POLISH ACADEMY OF SCIENCES

Systems Research Institute

APPLICATIONS OF INFORMATICS IN ENVIRONMENT ENGINEERING AND MEDICINE

Editors:

Jan Studzinski
Ludosław Drelichowski
Olgierd Hryniewicz

APPLICATIONS OF INFORMATICS IN ENVIRONMENT ENGINEERING AND MEDICINE

Polish Academy of Sciences • Systems Research Institute

Series: SYSTEMS RESEARCH
Vol. 42

Series Editor:

Prof. Jakub Gutenbaum

Warsaw 2005

This publication was supported
by POLISH MINISTRY OF SCIENCE IN INFORMATION SOCIETY TECHNOLOGIES

This book consist of the papers describing the applications of informatics in environment and health engineering and protection. Problems presented in the papers concern quality management of the surface waters and the atmosphere, application of the mathematical modeling in environmental engineering, and development of computer systems in health and environmental protection. In several papers results of the research projects financed by the Polish Ministry of Science and Information Society Technologies are presented.

Papers Reviewers:

Prof. Ludosław Drelichowski
Prof. Olgierd Hryniewicz
Dr. Edward Michalewski
Prof. Andrzej Straszak
Dr. Jan Studzinski

Text Editor: Anna Gostynska

Systems Research Institute of Polish Academy of Science
Newelska 6, PL 01-447 Warsaw

Section of Scientific Information and Publications
e-mail: biblioteka@ibspan.waw.pl

ISBN 83-89475-04-9
ISSN 0208-8029

APPLICATIONS OF INFORMATICS IN ENVIRONMENT ENGINEERING AND MEDICINE

Editors:
Jan Studzinski
Ludosław Drelichowski
Olgierd Hryniewicz

CHAPTER 3

Informatics and Economy in Environment and Health Protection

APPLICATION OF COMPUTER-AIDED DECISION TOOLS CONCERNING ENVIRONMENTAL POLLUTION WITH PHARMACEUTICALS

Kristina VOIGT
GSF-National Research Center for Environment and Health, Neuherberg, Germany
<kvoigt@gsf.de>

Rainer BRÜGGMANN
Leibniz- Institute of Fresh Water Ecology and Inland Fisheries, Berlin, Germany

Stefan PUDENZ
Criterion, Berlin, Germany

In recent years a large number of pharmaceutical residues have been reported to occur in the aquatic environment. Due to their pharmacological activity, their determination and understanding of their behavior and fate in the environment are important. Environmetrical and chemometrical studies should be applied to improve the assessment of environmental risks that these chemicals may pose. Here a literature study was performed comprising the subject of 12 pharmaceuticals which have been detected in seven environmental media. 75 relevant articles have been found in 29 different scientific journals. This means that a 75 objects (publications) x 7 attributes (environmental media) data-matrix is looked upon. Furthermore a deeper look at the attributes (environmental media) will be performed. As an evolving data-analysis method the Hasse Diagram Technique (HDT) is introduced. This is a multi-criteria data-analysis method derived from discrete mathematics. The HDT is suitable and applicable as a multi-criteria decision tool. This method has been widely elaborated by the second author, the corresponding software ProRank is commercialized by the third author. Applying the Hasse Diagram Technique Method on theconsideration of the drugs in environmental media in publications the most important publications are revealed as well as the least important ones. Furthermore the application of the evaluation by the results of the dissimilarity matrix are exemplified.

Keywords: Hasse Diagram Technique, multi-criteria decision support, environmetrics, chemometrics, pharmaceuticals in the environment, environmental pollution.

1. Introduction Pharmaceuticals in the Environment

During the past three decades, research on the impact of chemical pollution has focused almost exclusively on the conventional priority pollutants, i.e. persistent organic pollutants (POPs). Today these substances are less relevant for many first world countries because emissions have been substantially reduced through the adoption of appropriate legal measures and the elimination of many of the dominant pollution sources. The focus has consequently switched to compounds present in lower concentrations but which nevertheless might have the ability to cause harm (Larsen et al., 2004). These substances are pharmaceutically active compounds (PhACs), personal care products (PPCPs) and endocrine disrupting chemicals (EDCs). These chemicals are introduced into sewage to a high extent by households. Recently their occurrence on municipal sewage treatment plants (STPs) and the receiving water has been reported in many countries. Since pharmaceuticals are not totally removed during sewage treatment they are discharged in appreciable quantities into receiving waters through sewage plant effluents (Ternes et al., 2005). A recently published article gives further insight about their routes into the environment, their occurrence, their removal in waste water treatment plants (Ternes et al., 2004). The problem of pharmaceutical drinking water pollution is described by Jones et al. (2005). As the most important foodstuff our drinking water is already in places contaminated with pharmaceuticals there is an increasing widespread consensus that this kind of contamination might require legislative action sooner or later (Petrovic et al., 2003). The cooperation between science and decision making is strongly recommended (Quevauvill et al., 2005].

This reveals that the topic of pharmaceuticals and endocrine disruptors in the environment is an evolving issue and asks for further data evaluation strategies with the aim to support the decision making processes.

