem - Weiterentwicklung von Indikatorensystemen für die Umweltberichterstattung, pp.1-470, Umweltbundesamt, Uba-FB 97-022, Berlin.

Wasserforschungs e.V. (Hrsg.) (1998): Nachhaltige Wasserwirtschaft, Entwicklung eines Bewertungs- und Prüfsystems. Bericht zur Pilotphase des Forschungsvorhabens (AZ 12953), Gefördert durch die Deutsche Bundesstiftung Umwelt.

Acknowledgement

We thank for financial support of the Deutsche Bundesstiftung Umwelt AZ 1295 and all members of the PSW for data support and helpful discussions.

Environmetrical Approaches to Evaluate Internet Databases

-Hasse Diagram Technique / Multivariate Statistical Methods-

Kristina Voigt, Gerhard Welzl, Gerda Rediske

*GSF - National Research Center for Environment and Health
Institute of Biomathematics and Biometry
85764 Neuherberg, Germany
kvoigt@gsf.de, welzl@gsf.de*

Key words:

Hasse diagram technique, partial order, multivariate statistical methods, internet database, Bertin-strategy

Abstract

Constantly expanding chemical and environmental information sources increase the need for data analysis. This paper presents strategies to evaluate Internet databases by different mathematical and statistical approaches. The Hasse diagram technique, a method based on discrete mathematics and multivariate statistical methods are applied and the results compared and discussed. As an example a data-matrix of 21 objects (pesticide Internet resources) and 5 descriptors (evaluation criteria for environmental and chemical data-sources) is analyzed. The main focus of this paper lies on the statistical analysis and on the comparison of the results with those of the Hasse diagram technique. Application of Bertin-strategies on our data-set of 21 Internet resources results in the following ordering of descriptors: AN-SE-QI-ID-IP. The same method (Bertin-strategy) is applied to the sequence of objects (pesticide Internet resources). This procedure leads to a data-matrix of a more homogeneous structure than the original data-set. Furthermore we apply another method called POSAC (Partially Ordered Scalogram Analysis with Coordinates) to analyze the given data-set. The results of the Bertin-strategy and those of the POSAC-method are compared with the given Hasse diagram. A great correspondence is found in special regions of the multivariate statistical methods with maximal and minimal objects of the Hasse diagram.

1 Introduction

For most institutions, the existence of very large databases with important and critical information is not new and pulling out the data that is needed, when it is needed, has been an age-old challenge. This is especially valid when talking about environmental and chemical information. Hence follows that mining large data resources in chemistry and the environmental sciences will be a main task in the future. The need for applying data mining concepts is increasing due to a variety of factors:

- explosion in amount of information that is captured electronically.
- dramatic price decreases in data storage hardware
- focus on knowledge management in organizations increased pressure to share and use electronic data captured as a competitive advantage.

Data mining

Data mining can be defined as analyzing the data in large databases to identify trends, similarities, and patterns to support managerial decision making. Data mining technologies generally use algorithms and advanced statistical models to analyze data according to rules set forth by the particular application at hand. Data mining models fall into three basic categories: classification, clustering, and associations and sequencing [Zorn *et al* 1999]. According to this definition not only statistical models but also methods like the Hasse diagram technique can be regarded under the term "data mining".

Multivariate statistical methods

In this paper multivariate statistical methods will be applied to a data-set of 21 Internet pesticide resources evaluated by five different criteria. The results of this evaluation will be compared with those found by the application of the Hasse diagram technique to the same data-set. The main focus of this paper lies on the statistical analysis, as the Hasse diagram for an Internet database matrix was discussed during the 1. Workshop "Order Theoretical Tools in Environmental Sciences" in Berlin, November 16th, 1998 [Voigt 1998a]. Comparisons of multivariate statistical methods with methods of discrete mathematics (Hasse diagram technique) were carried out applying a data-matrix of 59 sites in the German state of Baden-Württemberg polluted with heavy metals and sulphur [Welzl 1998].

2 Environmetrics and Chemometrics

*Chemometrics,
environmetrics*

Internet-metrics

In chemistry and environmental sciences the terms chemometrics and environmetrics are used for scientific disciplines that apply mathematical, statistical, and other methods to provide maximum chemical or environmental information by analyzing chemical or environmental data [Einarsson 1997]. In the following sections we shall apply mathematical, statistical methods to pesticide Internet resources. Following the ideas of chemometrics, environmetrics we might talk about "database-metrics" or applied to the media "Internet-metrics".

Descriptive statistics

In statistical analysis two major statistical fields may be distinguished: descriptive and inferential statistics. Often it is indispensable to order, comprise and describe the data material before conclusions can be drawn from data-sets. Descriptive statistics - as part of chemometrics, environmetrics and biometrics - may be divided into:

1. representation in tables
2. graphical representation (e.g. histograms)
3. numerical characterization of data-sets by basic parameters (e.g. arithmetic mean, variance, standard deviation, correlation coefficient).

Inferential statistics

Inferential statistics aim at drawing right conclusions, decisions, and prognoses from the data-set available [Lorenz, 1996]. Some experts regard this distinction between descriptive and inferential statistics to be somewhat artificial, as the usefulness of results should be considered straight upon compiling the data-set [Bärlocher, 1999].

Bertin strategy

In the multivariate statistical analysis the emphasis lies in scaling or ordering variables or objects. In this respect several strategies are known. Cluster-analysis is intended to classify some objects into collective categories. The result is a partition of objects (or attributes). On the other side methods of ordination aim at ranging objects (or attributes) in some order. Ordination is a widely used technique in ecology and environmental sciences. The most well-known and established techniques are Principal Component Analysis (PCA) and Correspondence Analysis (CA). Another method based on the optimization of the sum of distances between objects (Internet resources) is known under the term Bertin strategy. The latter method will be applied in this paper.

3 Data-matrix 21 Pesticide Internet Resources, Five Evaluation Criteria

*Data-matrix 21 objects,
5 descriptors*

As an example a data-matrix of 21 objects (pesticide Internet resources) and 5 descriptors (evaluation criteria for environmental and chemical data-sources) is analyzed. The following evaluation criteria are applied: search possibilities (SE), quality of Internet resource (QI), number of chemicals (NU), identification parameters for chemicals (ID), and information parameters for chemical substances (IP). This data-matrix was published in the proceedings' volume of the 1. Hasse Workshop 1998 [Voigt 1998a].

Table 1. Scores for 21 Pesticide Internet Resources

<i>Acr.</i>	<i>Name of the Resource</i>	<i>SE</i>	<i>QI</i>	<i>NU</i>	<i>ID</i>	<i>IP</i>
AGR	AgrEvo MSDSs	0	1	0	2	2
APP	ARS Pesticides Properties Database	3	4	2	2	2
BEL	Bell Laboratories, Inc.	1	1	0	2	3
CPP	C & P Press Inc. MSDSs	3	1	4	2	2
CIA	Chemicals (Industrial/Agricultural)	1	0	1	1	1
DEM	Demise of the Dirty Dozen Pesticides	0	1	0	0	0
EXP	EXTOXNET, PIPs	3	2	1	1	4
FDA	FDA Glossary of Pesticide Chemicals	3	2	3	1	0
FMC	FMC MSDSs	0	1	0	2	2
IAP	Index of Authorised Plant Protection Products	5	4	4	0	1
ISA	ISA - Label Management System	4	1	4	1	3
LIP	LIPHATECH MSDSs	1	1	0	2	3
MEB	Methyl Bromide Phaseout Web Site	0	2	0	0	0
NOV	Novartis Crop Protection MSDSs	3	1	1	2	2
PMD	Pesticide Monitoring Database	3	2	1	0	1
ROH	Rohm and Haas MSDSs	1	1	0	2	3
SCO	Scotts Company MSDSs	0	1	1	2	2
SEP	SePRO Corporation MSDSs	0	1	0	1	2
SIN	Sinon CorpAgrochemical Department: Products	0	1	0	2	0
SUR	SureCo AllPro	1	1	0	2	2
WAC	Water and Air Contamination Limits	0	2	0	0	1

The scores run from 0 = insufficient to 5 = excellent. Details about the evaluation criteria and the corresponding scores are given by Voigt [Voigt, 1998b].

3.1 Evaluation of the Data-Matrix Applying the Hasse Diagram Technique

Hasse diagram

The corresponding Hasse diagram for the data-matrix is given in Figure 1. Again this diagram has already been published by Voigt, 1998a.

