Method of Evaluation by Order Theory (METEOR) applied on the Topic of Water Contamination with Pharmaceuticals

K. Voigt[1] and R. Brüggemann[2]
[1]GSF - National Research Center for Environment and Health, Institute of Biomathematics and Biometry, 85764 Neuherberg, Germany, kvoigt@gsf.de
[2]Leibniz-Institute of Freshwater Ecology and Inland Fisheries, 12587 Berlin, Germany, (brg@igb-berlin.de

Keywords: Partial order theory, Method of Evaluation by Order Theory (METEOR), Hasse Diagram Technique, water pollution, drinking water, groundwater, pharmaceuticals

Abstract:
A growing concern to be tackled today is the potential contamination of water resources by chemicals - including pharmaceuticals entering food chains and causing adverse effects in animals and humans. In this paper 75 publications from the years 2000-2004 are analysed concerning their information on 12 chosen pharmaceuticals. These pharmaceuticals are recognised as potential pollutants to the media water and soil. A 12x75 data-matrix is the basis for a mathematical evaluation (ranking) approach using the Method of Evaluation by Order Theory (METEOR).
The original data-matrix of 12 objects and 75 attributes is subject to several logical aggregation steps applying the Hasse Diagram Technique. It can be demonstrated that the aggregation to get three super indicators concerning different kinds of journals leads to a highly structured diagram with more comparabilities than incomparabilities. The aggregation steps with different weights, normalisation to 1 lead to linear order diagrams. These diagrams all show that the following ranking of the objects EES>IBU>BEZ>ROX>MET>FEN is given in all the linear order diagrams. All partial order diagrams and linear order diagrams demonstrate that the pharmaceutical EES: Estinyl Estradiol is always in a maximal position whereas FEN: Fenofibrate is always a minimal object.

1 Contamination of Water and Soil with Pharmaceuticals

Achieving sustainable development in the environmental and health sector it is absolutely necessary to keep the ground- and consequently the drinking water free of contaminants. Unfortunately several chemicals are detected in the media water: surface water, wastewater, groundwater, drinking water, sediments and soil. Groundwater contaminants come from two categories of sources: point sources and distributed, or non-point sources. Landfills, leaking gasoline storage tanks, leaking septic tanks, and accidental spills are examples of ***point sources***. Infiltration from farm land treated with pesticides and fertilizers is an example of a ***non-point source***. Wastewatersfrom households and hospitals are also considered as non-point sources.
The quality and availability of drinking water are essential for survival and the reality of life on earth. However, water is a limited resource everywhere in the world even in Europe. A growing concern to be tackled today is the potential contamination of water resources by chemicals - including pharmaceuticals entering food chains and causing adverse effects in animals and humans. High quality research can help to solve part of the problem. The quality and availability of drinking water are essential for survival and the reality of life on earth.
Pharmaceuticals, hormones and endocrine disruptors have become major issues in environmental chemistry, due to their presence in environmental waters (following incomplete removal in wastewater treatment or point-source contaminations) and concern about possible estrogenic and other effects, both on wildlife and humans [Richardson 2003]. The occurrence and fate of pharmaceutically active compounds (PhACs) in the aquatic environment has been recognized as one of the emerging issues in environmental chemistry. More than 80 PhACs from various prescription classes have been detected in the µg/l-level in sewage, surface and groundwater.

A systematic investigation of how effectively drinking water treatment technologies remove pharmaceutical products has found that the technologies being used in Germany appear to be rather effective [Ternes and Meisenheimer 2002]. However, some of the technologies used elsewhere in the world particularly in the United States—may be letting pharmaceuticals through.

Although research shows that pharmaceutically active products are found in surface waters throughout the United States [Kolpin and Furlong 2002] and Europe, there is as yet very little information on how effectively different drinking water treatment technologies remove these pharmaceutical residues.

Hence it is evident that environmetrical and chemometrical studies should be performed to get more knowledge out of the already existing data. This is the reason why we performed the study described in this paper.