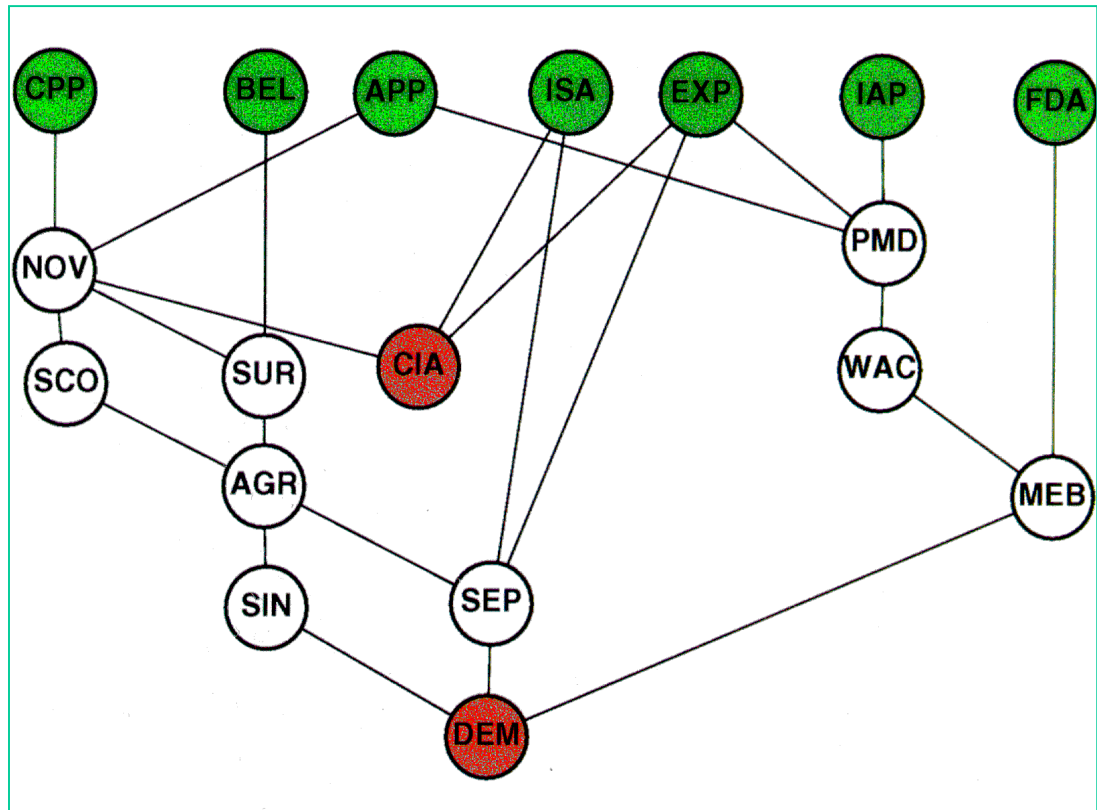


Figure 1. Hasse diagram for 21 pesticide Internet resources evaluated with 5 criteria

Maximal objects

Minimal objects

The Figure shows 7 maximal objects (APP, BEL, CPP, EXP, FDA, IAP, ISA), 2 minimal objects (CIA, DEM) and no isolated objects. Further can be stated that the diagram comprises 6 levels, the number of comparabilities is 94, whereas the number of incompatibilities is 240. Non-trivial equivalence classes are AGR/FMC, BEL/LIP/ROH.

This Hasse diagram technique compares and evaluates the Internet resources by ranking them according to their importance (good = maximal objects, bad = minimal objects). We can regard the Hasse approach as a data mining tool for environmental information and knowledge management.

3.2 Evaluation of the Data-Matrix Applying Multivariate Statistical Methods

Bertin-strategy

In the multivariate statistical analysis the scaling and ordering of objects or variables is envisaged. As mentioned above the Bertin-strategy is applied. This method is based on the optimization of the sum of distances between objects (or variables). This is an optimization problem ($N!/2$ different orders) which cannot have one exact solution. For solving this

problem - also known as Traveling Salesmen Problem - several approximate algorithms are applied, e.g. the methods of simulated annealing [Lawler, 1985].

First we order the variables by minimization the sum of distances between neighboring variables, then the ordering of objects applying the Bertin-strategy (minimization the sum of distances between neighboring objects) is performed.

3.2.1 Ordering of variables

The results of the ordering of the evaluation *criteria* is the following: NU-SE-QI-ID-IP. That means that the reverse sequence (IP-ID-QI-SE-NU) is also valid as only the distances between the variables are taken into consideration. Figure 2 illustrates the correlation between the evaluation criteria number of chemicals (NU) and search possibilities (SE). That indicates that NU and SE are similar and hence follows that they are neighbors in the ordering of the evaluation criteria.

Ordering of the evaluation criteria

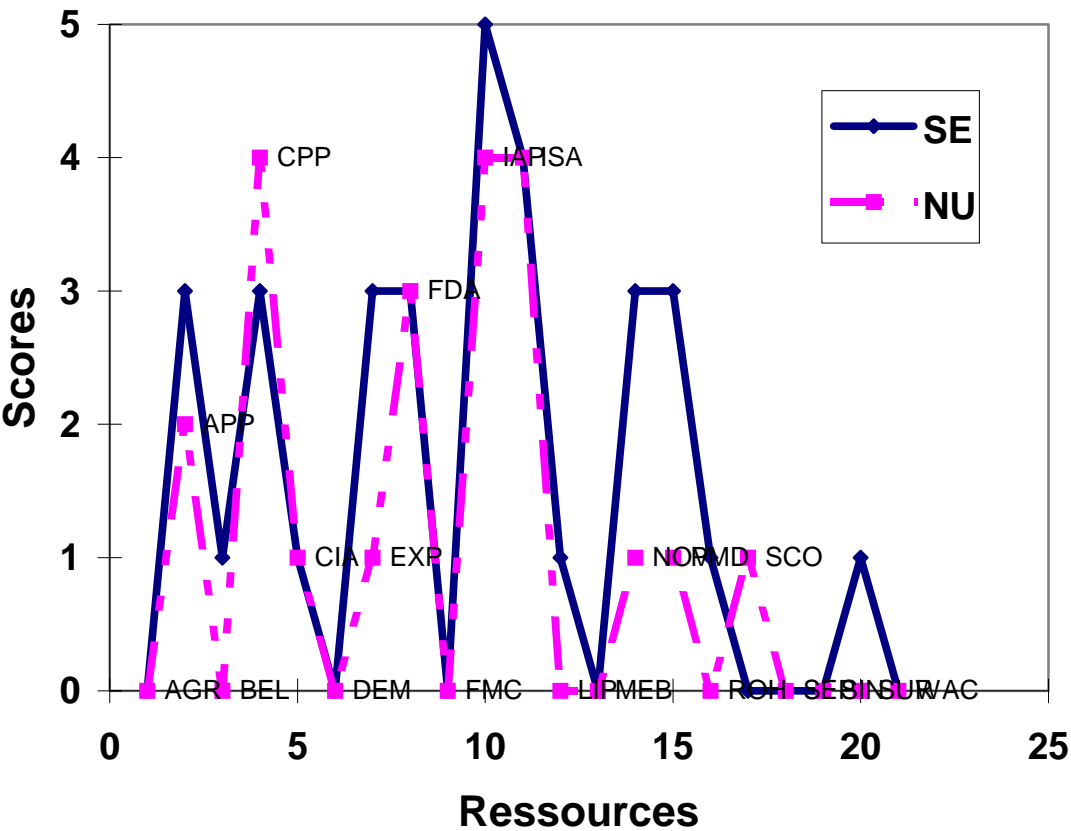


Figure 2. Correlation between NU and SE

W-matrix

The results given by the analysis of the W-matrix of the Hasse diagram technique, which describes the influence of the evaluation criteria on the ranking process, is as follows:

QI-ID-IP-NU-SE

QI-ID-IP-NU-SE.

The omission of the criterion QI leads to the highest number of changes in the Hasse diagram, whereas the omission of SE to the lowest number of changes. These results are published by Voigt [Voigt, 1998a].

Whereas the W-matrix focuses on the importance of variables in the ranking process, the Bertin-strategy applied here looks upon the similarities between two variables.

3.2.2 Analyzing the objects

The Bertin-strategy is applied to the data-matrix given. The minimization of the sum of distances between neighboring objects leads to a data-matrix of a more homogeneous structure as the original data-matrix. This is demonstrated in Figure 3. Several groups or classes of objects can be detected.

Minimization of the sum of distances

Group 1

On the left hand side of the diagram we can see **group 1** with scores 0,1 for the attributes NU, SE, QI and on the other hand scores 1,2 for the attributes ID and IP. The Internet pesticide resources CIA, SCO, SEP, LIP, ROH, BEL, SUR, AGR and FMC are found in this group.

Group 2

The next **group 2** is described by low scores not only for the attributes NU, SE and QI but also for the criteria ID, IP. The objects DEM, MEB, and WAC belong to this group.

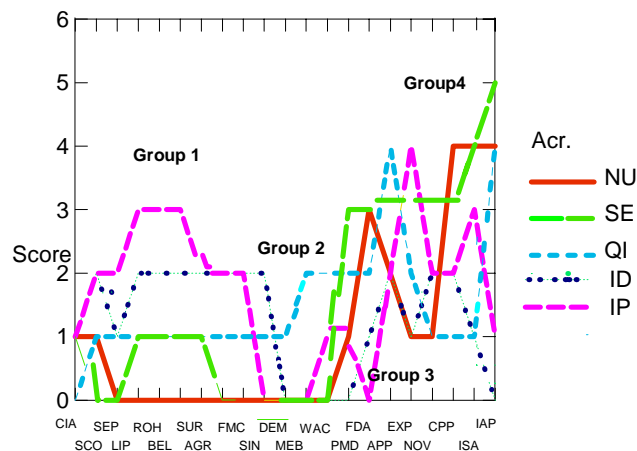


Figure 3. Parallel Coordinate Plot of Data-Matrix 21x5

Group 3

Group 3 comprises the objects PMD and FDA, which have scores of 1-3 for the attributes NU, SE, QI and scores 0,1 for the attributes ID and IP.

Group 4

Group 4 indicates objects, which show scores of 2-4 not only for the attributes NU, SE, and QI but also for ID and IP. These objects are given on the right hand side of the diagram (APP, EXP, NOV, CPP and ISA).

Two factors NU,
SE, QI/ID, IP

On the basis of this ordering procedure two factors are detected. One is characterized by the evaluation criteria NU, SE, QI and the second one by ID and IP. The combinations of two categories (high/low) of two factors leads to the four groups described above.

Comparing figure 3 with
the Hasse diagram
figure 1

Comparing this figure with the Hasse diagram in Figure 1 we can state that the group 4 corresponds in the objects APP, EXP, CPP and ISA with the maximals of the Hassediagram. The object NOV is found in the second highest level of the Hasse diagram. Group 2 corresponds in the object DEM with the minimals of the Hasse diagram. MEB is in the second lowest level of the Hasse diagram, whereas WAC is in the middle level position.

We now apply another method which can be regarded as a multivariate statistical tool as well as an approach in direction of the a discrete mathematical tool. This is called the POSAC-method.

3.3 Evaluation of the Data-Matrix Applying the POSAC-Method

POSAC

POSAC stands for Partially Ordered Scalogram Analysis with Coordinates. The POSAC module is found in the program package Systat 9. The POSAC module calculates a partial order scalogram analysis on a set of multicategory items.

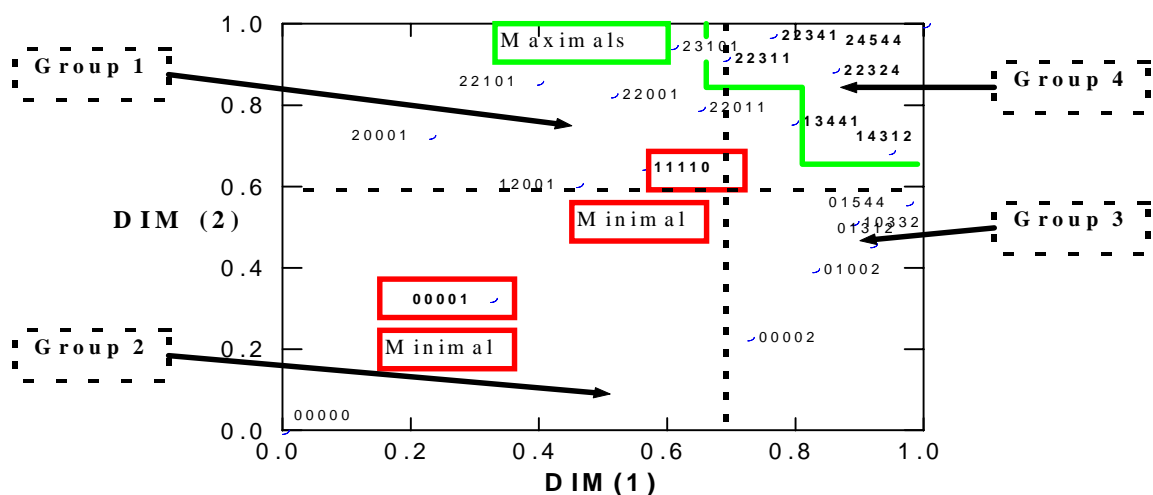


Figure 4. POSAC Profile Plot of Data-Matrix 21x5

Note: The sorting for the POSAC file is as follows: ID-IP-SE-NU-QI

2 DIM (Dimensions)

In this POSAC plot the 2 DIM (Dimensions) 1 and 2 represent two factors. The plot shows that DIM 1 is characterized by the attributes SE, NU, QI whereas DIM 2 is characterized by ID and IP. The four groups

found by the application of the Bertin-strategy to the data-matrix can easily be detected in Figure 4.

Comparing the POSAC plot with the Hasse diagram

Comparing the POSAC plot with the Hasse diagram (Figure 1) the following results can be drawn:

The POSAC method underlines the maximal objects found in the Hasse diagram. These are found at the right upper side of the POSAC plot. One of the minimals, the object DEM is also found in the lower left area of the POSAC profile plot. The other minimal CIA is situated in the middle of the profile plot.

4 Conclusions and Outlook

Multivariate statistical methods should be compared with Hasse diagram technique

Cooperational approaches

Analyzing environmental and chemical data-sets by different mathematical and statistical methods are useful and important approaches in environmental and chemometrics. They can be regarded as information and knowledge management tools in environmental sciences and chemistry. Such approaches are urgently needed in today's world of information and data overflow. Different techniques such as multivariate statistical methods e.g. the applied Bertin-strategy and POSAC method should be compared with methods part of discrete mathematics, e.g. Hasse diagram technique. The differences of the methods as well as their complementary aspects should be examined in depth in the future especially on environmental and chemical data-sets. First cooperational approaches are initiated with our working group Biostatistics at the Institute for Biomathematics and Biometry of the GSF - National Research Center for Environment and Health and the Institute of Fresh Water Ecology and Inland Fisheries in Berlin.

5 References

Bärlocher, F., (1999): Biostatistik, Georg Thieme Verlag, Stuttgart, pp. 206.

Einax J., Zwanziger H.W., Geiss S., (1997): Chemometrics in Environmental Analysis, VCH, Weinheim, pp. 384.

Lawler E.L., (1985): The traveling salesman problem, Wiley, Chichester.

Lorenz R. J., (1996): Grundbegriffe der Biometrie, pp. 338, Gustav Fischer Verlag, Stuttgart.

Voigt K., (1998a): Hasse-Diagram-Technique as an Effective Tool in Environmental Information Management, S. 55-66, 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, Institut für Gewässerökologie und Binnenfischerei, Berlin.

Voigt K., Brüggemann R., (1998b): Evaluation Criteria for Environmental and Chemical Databases, Online & CD-ROM Review, 22, 4, pp. 247-262.

Welzl G., Voigt K., Rediske G., (1998): Visualisation of environmental pollution - Hasse diagram technique and explorative statistical methods", in: Proceedings of the Workshop on Order Theoretical Tools in Environmental Sciences Berlin, November 16th, 1998, pp. 101-110, Institute of Fresh Water Ecology and Inland Fisheries, Berlin.

Zorn P., Emanol M., Marshall L., Panek M., (1999): Finding Needles in the Haystack: Mining Meets the Web, Online **23**, 5, 17-28.

Validation of a ranking model in relation to Danish monitoring data

Shima Dobel, Betty Bügel Mogensen and Peter B. Sørensen

*National Environmental Research Institute
Department of Environmental Chemistry
Frederiksborgvej 399
DK-4000 Roskilde
Denmark*

Abstract

The partial order ranking technique will be validated using monitoring data from Denmark. Pesticide data from surface waters in Denmark have been collected and the pesticides are ranked based on: The measured concentrations of pesticides in stream water, the limit of detection and the detection frequency. The result of this ranking is compared to a ranking model of the same pesticides, this ranking model being based on use data: The estimated recommended dose, the total sprayed area for each pesticide in Denmark and the adsorption coefficient, K_{oc} . In the Danish monitoring data there is no information about the actual dose used in the catchment areas of the streams but this dose is estimated from data on nation wide consumption. Best coincidence is found between the data on maximum concentration, detection frequency and limit of detection and the model using data on dosage and sprayed area. In this case the degree of agreement was 89%.

1 Origin of data

Data for the ranking of pesticides are collected from different counties in Denmark. The collected data consist of measured concentrations of pesticides in small streams in Denmark from 1994-1997. Concentrations are measured for thirty pesticides for a period with five to eleven measurements for each period. Data are collected for fourteen different streams in Denmark. Common for the fourteen streams is that the catchment areas of all the streams are used for agricultural purposes. One stream is on the island of Fyn, one on the island of Lolland and twelve are in Jylland (Table 1).

Table 1. Streams where the pesticides are measured.

<i>Streams</i>	<i>Abbreviation</i>	<i>Geographical site</i>	<i>Soil type of the catchment areas</i>	<i>Year</i>
Lillebækken	Lille	Funen	Clayey soil	1994-1996
Ellerup	Ell	Jutland, Århus	Clayey soil	1997
Horndrup	Horn	Jutland, Århus	Clayey soil	1997
Jaungyde	Jaun	Jutland, Århus	Clayey soil	1997
Støvlbæk-Vemb	Støvl	Jutland, Ringkøbing	Sandy soil	1996
Sunds Møllebæk	Sund	Jutland, Ringkøbing	Sandy soil	1996
Green Bæk	Green	Jutland, Ringkøbing	Sandy soil	1996
Herborg Bæk	Herb	Jutland, Ringkøbing	Loamy soil	1996
Lambæk	Lam	Jutland, Ringkøbing	Loamy soil	1996
Skødbæk	Skø	Jutland, Ringkøbing	Clayey soil	1996
Fald Å	Fald	Jutland, Ringkøbing	Clayey soil	1996
Vejrum Bæk	Vej	Jutland, Ringkøbing	Clayey soil	1996
Ellebæk	Elle	Jutland, Ringkøbing	Clayey soil	1996
Højvads Rende	Høj	Lolland, Storstrøms Amt	Clayey soil	1996

The type of soil of the catchment areas can have an influence on the transportation of pesticides into streams. In clayey soils the transport of pesticides to ground water will often be less compared to the transport of pesticides to ground water in sandy soils because of a higher field capacity of the top soil resulting in an increased evaporation taken as a long term average and a prolonged retention time of the pesticides in the top soil. On the other hand in clayey soils transportation of pesticides to streams happens to a higher degree by drain and by surface run-off compared to areas with sandy soil types. However, in some cases sandy soil types can have a higher degree of organic matter than clayey soils which will give a higher adsorption and in a dry clayey soil water and pesticides can run through cracks in the soil matrix with little adsorption.

2 Ranking of pesticides

The ranking of data is compared to different ranking models. Ranking of the data is based on: The measured concentration of the pesticides in the streams, the limit of detection and the detection frequency.

The other ranking is the ranking model based on use data: The sprayed area of the pesticide, the estimated recommended dose of pesticides in Denmark and the adsorption coefficient, Koc, where high Koc values indicate high adsorption to organic carbon and thereby a limited occurrence in the surface water. It is also possible to use other parameters, e.g. the dissipation half life, DT50, which indicates the potential of degradation for the pesticide where high values indicate a slow degradation time and thereby a higher risk for transportation to surface water. However, the Koc parameter is identified by *Sørensen et al. (1999)* as the most important physicochemical parameter and other parameters will be considered only if the three chosen parameters are insufficient to make a good agreement to the data ranking result. The use data can be seen from Table 2.