2 Literature Survey for 12 Pharmaceuticals

We took 12 pharmaceuticals of seven different groups of drugs into consideration. Contrast media and cytostatic drugs were left out as they are mainly found in hospital sewage. Pharmaceuticals which have been chosen for the approach belong to the following groups of pharmaceuticals:

1. Lipid regulators (Bezafibrate, Chlofibric acid, Fenofibrate)
2. Antiphlogostics (Diclofenac, Ibuprofen, Phenazone)
3. Betablocker (Metoprolol)
4. Phychiatric drugs (Diazepam)
5. Antiepileptic (Carbamazepine)
6. Antibiotics (Roxithromycine, Sulfamethoxazole)
7. Estrogens (Ethinylestradiol)
8. Contrast media
9. Cytostatic drugs

For the chosen 12 pharmaceuticals an intensive literature survey was performed during the month of February 2004. For the evaluation approach only the years 2000 – 2004 were considered. Relevant articles were found in 29 different scientific journals. The journals with a high number of publications were the following journals:

J03: Anal. Chem:	8 publications
J05: Chemosphere	5 publications
J07: Environ. Sci. Technol.	13 publications
J16: Journal of Chromatography A:	12 publications

75 publications were considered. Most of them were published in 2003. Germany has 24 publications in the years 2000-2004 followed by the US and Spain with 12 publications. The dominant position of German studies is striking as international journals were evaluated and only one paper was found in a German journal.

The aim of this study is to get further insight into the problem of environmental pollution with pharmaceuticals. This first approach takes a look at the pharmaceuticals with respect to their occurrence in the environmental media water and soil.

3 Mathematical Evaluation Methods

The Hasse Diagram Technique is well explained in a variety of different environmental and chemical as well as statistical journals. A rather comprehensive description can be found in [Brüggemann and Welzl 2002]. A comparison of the Hasse Diagram Technique with multi-variate statistical methods is given by Voigt [Voigt and Welzl 2004a]. Therefore only some aspects are picked out, which will be useful in the subsequent application. Hasse Diagrams visualize the order relations within objects: Two objects, also called elements (if the aspect of belonging to sets is important) x, y of an object set are considered as being ordered, e.g. $x \leq y$, if all scores of x are less or equal than those of y. Hasse Diagrams are acyclic digraphs and objects are drawn as small circles together with an appropriate identifier. The edges of this graph are the cover-relations; that means, edges which express simply the transitivity are omitted, as they bear redundant information. In our applications the circles near the top of the page (of the Hasse Diagram) indicate objects that are the "better" objects according to the criteria used to rank them: The objects not "covered" by other objects are called maximal objects. Objects which do not cover other objects are called minimal objects. In some diagrams there exist also isolated objects which can be considered as maximal and minimal objects at the same time. Sometimes it is useful to call those elements as ‘proper’, which are not at the same time both, maximal and minimal elements.

When there is exactly one maximal and one minimal element resp., then these unique objects are called greatest and least element, resp.

The WHasse program is developed, improved and updated by Rainer Brüggemann (a brief technical information about the WHasse-program, written in DELPHI, can be found in a publication of the second author [Brüggemann and Bartel 1999]) and is available for non-commercial use from the second author. For commercial applications it is recommended to contact the company Criterion – Evaluation and Information Management [Criterion 2004].

3.1 Hasse Diagrams as mathematical objects

The basis of the Hasse Diagram technique (named HDT for short) is the assumption that a ranking can be performed while avoiding the use of an ordering index [Halfon and Reggiani 1986]. For an evaluation of the objects they must be compared. The comparison is done by examining characteristic properties (attributes, descriptors) of these objects. If the evaluation is aimed to assess criteria, then the attributes or (synonyms: descriptors) are thought of as measures, how well a criterion is fulfilled. Note that the concepts "criterion" and an assigned measure "attribute"' should kept well separated. Attributes are -in the case of the object "x" denoted as q(1,x), q(2,x),...,q(m,x) and often written as a tuple **q**(x). We avoid the term vector, because the properties of a linear space are not needed in the HDT. Note, however that within the concept of METEOR a linear space is needed; see below. Often the properties are gathered to a set without reference to actual values realized by the objects. This set of properties is called an **information base IB**. Often subsets of IB are needed, especially if METEOR is to be applied. Consider now two objects x and y, then we say $y \geq x$ (with respect to the m properties of interest) if
$q(i,x) \leq q(i,y)$ <u>for all</u> i =1 ,2,..,m and there is at least one i* , for which $q(i^*,x) < q(i^*,y)$ (because of the demand "for all" this definition is denoted as "generality principle").
If $q(i,x) \leq q(i,y)$ for all i=1,..,m then the objects x and y are comparable. The mere fact that x is comparable with y is often denoted as $x \perp y$.
Often however one finds
$q(i,x) < q(i,y)$ for one index set I' and
$q(i,x) > q(i,y)$ for another index set I'' with $I' \cap I'' = \emptyset$.
In that case, the objects x and y are incomparable and one writes: $x \parallel y$. However, the order relation defined here is known as product order. There are many other ways to define order relations.
The main frame of HDT is therefore (the four-point-program):