Table 2. Use data for the ranked pesticides. The estimated dose and the area is an average for the period 1994-1997. The data for dose and area are from the Danish EPA, 1994, 1995, 1996, 1997. The data for Koc are from *Lindhardt et al., 1999* if no other reference is given. **Kreuger & Törnqvist, 1998*, ***Montgomery, 1993*, ****Linder et al., 1994*.

<i>Active ingredient</i>	<i>Dose (kg)</i>	<i>Area (ha)</i>	<i>Koc</i>
2,4-D	1,110	49048	33
Atrazin	0,747	890	170**
Benazolin-ethyl	0,000	0	30*
Bentazone	0,671	480938	91
Carbofuran	0,667	7107	24
Clopyralid	0,138	358246	8
Cyanazin	0,342	138490	94*
Diazinon	0,000	0	1000**
Dichlobenil	0,000	0	153**
Dichlorprop	3,335	51743	121
Dimethoate	0,300	513116	32
Dinoseb	0,000	0	320**
Diuron	0,000	0	551**
DNOC	0,000	0	42***
Esfenvalerate	0,010	1644200	2083
Ethofumesate	0,668	212572	219
Fenpropimorph	0,483	2235741	4382
Glyphosat	1,033	1807769	25424
Hexazinon	0,000	0	99**
Ioxynil	0,245	1406561	828
Isoproturon	1,236	1508665	53
Linuron	1,000	26002	550
MCPA	1,479	678576	55
Mecorprop	45,979	17793	20
Metamitron	3,500	59228	358
Metazachlor	0,000	0	80*
Metsulfuronmethyl	0,005	44600	57
Pendimethalin	1,337	467668	14033
Pirimicarb	0,136	852460	290
Propiconazole	0,387	863552	770
Propyzamid	0,645	153981	944
Simazin	0,000	0	138**
Terbutylazin	1,127	135906	220

The dose in Table 2 is an estimated dose. The dose is estimated as the sold amount of pesticide in a certain year divided by the sprayed area at the same year. The sold amount of pesticide is the total sold amount for each pesticide and the sprayed area is the total sprayed area for each pesticide.

2.1 Data ranking

The ranking of pesticide occurrence data is done in three different ways: With the average measured concentrations of pesticides, with the weighted average concentrations, and with the average of the maximum concentrations for each pesticide. It is then tested which of the three ways of ranking has the highest degree of agreement with the ranking model. So different data rankings will be compared to different ranking models and it is assumed that the most correct data ranking and the most correct ranking model together will give the best agreement between data and model.

The data can be treated like a continuous relationship or a stochastic relationship (Figure 1).

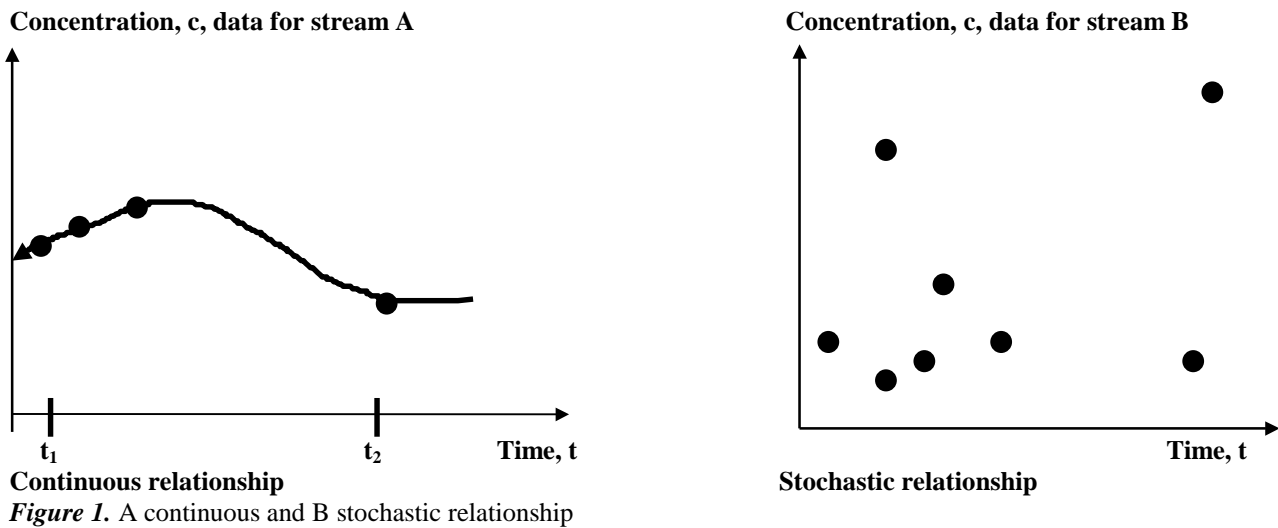


Figure 1. A continuous and B stochastic relationship

If the relationship is stochastic the average concentration is estimated as:

$$\bar{c} = \frac{1}{n_i} \sum c_i$$

and if the relationship is continuous the average concentration is estimated as: $\bar{c} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} c dt$. As can be seen from Figure 1 a

continuous relationship can be described by a function and a stochastic relationship is a random set of data which can not be described by a continuous mathematical function. When treating data that follows a stochastic relationship a simple average of the measured concentration is estimated.

If the data can be described by a function as is the case for stream A in Figure 1 then the data follows a continuous relationship and should be treated like that by weighting the concentration. The weighted concentration for a continuous set of data is the area under the curve divided by the period of time.

To estimate the weighted concentration of each pesticide the area under the curve for each stream for each pesticide is estimated.

Concentration, C, stream A

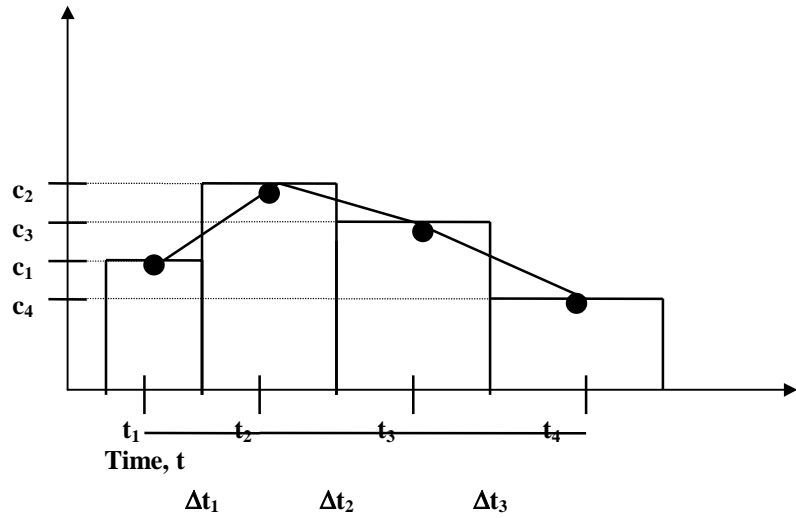


Figure 2.

The area under this curve is estimated as:

$$\text{area} = \Delta t_1 \times c_1 + \Delta t_n \times c_n + \frac{1}{2} \sum_{i=2}^{n-1} c_i (\Delta t_{i-1} + \Delta t_i), \quad n = \text{the number of}$$

measurements in the series (Figure 2). The weighted concentration is estimated as the sum of the areas for each pesticide divided by the sum of

the Δt for each pesticide: $\text{Conc} = \frac{\sum \text{area}}{\sum \Delta \text{time}}$

The pesticides will hereafter be ranked by the three different ways and this is compared to the ranking model to see if the data follows a continuous or a stochastic relationship.

2.2 Ranking by continuous relationship

The data is now treated like they follow a continuous set of data and the weighted average concentrations are estimated. The weighted average concentrations, the detection frequency and the limit of detection can be seen from Table 3.

Table 3. The weighted average concentration, the detection frequency and the limit of detection for each pesticide. The data are from *Ringkjøbing Amtskommune, 1997; Storstrøms Amt, 1996; Wiberg-Larsen et al., 1997; Wiggers, 1999.*

<i>Pesticide</i>	<i>Weighted average concentration µg/L</i>	<i>Detection frequency %</i>	<i>Limit of detection µg/L</i>	<i>Streams</i>
2,4-D	0,373	3	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Atrazin	0,143	18	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Benazolin-ethyl	0,000	0	0,010	Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Bentazone	0,105	43	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Carbofuran	0,000	0	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Clopyralid	0,000	0	0,200	Høj
Cyanazin	0,162	3	0,018	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Diazinon	0,000	0	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dichlobenil	0,106	40	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dichlorprop	0,099	20	0,017	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Dimethoate	0,162	8	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dinoseb	0,067	3	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Diuron	0,035	33	0,035	Høj
DNOC	0,098	35	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Esfenvalerate	0,272	9	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Ethofumesate	0,015	15	0,015	Ell, Horn, Jaun, Høj
Fenpropimorph	0,159	30	0,023	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Glyphosat				
Hexazinon	0,178	3	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Ioxynil	0,017	6	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Isoproturon	0,231	45	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Linuron				
MCPA	0,228	26	0,017	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Mecorprop	0,164	49	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Metamitron	2,290	6	0,029	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Metazachlor				
Metsulfuronmethyl	0,000	0	0,030	Høj
Pendimethalin	0,074	32	0,033	Ell, Horn, Jaun, Høj
Pirimicarb	0,131	11	0,029	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Propiconazole	0,239	18	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Propyzamid	0,041	23	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Simazin	0,197	22	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Terbutylazin	0,099	13	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle

There are no weighted average concentrations for Linuron, Glyphosat and Metazachlor because data for these pesticides are only available for “Lillebæk” stream and data for “Lillebæk” were only given as the maximum concentrations. Data for these three pesticides were therefore not included in the ranking with the weighted data.