1. Selecting a set of elements of interest which are to be compared, *E*. The so-called ground set.
2. Selecting a set of properties, by which the comparison is performed, called the information base *IB*.
3. Find a common orientation for all properties; according to the criteria they are assigned.
4. Analysing $x,y \in E$ whether one of the following relations is valid:
   - $x \sim y$ (equivalence, we call the corresponding equivalence relation *R*, the equality of two tuples **q**(x), **q**(y))
   - $x \leq y$ or $x \geq y$ (comparability)
   - $x \parallel y$ (incomparability, there is a "contradiction in the data of x and y").

The relation defined above among all objects is indeed an order relation, because it fulfils the axioms of order, namely

- reflexivity (one can compare each object with itself)
- antisymmetry (if x is preferred to y then the reverse is only true, if the two objects are equal (or equivalent)
- transitivity (if x is better than y, and y is better than z, then x is better than z).

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 $(E, \leq)$.
We note: A set *E* equipped with a partial order is often written as $(E,\leq)$. Because the $\leq$-comparison depends on the selection of the information base (and of the data representation (classified or not, rounded, etc.) we also write (*E*, *IB*) to denote this important influence of the *IB* for any rankings.
Sometimes it is useful to refer to the quotient set, which is induced by the equivalence relation of equality, *R* (see for details: Brüggemann [Brüggemann and Bartel 1999]). As usual we write *E/R* for the quotient set, and (*E/R, **IB***) for the partially ordered quotient set.

When there are no paths between parts of the non-directed graph, these parts, supplied with the inherited order relation are called (isolated) hierarchies or -if only one object (one equivalence class) constitutes a part of the Hasse Diagram - an isolated element. Different hierarchies indicate quite specific data structures.

If empirical posets are to be examined it is important to establish orientation rules, i.e. which value of attributes is considered to contribute to badness and which values to goodness. Concerning the evaluation of the pharmaceuticals in environmental media by their publication in scientific journals
1 means: "good", pharmaceutical is analyzed.
0 means: "bad", pharmaceutical is not treated in this publication.

3.2 W-matrix : Dissimilarity-matrix

The W-matrix describes the influence of the attributes on the Hasse Diagram.
The entries of the W-matrix are a measure for the metric distance among posets, based on the same ground set of objects, but induced by different subsets of IB of m-1 attributes, i.e. subset generated by IB – $\{q_i\}$, i= 1,…, m. The definitions of the entries of the W-matrix depend on the actual selected subset of elements of *E*. Mostly the full ground set *E* is used. More details can be found in Brüggemann et al. [Brüggemann and Halfon 2001a]. For further reading we refer to background publications by Brüggemann [Brüggemann and Welzl 2002], [Brüggemann and Bartel 1999].

3.3 METEOR (Method of Evaluation by Order Theory)

Aggregation procedures of the data-matrix will be performed by applying by METEOR (Method of Evaluation by Order Theory). The basic idea is that subsets of the IB can be combined by weighted sums; see Brüggemann et al. [Brüggemann and Pudenz 2001b]. Therefore the columns of the data matrix (rows: the elements, columns the attributes) can be considered as vectors of a linear space. In order to combine them freely, a common scaling level must be assumed. Each positive monotonous combination of -say- two attributes, leading to a “superattribute” corresponds order theoretically to an order preserving map. Therefore the role of weighting can be traced back, when the final result, a linear order, is found by a stepwise aggregation. Furthermore, checking the local incomparability of any element, it is possible to identify weight-sensible and weight-insensible elements of the ground set *E* and *E/R*, resp., see Brüggemann et al. [Brüggemann 2001a].

4 Application of the Hasse Diagram Technique (HDT) on the 12x75 Data-matrix

12 pharmaceuticals (objects) are evaluated with 75 publications (attributes) applying the data-driven evaluation method, HDT (see Figure 1, left). It was found that the pharmaceuticals CAR, CLO, DIC, SUL, ROX and IBU show best results. These objects are proper maximal objects. There are no other pharmaceuticals which are better in all aspects than these proper maximal objects. FEN shows bad results in comparison to all other pharmaceuticals. It is a least object. The pharmaceuticals BEZ, DAP, EES and PHE are so-called isolated objects. They cannot be compared to any other object. Hence six proper maximal, one proper minimal and four isolated objects are found in the diagram of Figure 1 left hand side.

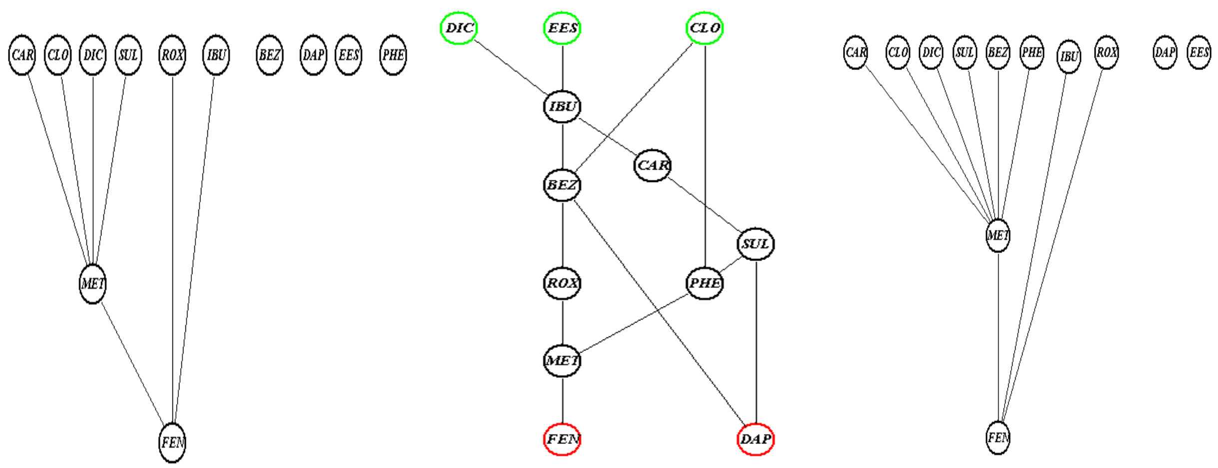

Figure 1: Hasse Diagrams of 12x75 data-matrix (left), 12x3 data-matrix (middle), 12x47 data-matrix

This diagram shows only 3 levels and has many incomparabilities. Differences concerning the numbers of successors can be detected in the diagram, e.g. the maximal objects CAR, CLO, DIC, and SUL are connected and hence comparable with two other objects, namely MET and FEN, whereas the maximal objects ROX and IBU are only connected with one other object, namely FEN.
Furthermore, for example MET is one of the articulation points (the other one is FEN). An articulation point is a vertex of the transitive hull of the digraph whose elimination would increase the number of hierarchies. Further information on this subject can be found in a recent publication [Voigt and Brüggemann 2004b]. By elimination of the object MET only one hierarchy would appear. All other objects would be isolated ones.
As the 12x75 data-matrix is very "broad" on the attribute side, the next logical step is therefore to perform data reduction procedures e.g. logical aggregations of attributes.

5 Application of METEOR

5.1 Weighting schemes (overview)

The original data-matrix of 12 pharmaceuticals (objects) and 75 parameters (attributes) will be subject to several logical aggregation steps. The aim of the aggregation procedure which can be performed by applying the Hasse Diagram Technique Program [Brüggemann and Welzl 2002] is to get a unique prioritisation scheme.
Several different weighting procedures are considered and performed (Figure 2).