The degree of agreement depends on the attributes used in the model (Table 4). The best agreement is found for the ranking model with the two attributes dose and area. With these two attributes there are 87 agreements and 20 disagreements, which means that 81% of the predictions in the model, which can be controlled by the data, are correct. If Koc is used as a third attribute the agreement with the ranking model decreases. When Koc is introduced as a third attribute there are only 50

agreements and 17 disagreements. There are also fewer comparisons because of the extra attribute which weakens the ranking. It is desirable to have as many comparisons as possible. Fewer comparisons and fewer agreements imply that Koc in this situation does not improve the prediction in the ranking model.

Table 4. Agreements and disagreements with different numbers of attributes with the data treated as a continuous set of data.

<i>Attributes used in the model</i>	<i>Number of comparisons</i>	<i>Number of agreements</i>	<i>Number of disagreements</i>	<i>Degree of agreements</i>
Dose	137	102	35	0.74
Area	137	100	37	0.73
Koc	178	76	82	0.48
Dose, area	107	87	20	0.81
Dose, Koc	96	62	34	0.65
Koc, area	74	53	21	0.72
Dose, area, Koc	67	50	17	0.75

2.3 Ranking by stochastic relationship

The data can be treated like a stochastic set of data in two ways. One way is by taking a simple average of the maximum concentration of each pesticide in each stream. The other way is by taking a simple average of all the data for each pesticide.

2.3.1 Ranking by average maximum concentration

The data is now treated like a stochastic set of data and in the data ranking a simple average of the maximum measured concentration for each pesticide is estimated. The maximum measured concentration, the limit of detection and the detection frequency can be seen in Table 5.

Table 5. Average of maximum measured concentrations ($\mu\text{g/L}$), detection frequency (%) and limit of detection ($\mu\text{g/L}$) of pesticides in streams in Denmark. The data are taken from Ringkjøbing Amtskommune, 1997; Storstrøms Amt, 1996; Wiberg-Larsen et al., 1997; Wiggers, 1999.

<i>Pesticide</i>	<i>Maximum average concentration $\mu\text{g/L}$</i>	<i>Detection frequency %</i>	<i>Limit of detection $\mu\text{g/L}$</i>	<i>Abbreviation</i>
2,4-D	0,329	3	0,036	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Atrazin	0,658	32	0,029	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Benazolin-ethyl	0,000	0	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Bentazone	4,267	30	0,025	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Carbofuran	0,000	0	0,073	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Clopyralid	0,133	1	0,225	Lille, Høj
Cyanazin	0,150	6	0,043	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Diazinon	0,000	0	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dichlobenil	0,697	40	0,028	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dichlorprop	0,275	13	0,068	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Dimethoate	0,350	8	0,033	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dinoseb	0,120	2	0,015	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Diuron	1,018	16	0,063	Lille, Høj
DNOC	0,590	27	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Esfenvalerate	0,287	13	0,075	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Ethofumesate	0,216	24	0,052	Lille, Ell, Horn, Jaun, Høj
Fenpropimorph	0,376	25	0,040	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Glyphosat	0,280	56	0,013	Lille, Ell, Horn, Jaun
Hexazinon	1,460	21	0,032	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Ioxynil	0,043	23	0,032	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Isoproturon	1,415	37	0,049	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Linuron	0,600	5	0,100	Lille
MCPA	1,150	21	0,055	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Mecorprop	1,635	51	0,036	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Metamitron	3,590	8	0,023	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Metazachlor	0,100	2	0,075	Lille
Metsulfuronmethyl	0,000	0	0,030	Høj
Pendimethalin	0,310	19	0,055	Ell, Horn, Jaun, Høj
Pirimicarb	23,544	9	0,035	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Propiconazole	0,781	20	0,035	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Propyzamid	0,090	20	0,040	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Simazin	0,715	20	0,050	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Terbutylazin	0,167	23	0,028	Lille, Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle

As for the previous ranking the degree of agreement between the two rankings depends on which attributes are used in the model (Table 6). The best model ranking is made with the attributes dose and area with 111 agreements and 14 disagreements, which means that 89% of the prediction in the ranking model with the use data are correct. When the model ranking is compared to the data ranking and a third attribute, Koc, is used, there are 65 agreements and 12 disagreements. This is the same situation as for the data treated like a continuous relationship that Koc does not improve the ranking model.

Table 6. Agreements and disagreements with different numbers of attributes.

<i>Attributes used in the model</i>	<i>Number of comparisons</i>	<i>Number of agreements</i>	<i>Number of disagreements</i>	<i>Degree of agreements</i>
Dose	148	123	25	0.83
Area	148	122	26	0.82
Koc	167	86	81	0.51
Dose, area	125	111	14	0.89
Dose, Koc	83	67	16	0.81
Koc, area	94	72	22	0.77
Dose, area, Koc	77	65	12	0.84

The result of the ranking with the weighted data gave fewer agreements compared to disagreements than the ranking with the data treated like a stochastic set of data with the maximum average concentration.

2.3.2 Ranking by the average concentration

Ranking of the data is also done by taking a simple average of the measured concentrations of pesticides in the streams and comparing this ranking with the ranking model. The data used for ranking the data by taking a simple average of the measured concentrations can be seen from Table 7.

Table 7. Average measured concentrations ($\mu\text{g/L}$), detection frequency (%) and limit of detection ($\mu\text{g/L}$) of pesticides in streams in Denmark. The data are taken from *Ringkjøbing Amtskommune, 1997; Storstrøms Amt, 1996; Wiberg-Larsen et al., 1997; Wiggers, 1999.*

<i>Pesticide</i>	<i>Average concentration $\mu\text{g/L}$</i>	<i>Detection frequency %</i>	<i>Limit of detection $\mu\text{g/L}$</i>	<i>Streams</i>
2,4-D	0,016	3	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Atrazin	0,036	18	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Benazolin-ethyl	0,000	0	0,010	Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Bentazone	0,052	43	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Carbofuran	0,000	0	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Clopyralid	0,000	0	0,200	Høj
Cyanazin	0,007	3	0,018	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Diazinon	0,000	0	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dichlobenil	0,039	40	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dichlorprop	0,022	20	0,017	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Dimethoate	0,016	8	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Dinoseb	0,002	3	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Diuron	0,007	33	0,035	Høj
DNOC	0,036	35	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Esfenvalerate	0,026	9	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Ethofumesate	0,003	15	0,015	Ell, Horn, Jaun, Høj
Fenpropimorph	0,048	30	0,023	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Glyphosat	0,213	56	0,013	Ell, Horn, Jaun
Hexazinon	0,012	3	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Ioxynil	0,001	6	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Isoproturon	0,137	45	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Linuron				
MCPA	0,067	26	0,017	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Mecorprop	0,086	49	0,013	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle, Høj
Metamitron	0,315	6	0,029	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Metazachlor				
Metsulfuronmethyl	0,000	0	0,030	Høj
Pendimethalin	0,030	32	0,033	Ell, Horn, Jaun, Høj
Pirimicarb	0,998	11	0,029	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Propiconazole	0,048	18	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Propyzamid	0,009	23	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Simazin	0,035	22	0,014	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle
Terbutylazin	0,018	13	0,010	Ell, Horn, Jaun, Støvl, Sund, Green, Herb, Lam, Skø, Fald, Vej, Elle

As with the previous ranking, the number of agreements and disagreements depends on the used attributes (Table 8). When the ranking is done by taking a simple average of the measured concentrations the best agreement between the data and the ranking model is also achieved with a model with dose and area as attributes. In this ranking there are 118 agreements and 24 disagreements which means a degree of agreement of 82% between the two rankings. If Koc is used as a third attribute, again there are fewer comparisons and the degree of agreement is only 73%. Nor does Koc improve the ranking model, when taking a simple average of the measured concentrations does the data ranking.

Table 8. Agreements and disagreements with different numbers of attributes.

<i>Attributes used in the model</i>	<i>Number of comparisons</i>	<i>Number of agreements</i>	<i>Number of disagreements</i>	<i>Degree of agreements</i>
Dose	202	145	57	0.72
Area	202	149	53	0.74
Koc	221	98	123	0.44
Dose, area	144	118	24	0.82
Dose, Koc	126	79	47	0.63
Koc, area	94	65	29	0.69
Dose, area, Koc	81	59	22	0.73

3 Odense Å

Data from a larger stream, Odense Å have also been ranked to see if the ranking model also agrees with data from such larger streams.

Odense Å is a stream on Fyn. The catchment area of Odense Å consists of both agricultural and urban areas and this is an important difference between this stream and the above streams And this difference may influence the predictability of the ranking model.

3.1 Ranking of data from Odense Å

Ranking of the data from Odense Å is done by ranking the maximum measured concentration of pesticides found in Odense Å with the detection frequency and the limit of detection. This ranking is compared to a ranking model like the rankings of the pesticides from the small streams and the two rankings are compared. If the ranking model has a good agreement with the data ranking the ranking model can also be used to indicate which pesticides can be found in larger streams with catchment areas of mixed land use . The ranked data can be seen from Table 9

Table 9. Maximum measured concentrations ($\mu\text{g/L}$), detection frequency (%) and limit of detection ($\mu\text{g/L}$) of pesticides measured in Odense Å. The measurements are made from 1994-1996. The pesticides dinoseb, DNOC, metamitron, metsulfuronmethyl and pendimethalin were not measured in Odense Å. The data are taken from *Wiberg-Larsen et al., 1997*.