1. Aggregation to get three super indicators concerning analytical journals (ANAL), the scientific journal Environmental Science and Technology (ESTE), and miscellaneous journals (MISC) (Section 5.1.1)
2. W-Matrix: 1 attribute is left out, all other 74 attributes are kept (Section 5.1.2). Based on this finding, the remaining 74 attributes are equally weighted, whereas the weight of the omitted attribute is formally considered as 0. This is an example of an extreme case of weighting (Section 5.1.3)
3. Equal weighting (Section 5.1.3)
4. Two different weighting, normalization to 1, n=75 (Section 5.1.3)

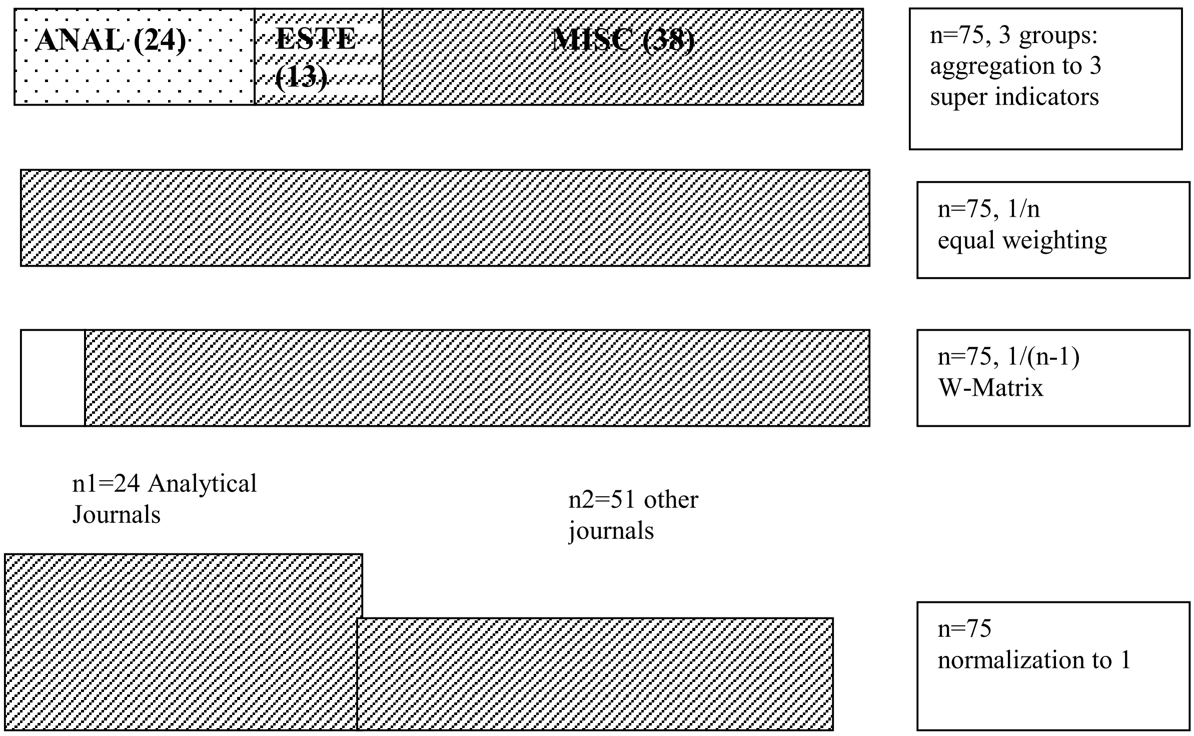

Figure 2: Different weighting schemes

5.1.1 Aggregation Equal Weight of Analytical and Environmental Journals

The aggregation of the data-matrix will be performed and the results presented by the application of METEOR (Method of Evaluation by Order Theory). The criteria (attributes) encompass scientific journals which have different focuses. We aggregate the three main analytical journals: Analytical Chemistry, Journal of Chromatography A and Trends in Analytical Chemistry into one super indicator, named ANAL. The second one is given by the journal Environmental Science and Technology (ESTE) whereas the third super indicator encompasses all the other journals (MISC). Hence we cope with three aggregation groups; where each of the following super indicators "ANAL", "ESTE", and "MISC" are calculated by a sum with equal weights. For example:

$$ANAL = \sum_{i=1}^{i=24} w_i \cdot q_i \quad q_i \in IB_{Analyticaljournals} = \{P03, P07, ..., P32\} \subset IB \quad w_i = 1/24 \qquad (1)$$

Aggregation of 24 publications in analytical journals:	ANAL
Aggregation of 13 publications in Environ. Sci. Tech.:	ESTE
Aggregation of 38 publications in miscellaneous journals:	MISC

The Hasse Diagra of the 12x3 data-matrix is given in Figure 1, middle section. A lot of visible changes took place comparing the original Hasse Diagram of the 12x75 data-matrix (Figure 1 left hand side) with the reduced 12x3 data-matrix (Figure 1 middle). The enhancement of the levels and comparabilities is enormous. No isolated object is found in this diagram. The former isolated object EES is a maximal object now and shows 9 successors. The other two maximal objects also encompass many successors. DIC has also nine successors whereas CLO has 8 ones.

5.1.2 W-Matrix: Dissimilarity-Matrix (leaving out 1 attribute)

The W-matrix is calculated for all objects given in the original diagram in Figure 1, left hand side. It reveals that the attribute 48 [publication Ternes et al. 20003] is the most important attribute in this approach. Four changes take place leaving out this attribute. The Hasse Diagram of Case 48 is given in Figure 1 right hand side. A few visible changes took place comparing this diagram with the results of the 12x75 data-matrix. Now we find only two isolated objects, namely DAP and EES. PHE and BEZ are maximal objects in this reduced data-matrix. Several differences can be detected while reducing the initial 12x27 data-matrix applying the METEOR method and the W-Matrix. In our examples we reduced the data-matrix to a matrix of 12x3 (equal weighting) and to 12x74 (W-Matrix).

All three diagrams in Figure 1 show a lot of differences but also similarities. In none of the diagrams equivalent objects are found. The Hasse diagram of the original 12x75 data-matrix as well as of the 12x74 data-matrix show only 3 levels and a huge number of incomparabilities and a small number of comparabilities. The diagram of the 12x3 data-matrix is well structured into 7 levels and gives more comparabilities than incomparabilities.

5.1.3 Aggregation Different Weights, Normalisation to 1

The next logical step is the aggregation of all attributes 1/74 = 0.01351 and the setting of P48 = 0. The linear order Hasse Diagram is given in Figure 3, left hand side.

The next step is to weight appropriately selected attribute-groups differently. For example: One might be more interested in analytical journals than in environmental ones or verse visa. The "ANAL" journals are looked upon and the "ESTE" and "MISC" are put together as "ENVI" journals.

For the first attempt we weight the analytical publications double with respect to the environmental ones.

Let n1 and n2 be the numbers of attributes in two attribute groups (here: analytical and environmental journals, resp.), w1 and w2 the weights by which the attributes of the one group and those of the other group are combined, and m the relative weight.

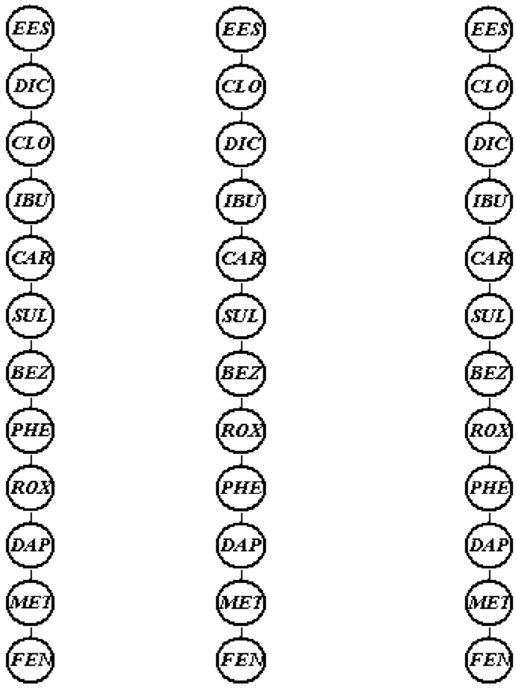

Figure 3: Hasse Diagram P48=0 and all other 74 objects 1/74 (left), middle and right: Analytical and environmental journals aggregated with different weights (see text)

Then, maintaining the normalization to 1 the weights can be calculated as follows:

$$w_2 = \frac{1}{(n1 \cdot m) + n2} \qquad (2)$$

$$w_1 = m \cdot w_2 \qquad (3)$$

As the analytical journals are considered to be more important than the environmental journals, m must be selected > 1. Here for m, arbitrarily the value 2 is given:

w_1 = (weight of analytical journals) = 0.01010
w_2 = (weight of environmental journals) = 0.02020

The corresponding Hasse Diagram is given in the Figure 3 in the middle section.
Clearly this Hasse Diagram comprises a linear order, not a partial order.