<i>Pesticide</i>	<i>Weigthed average concentration $\mu\text{g/L}$</i>	<i>Detection frequency %</i>	<i>Limit of detection $\mu\text{g/L}$</i>
2,4-D	0,100	11	0,075
Atrazin	0,400	34	0,065
Benazolin-ethyl			
Bentazone	0,300	12	0,030
Carbofuran	0,400	2	0,200
Clopyralid	0,000	0	0,250
Cyanazin	0,300	2	0,100
Diazinon			
Dichlobenil	0,200	28	0,065
Dichlorprop	0,200	19	0,150
Dimethoate	0,100	2	0,075
Dinoseb			
Diuron	1,000	45	0,075
DNOC			
Esfenvalerate	0,000	0	0,200
Ethofumesate	0,100	13	0,115
Fenpropimorph	0,000	0	0,075
Glyphosat	0,200	100	0,015
Hexazinon	0,080	32	0,075
Ioxynil	0,000	0	0,075
Isoproturon	1,000	30	0,125
Linuron	0,000	0	0,100
MCPA	0,200	23	0,100
Mecorprop	0,400	42	0,075
Metamitron			
Metazachlor	0,000	0	0,075
Metsulfuronmethyl			
Pendimethalin			
Pirimicarb	0,070	26	0,060
Propiconazole	0,080	2	0,085
Propyzamid	0,800	4	0,100
Simazin	0,300	19	0,125
Terbutylazin	0,100	32	0,065

Data for dinoseb, DNOC, metamitron, metsulfuronmethyl, pendimethalin were not available for Odense Å. The number of agreements between the data ranking and the ranking model for Odense Å is also dependent on which attributes are used and how many attributes are used (Table 10). From Table 10 it can be seen that the ranking model doesn't agree as well with the data ranking for Odense Å as it did with data from small streams. The ranking model can in this case not be used to predict which pesticides will most probably be found in Odense Å.

Table 10. Agreements and disagreements with different numbers of attributes for Odense Å.

<i>Attributes used in the model</i>	<i>Number of comparisons</i>	<i>Number of agreements</i>	<i>Number of disagreements</i>	<i>Degree of agreements</i>
Dose	118	78	40	0.66
Area	118	41	77	0.35
Koc	123	81	42	0.66
Dose, area	67	34	33	0.51
Dose, Koc	82	59	23	0.72
Koc, area	53	26	27	0.49
Dose, area, Koc	42	22	20	0.52

4 Discussion

Treating the data like a stochastic set of data and using the average maximum concentration gave the best agreement with the model and will therefore be the best way to treat the data when assuming that the ranking model is right. The measured concentrations have to be treated like a stochastic set of data and can not be fitted as a function but the data are randomly distributed.

The rankings made with this set of data did not have as many agreements as the ranking model from the Swedish data. (Sørensen et al 1999). A reason may be that the actual doses and sprayed areas are not known in this model but is estimated from the amount of pesticide sold in Denmark and the sprayed area of the pesticide. In the data from Sweden the actual use data were known for the specific catchment and samples were collected more frequently during a longer period of time.

With these collected data from small streams in Denmark with agricultural catchments the model will in 89% of all cases predict which pesticides will be found most frequently and in the highest concentrations in the surface water in streams. These results indicate that it is possible in general to rank pesticides to predict which pesticides will most probably be found in small streams in rural areas.

The lack of agreement between the data ranking and the ranking model for Odense Å is probably due to the mixed land use within the catchment area, the use pattern of pesticides being different in rural and urban areas.

5 Conclusion

A ranking model was constructed from the data: The estimated recommended dose, the sprayed area and Koc. This model is compared to a ranking with the measured concentrations of pesticides in Danish streams, the detection frequency and the limit of detection.

The ranking of the measured data is made on data from small Danish streams with agricultural catchment areas and from one big stream with mixed land use in the catchment area. The ranking model showed that it can be used to predict which pesticides will most probably be found most frequently and/or in highest concentrations in small streams in Denmark. If only the estimated recommended dose and the treated area are used in the model and if the maximum measured concentration is used in the data ranking there is 89% degree of agreement between the ranking model and the data ranking. The ranking model would probably be a little better if the exact dose used in the catchment areas were known.

The same ranking model didn't agree very well with the data from the bigger stream, Odense Å having a catchment area of mixed land use. This is probably a consequence of the different use patterns for pesticides in agricultural and urban areas.

The ranking model based on the estimated recommended dose of pesticides and the total area the pesticide is sprayed on can be used to indicate which pesticides will be found most in small Danish streams with agricultural catchment areas. The ranking model can not be used to indicate which pesticides will be found most in larger streams with catchment areas with mixed land use.

6 References

Kreuger, J. & Törnqvist, L. (1998): “Multiple Regression Analysis of Pesticide Occurrence in Streamflow Related to Pesticide Properties and Quantities Applied”, *Chemosphere*, Vol. 37, No. 2, pp. 189-207.

Linders, J.B.H.J., Jansma, J.W., Mensink, B.J.W.G., Otermann, K. (1994): “Pesticides: Benefaction or Pandora’s Box? A synopsis of the environmental aspects of 243 pesticides”, National Institute of Public Health and Environmental Protection Bilthoven, The Netherlands.

Lindhardt, B., Sørensen, P.B., Elmegård, N., Fomsgård, I.S., Nielsen, E., Hart, J., Gravesen, L., Nielsen, L.A., Lorenzen, L., Madsen, K.J. (1999): “Teknisk baggrundsrapport om identifikation af de mest miljø- og sundhedsskadelige af de på nuværende tidspunkt godkendte pesticider”, Bichel-udvalget, “Rapport fra underudvalget om Miljø og sundhed”.

Montgomery, J.H. (1993): “Agrochemicals Desk Reference, Environmental Data”, Lewis Publishers.

Orientering fra Miljøstyrelsen (1995): “Bekæmpelsesmiddelstatistik, 1994”, Nr. 8.

Orientering fra Miljøstyrelsen (1996): “Bekæmpelsesmiddelstatistik, 1995”, Nr. 8.

Orientering fra Miljøstyrelsen (1997): “Bekæmpelsesmiddelstatistik, 1996”, Nr. 10.

Orientering fra Miljøstyrelsen (1998): “Bekæmpelsesmiddelstatistik, 1997”, Nr. 6.

Ringkjøbing Amtskommune (1997): Vandmiljøafdelingen, Notat: “Undersøgelse for pesticider i vandløb i Ringkjøbing Amt 1996.

Storstrøms Amt, Teknik- og Miljøforvaltningen (1997): Analyseprotokol.

Sørensen, P.B., Mogensen, B.B., Gyldenkerne, S., Rasmussen, A.G. (1998): “Pesticide Leaching Assessment Method for Ranking Both Single Substances and Scenarios of Multiple Substance use”, *Chemosphere*, Vol. 36. No. 10, pp. 2251-2276.

Sørensen, P.B., Brüggemann, R., Mogensen, B.B., Carlsen, L., Pudenz, S. (1999): “Compound data analysis of monitoring data of pesticides in surface waters. Data interpretation and model development using partial order ranking theory”, unpublished.

Tomlin, C.D.S. (1997): “The Pesticide Manual, British Crop Protection Council, 49 Downing Street Farnham, Surrey GU9/PH, UK.

Wiberg-Larsen, P., Pedersen, S.E., Tornbjerg, N.H., Sode, A., Muus, K., Wehrs, M. (1997): “De Fynske Vandløb. VANDMILJØovervågning. Tema: Ferskvand”, Fyns Amt.

Wiggers, L. (1999): “Pesticider i Vandløb, Kilder og Søer i Århus Amt”, Århus Amt.

Internet Forum for Partial Order Method - Hasse Web Site -

Kristina Voigt and Efraim Halfon²

*GSF - National Research Center for Environment and Health
Institute of Biomathematics and Biometry
85764 Neuherberg, Germany*

² *Visiting Professor, Institute of Fresh Water Ecology and Inland Fisheries
Abteilung Ökohydrologie
Rudower Chaussee 6A
12489 Berlin, Germany*

Key words:

Hasse diagram technique, partial order, discrete mathematics, Web page, Internet, information management

Abstract

The Internet is one of the main tools for disseminating and finding information. This paper presents a preliminary discussion and the possible layout of a web site dedicated to the use of Hasse software developed at our institute and elsewhere. This discussion is important since the Hasse software is being developed concurrently at several institutions worldwide. We feel that it is very important to inform potential users about the issues concerning the background and contents of Hasse diagram technique, the workshops, publications, the availability of expertise in this field and the development and distribution of Hasse software. In the future the active cooperation of interested users will be encouraged. The Web site consists of an introduction to the topic, workshops, publication references, list of experts, and software.

1 Introduction

The World Wide Web may very well be the most elegant real-world manifestation of the central metaphor of chaos theory. From its unassuming beginnings in 1989, the Web has picked up so much momentum that it is now "cracking the barrier of the largest information collection ever assembled by humans" [Sherman 1999]. With 201 million Internet users (as estimated in September 1999 [NUA 1999]), many individuals worldwide have the tools to use these data sources.

*201 million Internet users
in September 1999*

2 Structure of Hasse Web site

2.1 Hasse Home Page

The Hasse Web page will be written in HTML (Hyper Text Markup Language) with additional code in Javascript and Perl. The participants of the Workshop "Order Theoretical Tools in Environmental Sciences" agreed upon the usefulness of such a Web page to initiate and maintain communication. The Web site will be designed using frames. Figure 1 shows the planned Hasse Web site home page.