Alternatively we weight the environmental journals double with respect to the analytical journals, i.e. m=0.5.
w_2 = (weight analytical journals) = 0.0079
w_1 = (weight environmental journals) = 0.0158

The corresponding Hasse Diagram is given in the Figure 3, right hand side.

The ranking of the objects EES>IBU>BEZ>ROX>MET>FEN of the 12x3 data-matrix (partial order) is also kept in all linear order diagrams. It is striking that the two Hasse diagrams with different weights show the same linear orders. There is no difference if we weigh the analytical journals twice than the environmental journals or the other way round. With respect to the presence of pharmaceuticals in environmental media this does not make any difference. This means that all journals treat the same pharmaceuticals only different aspects of the topic, the analytical journals go more into depth into the analytical methods whereas the environmental journals take a look at the monitoring side or biodegradation side of the issue.
In the diagram in which the object P48=0 two changes in positions DIC-CLO and PHE-ROX are detected (see Figure 3 left hand side). EES = Estinyl Estradiol is always in a maximal position whereas FEN:

Fenofibrate is always a minimal object. This means that EES is the best scrutinized pharmaceutical and FEB the least analysed one within this information base IB.

6 Discussion and Outlook

As the topic of contamination of groundwater and drinking water becomes more and more important, the analyses of the existing results of the past years seem to be useful and necessary. In this first approach the availability of 12 well-known pharmaceuticals in 75 scientific publications was scrutinized. It was demonstrated that the Estrogen Ethinylestradiol was the best researched drug whereas for the Lipid regulator Fenofibrate a relatively small number of publications and hence analyses were found. The two performed aggregation procedures: aggregation different weight and normalization to 1 did not show different results in the linear order diagrams. This means that the specification of the journals (analytical or general environmental one) does not make any difference in the results. Both types of journals scrutinize the same pharmaceuticals. Further studies will be performed regarding the analyses of the media drinking water, groundwater, surface water, wastewater, sediment, soil. These data-analyses will show which media are more contaminated (more publications found) than other media.

7 References

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

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

Brüggemann, R. and S. Pudenz. 2001b. Hassediagrammtechnik zur Bewertung wasserwirtschaftlicher Massnahmen, Wasserforschung e.V., *Schriftenreihe Wasserforschung 7, Nachhaltige Entwicklung in der Wasserwirtschaft***.**1-15. Berlin: Wasserforschung e.V.

Brüggemann, R. and G. Welzl. 2002. Order Theory Meets Statistics – Hasse Diagram Technique. In: Voigt K, Welzl G. (eds.) Order Theoretical Tools in *Environmental Sciences, Order Theory (Hasse Diagram Technique) Meets Multivariate Statistics*. pp 9-40, Aachen: Shaker-Verlag,

Criterion. 2004. WHasse Software, http://www.criteri-on.de/hdtws-site/index.html

Halfon, E. and M. G. Reggiani. 1986. On Ranking Chemicals for Environmental Hazard. *Environ. Sci. Technol.* 20, pp 1173-1179.

Kolpin, D. W., W. T. Furlong, Meyer, M.T., Thurman, E.M., Zaugg, S.D., Barber, L.B., Buxton, H.T. 2002. Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in U.S. Streams, 1999-2000: A National Reconnaissance. *Environ. Sci. Technol.* 36, pp. 1202-1211.

Richardson, S. D. 2003. Disinfection by-products and other Emerging Contaminants in Drinking Water. *Trends in Analytical Chemistry* 22, pp. 666-684.

Ternes, T., M. Meisenheimer, McDowell, D., Sacher, F., Brauch, H.-J. 2002. Removal of Pharmaceuticals during Drinking Water Treatment *Environmental Science & Technology* 36, pp. 3855-3863.

Ternes, T., J. Stüber, Herrmann, N., McDowell, D., Ried, A., Kampmann, M., Teiser, B. 2003. Ozonization: A Tool for Removal of Pharmaceuticals, Contrast Media and Musk Fragrances from Wastewater *Water Research* 37, pp.1976-1982.

Voigt K., G. Welzl, R. Brüggemann. 2004a. Data Analysis of Environmental Air Pollutant Monitoring Systems in Europe *Environmetrics*, in press.

Voigt K., R. Brüggemann. 2004b. Chemical Databases Evaluated by Order Theoretical Tools *Analytical and Bioanalytical Chemistry*, submitted.