*Hasse Web page
HTML*

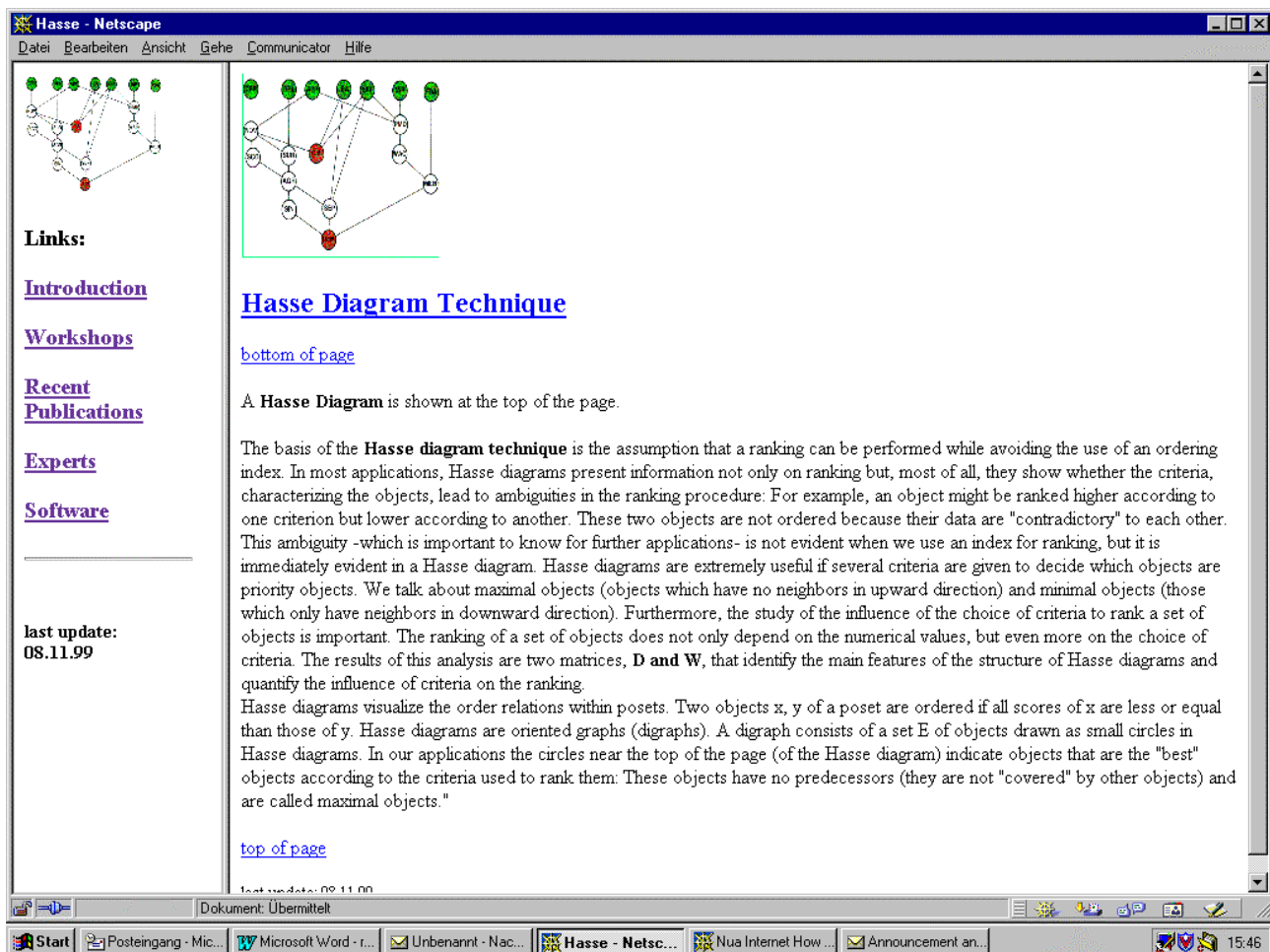


Figure 1. Hasse Home Page (200e.htm)

Hasse Home Page

The left side of the Hasse Home Page includes possible links. These might be:

- Introduction (Basic aspects of the Hasse diagram technique) [theory.htm]
- Workshops (recent and future workshops with call for papers, abstracts of proceedings etc.) [work.htm]
- Recent publications [pube.htm]
- Experts (main developers and users of the Hasse program) [expert.htm]
- Software (Hasse program can be ordered) [soft.htm]

The names of the htm files are given in brackets.

Meta-data

Meta-data are data about data; for example, tags that indicate the subject of a WWW document [Computer Currents 1999]. Meta-data are extremely important since search-engines look them up when building their indices [Webopedia 1999]; users of search engines can find Web pages by applying these meta-data as keywords. Meta-data are also called meta-tags. A special HTML tag that provides information about a Web page. Unlike normal HTML tags, meta-tags do not affect how the page is displayed. Instead, they provide information such as who created the page, when it was last updated, what the page is about, and which keywords represent the page's content.

Meta-tags

The meta-data or meta-tags included in the home page are Hasse Diagram, data mining, data analysis, ranking system, discrete mathematics, *Kristina Voigt, K. Voigt, E. Halfon, Efraim Halfon, Rainer Brüggemann, R. Brüggemann.*

2.2 Introduction to Hasse Diagram Technique

The following text is included as an introduction to Hasse Diagrams and a Hasse Diagram is shown at the top of the page.

The basis of the Hasse diagram technique is the assumption that a ranking can be performed while avoiding the use of an ordering index. In most applications, 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 because their data are "contradictory" to each other. This ambiguity -which is important 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. We talk about maximal objects (objects which have no neighbors in upward direction) and minimal objects (those which only have neighbors in downward direction). Furthermore, the study of the influence of the choice of criteria to rank a set of objects is important. The ranking of a set of objects does not only depend on the numerical values, but even more on the choice of criteria. The 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.

Hasse diagrams visualize the order relations within posets. 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 (digraphs). A digraph consists of a set E 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 "best" objects according to the criteria used to rank them. These objects have no predecessors (they are not "covered" by other objects) and are called "maximal objects."

The text above is an excerpt from the publication concerning the ranking of environmental databases applying the Hasse diagram technique [Brüggemann 1996].

Hasse diagrams

Two matrices, D and W

Maximal objects

2.3 Workshops

Order Theoretical Tools in Environmental Sciences

Two workshops concerning "Order Theoretical Tools in Environmental Sciences" have taken place, the first one in Berlin, Germany 1998 and the second one in Roskilde, Denmark in 1999. The third one will most probably be held in Berlin in the year 2000. The participants of the last workshop in Roskilde were of the opinion that holding the workshops on a one-year basis should be preferred to the two year time span. Figure 2 gives the html page of the workshops.

```
html>
<head>
<IMG src="hassevoigt.gif" ALT="[Hassevoigt]" WIDTH=160 HEIGHT=150>
<a href="/work.htm">
<title> Hasse</title>
</head>
<body>
<h1>Hasse Diagram Technique</h1>
<a name="oben"></a>
<a href="#unten">bottom of page</a>
<h1>Workshops and Conferences</h1>
<hr>
<h2>The following Workshops and Conferences were/are held:</h2>
<ul type="square">
<br>
<li> <a href="#Berlin"> A. 1.Workshop on Order Theoretical Tools in Environ-
mental Sciences, Berlin, Germany 1998</a><br>
<li> <a href="#Roskilde"> B. 2. Workshop in Partial Order Ranking Methods, Ro-
skilde, Denmark, 1999</a><br>
</ul>
<h3><a name="Berlin"> A. Workshop on Order Theoretical Tools in Environ-
mental Sciences</a></h3><br>
held in Berlin, Germany on <b>16.11.98 </b> at the <a href="http://www.igb-
berlin.de/www/abt1/abt1.htm" target=_top>
Institute for Fresh Water Ecology and Inland Fisheries, Dr. Rainer
Br&uuml;ggemann
</a><br><p>
<br><br>
<h3><a name="Roskilde"> B. 2. Workshop in Partial Order Ranking Methods or
Application
and Method Development</a></h3><br>
held in Roskilde, Denmark
on <b>22.10.99 </b> at the <a href="http://www.dmu.dk/EnvironmentalChemistry/"
target=_top>
NERI (National Environmental Research Institute), Department of Environmental
Chemicals
Dr. Peter Soerensen </a><br><p>
</a><br>
<br><br>
<hr>
<a href="#oben">top of page</a><br><p>
<a name="unten"></a>
<small>last update: 08.11.99 </small>
</body>
</html>
```

Figure 2. Hasse Workshops html format

This web page contains links to the institutions which organized or will organize the workshops. Call for papers, proceedings' abstracts etc. could be included here to advertise future events. The corresponding Internet layout (abbreviated) is given in Figure 3.



Figure 3. Hasse Workshops (work.htm)

2.4 Recent Publications

Hasse publications

The Hasse publications are divided into the following categories:

- A. Refereed journal papers
- B. Articles in books and refereed conference papers
- C. Reports

As of November 1999, we list 15 papers in scientific journals, 20 articles in books and proceedings, and 2 reports. This list consists mainly of publications from *Rainer Brüggemann*, *Kristina Voigt* and *Efraim Halfon*. This list needs to be urgently updated. We will include a page where anyone can add his or her information for posting. Given the interactive needs this page should be programmed in Javascript with additional code on our server in CGI and Perl.

2.5 Experts

Experts

The following experts and institutions are working in the field of HDT and can be contacted by e-mail:

- A. Institut für Gewässerökologie und Binnenfischerei, Dr. Rainer Brüggemann (brg@igb-berlin.de)
- B. Dr. Efraim Halfon, Burlington, Canada (info@butx.com)
- C. National Environmental Research Institute Denmark, Roskilde, Dr. Peter Soerensen (PBS@DMU.dk)
- D. GSF- National Research Center for Environment and Health, Dr. Kristina Voigt (kvoigt@gsf.de)

The address with e-mail and URL are provided here. Again this list has to be completed by naming other experts.

2.6 Software

Hasse diagram software

Most important to the visitors of this Web page is the information how and where to obtain the Hasse diagram software. The latest version is only available on CD-ROM from the Institute of Fresh Water Ecology and Inland Fisheries in Berlin, Dr. Rainer Brüggemann. An abridged Internet version of the Hasse program is under development at the same institution.

3 Conclusions and Outlook

Webmaster

The development of this web site will be a collaborative effort and needs input from other experts working with the Hasse diagram technique. The choice of a Webmaster and the location of the server are still open for discussion but should be resolved soon since this site is urgently needed to inform the Internet scientific community about this important and useful mathematical tools to rank and evaluate objects.

4 References

Brüggemann R., Voigt K., Halfon E., 1996, Partially Ordered Sets - a Computerized Tool to Compare Environmental Databases in: Denzer R., Schimak G., Russell D., Environmental Software Systems, Proceedings of the International Symposium on Environmental Software Systems 1995, pp. 185-195, Chapman & Hall, London.

Computer Currents High-Tech Dictionary, 1999,
<http://www.currents.net/resources/dictionary/dictionary.phtml>

NUA Internet Surveys, 1999,
http://www.nua.ie/surveys/how_many_online/index.html, September 1999.

Sherman C., 1999, The Future of Web Search, Online **23**, 3, 54-61.

Webopedia, 1999, <http://www.pcwebopaedia.com/>.

National Environmental Research Institute

The National Environmental Research Institute, NERI, is a research institute of the Ministry of Environment and Energy. In Danish, NERI is called *Danmarks Miljøundersøgelser (DMU)*.

NERI's tasks are primarily to conduct research, collect data, and give advice on problems related to the environment and nature.

Addresses:

URL: <http://www.dmu.dk>

National Environmental Research Institute
Frederiksborgvej 399
PO Box 358
DK-4000 Roskilde
Denmark
Tel: +45 46 30 12 00
Fax: +45 46 30 11 14

Management
Personnel and Economy Secretariat
Research and Development Section
Department of Atmospheric Environment
Department of Environmental Chemistry
Department of Policy Analysis
Department of Marine Ecology and Microbiology

National Environmental Research Institute
Vejlsøvej 25
PO Box 314
DK-8600 Silkeborg
Denmark
Tel: +45 89 20 14 00
Fax: +45 89 20 14 14

Department of Lake and Estuarine Ecology
Department of Terrestrial Ecology
Department of Streams and Riparian areas

National Environmental Research Institute
Grenåvej 12, Kalø
DK-8410 Rønde
Denmark
Tel: +45 89 20 17 00
Fax: +45 89 20 15 14

Department of Landscape Ecology
Department of Coastal Zone Ecology

National Environmental Research Institute
Tagensvej 135, 4
DK-2200 København N
Denmark
Tel: +45 35 82 14 15
Fax: +45 35 82 14 20

Department of Arctic Environment

Publications:

NERI publishes professional reports, technical instructions, and the annual report. A R&D projects' catalogue is available in an electronic version on the World Wide Web.

Included in the annual report is a list of the publications from the current year.

Faglige rapporter fra DMU/NERI Technical Reports

1999

- Nr. 287: The Danish CORINAIR Inventories. Time Series 1975-1996 of Emissions to the Atmosphere. By Winther, M., Illerup, J.B., Fenhann, J. & Kilde, N. 81 pp., 100,00 DDK.
- Nr. 288: Mere og bedre natur i landbrugslandet - dokumenteret grundlag for en ekstra indsats. Reddersen, J., Tybirk, K., Halberg, N. & Jensen, J. 109 s., 120,00 kr.
- Nr. 289: Atmosfærisk deposition af kvælstof 1998. NOVA 2003. Af Skov, H., Hertel, O., Ellermann, T., Skjødt, C.A. & Heidam, N.Z. 102 s., 110,00 kr.
- Nr. 290: Marine områder - Status over miljøtilstanden i 1998. NOVA 2003. Af Markager, S. et al. 161 s., 150,00 kr.
- Nr. 291: Søer 1998. NOVA 2003. Af Jensen, J.P., Søndergaard, M., Jeppesen, E., Lauridsen, T.L. & Sortkjær, L. 106 s., 125,00 kr.
- Nr. 292: Vandløb og kilder 1998. NOVA 2003. Af Bøgestrand, J. (red.) 130 s., 150,00 kr.
- Nr. 293: Landovervågningsoplande 1998. NOVA 2003. Af Grant, R. et al. 152 s., 150,00 kr.
- Nr. 294: Bilparkmodel. Beregning af udvikling og emissioner. ALTRANS. Af Kveiborg, O. 84 s., 75,00 kr.
- Nr. 295: Kvalitetsparametre for haglammunition. En undersøgelse af spredning og indtrængningsevne som funktion af haglenes størrelse og form. Af Hartmann, P., Kanstrup, N., Asferg, T. & Fredshavn, J. 34 s., 40,00 kr.
- Nr. 296: The Danish Air Quality Monitoring Programme. Annual Report for 1998. By Kemp, K. & Palmgren, F. 64 pp., 80,00 DKK.
- Nr. 297: Preservatives in Skin Creams. Analytical Chemical Control of Chemical Substances and Chemical Preparations. By Rastogi, S.C., Jensen, G.H., Petersen, M.R. & Worsøe, I.M. 70 pp., 50,00 DKK.
- Nr. 298: Methyl t-Butylether (MTBE) i drikkevand. Metodeafprøvning. Af Nyeland, B., Kvamm, B.L. 51 s., 50,00 kr.
- Nr. 299: Blykontaminering af grønlandske fugle - en undersøgelse af polarlomvie til belysning af human eksponering med bly som følge af anvendelse af blyhagl. Af Johansen, P., Asmund, G. & Riget, F.F. 27 s., 60,00 kr.
- Nr. 300: Kragefugle i et dansk kulturlandskab. Feltundersøgelser 1997-99. Af Hammershøj, M., Prang, A. & Asferg, T. 31 s., 40,00 kr.
- Nr. 301: Emissionsfaktorer for tungmetaller 1990-1996. Af Illerup, J.B., Geertinger, A., Hoffmann, L. & Christiansen, K. 66 s., 75,00 kr.
- Nr. 302: Pesticider 1 i overfladevand. Metodeafprøvning. Af Nyeland, B. & Kvamm, B.L. 322 s., 150,00 kr.
- Nr. 303: Ecological Risk Assessment of Genetically Modified Higher Plants (GMHP). Identification of Data Needs. By Kjær, C., Damgaard, C., Kjellsson, G., Strandberg, B. & Strandberg, M. 32 pp., 50,00 DKK.
- Nr. 304: Overvågning af fugle, sæler og planter 1998-99, med resultater fra feltstationerne. Af Laursen, K. (red.). 81 s., 70,00 kr.
- Nr 306: Mercury in Soap in Tanzania. By Glahder, C.M., Appel, P.W.U. & Asmund, G. 19 pp., 60,00 DKK.

2000

- Nr. 307: Cadmium Toxicity to Ringed Seals (*Phoca hispida*). An Epidemiological Study of possible Cadmium Induced Nephropathy and Osteodystrophy in Ringed Seals from Qaanaaq in Northwest Greenland. By Sonne-Hansen, C., Dietz, R., Leifsson, P.S., Hyldstrup, L. & Riget, F.F. (in press)
- Nr. 308: Økonomiske og miljømæssige konsekvenser af merkedordsordningerne i EU's landbrugsreform. Agenda 2000. Af Andersen, J.M., Bruun et al. 63 s., 75,00 kr.
- Nr. 309: Benzene from Traffic. Fuel Content and Air Concentrations. By Palmgren, F., Hansen, A.B., Berkowicz, R. & Skov, H. 42 pp., 60,00 DKK
- Nr. 310: Hovedtræk af Danmarks Miljøforskning 1999. Nøgleindtryk fra Danmarks Miljøundersøgelser jubilæumskonference Dansk Miljøforskning. Af Secher, K. & Bjørnsen, P.K. (i trykken)
- Nr. 311: Miljø- og naturmæssige konsekvenser af en ændret svineproduktion. Af Andersen, J.M., Asman, W.A.H., Hald, A.B., Münier, B. & Bruun, H.G. (i trykken)
- Nr. 312: Effekt af døgnregulering af jagt på gæs. Af Madsen, J., Jørgensen, H.E. & Hansen, F. 64 s., 80,00 kr.
- Nr. 313: Tungmetalledfald i Danmark 1998. Af Hovmand, M. & Kemp, K. (i trykken)
- Nr. 314: Future Air Quality in Danish Cities. Impact Air Quality in Danish Cities. Impact Study of the New EU Vehicle Emission Standards. By Jensen, S.S. et al. (in press)
- Nr. 315: Ecological Effects of Allelopathic Plants – a Review. By Kruse, M., Strandberg, M. & Strandberg, B. 64 pp., 75,00 DKK.
- Nr. 316: Overvågning af trafikens bidrag til lokal luftforurening (TOV). Målinger og analyser udført af DMU. Af Hertel, O., Berkowicz, R., Palmgren, F., Kemp, K. & Egeløv, A. (i trykken)
- Nr. 317: Overvågning af bæver *Castor fiber* efter reintroduktion på Klosterheden Statsskovdistrikt 1999. Red. Berthelsen, J.P. (i trykken)
- Nr. 318: Order Theoretical Tools in Environmental Sciences. Proceedings of the Second Workshop October 21st, 1999 in Roskilde, Denmark. By Sørensen, P.B. et al. (in press)
- Nr. 319: Forbrug af økologiske fødevarer. Del 2: Modellering af efterspørgsel. Af Wier, M. & Smed, S. (i trykken)

[Blank page]

This is a proceeding form: The second workshop on Order Theoretical Tools in Environmental Sciences. The principle of partial ranking is presented and discussed as a tool for analyzing of environmental problems. The method is show to be effective for analysis of complex systems where incomparability exist among different key parameters.

Ministry of Environment and Energi
National Environmental Research Institit

ISBN 87-7772- 548-4
ISSN (Print) 0905-815X
ISSN (Electronic) 1600-0048