2. Literature study on 12 selected pharmaceuticals in environmental media

That is the reason why we performed a comprehensive literature survey for seven environmental targets and the presence of 12 selected drugs therein. We chose the following 12 drugs for an intensive literature study: Bezafibrate (lipid regulator), Carbamazepine (anti epileptic), Clofibric acid (blood lipid regulator), Diclofenac (anti rheumatic), Diazepam (psychiatric drug), Fenofibrate (lipid regulator), Ibuprofen (analgetic), Metoprolol (beta blocker), Phenazone (analgetic), Ethinyl Estradiol (sex hormone, steroid), Roxithomycine (antibiotic), Sulfamethoxazole (antibiotic). The contraceptive Ethinyl Estradiol proved to be an endocrine disruptor. We performed the literature study early 2004 looking for current, that is to say articles published in the time period 2000-2004. We found 75 articles in international scientific journals which formed the basis for the evaluation approach. Attributes to be looked upon were the availability of the pharmaceuticals in the media surface water (SF), groundwater (GW), drinking water (DW), sewage sludge (SS), wastewater (WW), soil (SO) and sediment SE). The consideration of the environmental media in any publication is coded by 0 (not available) or 1 (available). Hence a 75 (objects) x 7 (attributes)

data-matrix with the entries c_{ij} (i=1,...75, j=1,..7)) results.

$$c_{ij} = \begin{cases} 1 \textit{ if the target } j \textit{ and one of the } 12 \textit{ drugs is discussed in the } i^{th} \textit{ publication} \\ 0 \textit{ else} \end{cases}$$

The data-matrix has already been evaluated with the main focus on the comparison of different multi-variate evaluation methods (Voigt et al., 2005).

3. Multi-Criteria Decision Tool: Hasse Diagram Technique (HDT)

3.1 Background and software

One consequence of our information society is an enormous increase in available data, with many people attempting to gain as much information as possible from these data, especially for comparative evaluations and related decision processes. Various tools are available to support these activities. However, a common difficulty with the evaluation step is that many of the methods mask and aggregate the data, and therefore both valuable information and transparency are lost. This translates to potential impacts on acceptance of the final decision. The Hasse Diagram Technique presents a rather new approach based on partial ordered sets that can be used to avoid the merging of data and thus preserve important elements of the evaluation and decision-making processes (Criterion, 2005).

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, Welzl, 2002). A comparison of the Hasse Diagram Technique with multi-variate statistical methods is given by Voigt et al. (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.

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 from the second author (Brüggemann et al., 1999) and is available for non-commercial use from the second author. The mathematical basis is described in Brüggemann et al. (2001).The commercial software is called ProRank Software for multi-criteria evaluation and decision support. For commercial applications it is recommended to contact the company Criterion – Evaluation and Information Management (Criterion, 2005).

Further theoretical developments concerning order theoretical tools in environmental sciences and their applications are discussed in regularly held workshops. For further information and for obtaining proceedings volumes please contact one of the authors.

3.2 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. 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. 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 || 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 $\mathbf{q}(x)$, $\mathbf{q}(y)$)
   - $x \leq y$ or $x \geq y$ (comparability)
   - x || 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 fulfills 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 (Brüggemann and Welzl, 2002).

Sometimes it is useful to refer to the quotient set, which is induced by the equivalence relation of equality, *R* (see for details: Brüggemann and Bartel, 1999). As usual we write *E/R* for the quotient set, and (*E/R*, *IB*) for the partially ordered quotient set.

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 ecotoxicity of environmental chemicals by lethal concentrations i.e. LC50 values for example, the orientation is the other way round. Here the following situation arises:

small values: "good", relatively unhazardous.

large values: "bad", relatively hazardous.

Concerning the evaluation of publications on pharmaceuticals with respect to environmental media the value 1 means available information, hence "good", the value 0 means information unavailable, hence "bad".

The total number of comparabilities I and incomparabilities U and their local analogues (i.e. the no of comparabilities V(x) and incomparabilities U(x) of a certain element x are useful quantities for the documentation of the Hasse Diagram and for the estimation of ranking uncertainties (Brüggemann and Welzl, 2002).

3.3 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. For further reading we refer to background publications by Brüggemann (2002, 1999).

4. Evaluation of the Publications Comprising Information on Environmental Contamination of Selected Pharmaceuticals

4.1 Evaluation by Hasse Diagram Technique (Complete Data-matrix)

The Hasse Diagram for the 75x7 data-matrix is calculated with the commercial program package ProRank (Criterion, 2005) and is shown in Fig. 1.

The Hasse Diagram in Figure 1 is structured into six levels. Some objects are denoted by K which means they comprise several objects (so-called equivalence classes) and other objects are called P which means publication number x. By equivalence relations there is a considerable reduction of the size of the set of objects: From 75 publications the evaluation by the criteria mentioned above a quotient set results, which has only 28 elements (classes). The organization into six levels shows that a rather good separation into different degrees of occurrences in the publications appears.

Furthermore the diagram shows one maximal object K10 $\in$ E/R which means that there are no publications dominating the equivalent publications P26, P31, P38, P74. As K10 is connected with all other objects, this implies that these four publications have more information than all other objects. These articles are review articles which cover many environmental media. We can detect the following six minimal objects K1, K5, K11, K12, P35 and P49; four of them form equivalence classes. The information in these articles is very small and/or specialized. Taking a look at all objects (quotient set) 258 incomparabilities, 120 comparabilities are counted, this means that there are more than twice as many incomparabilities than comparabilities.

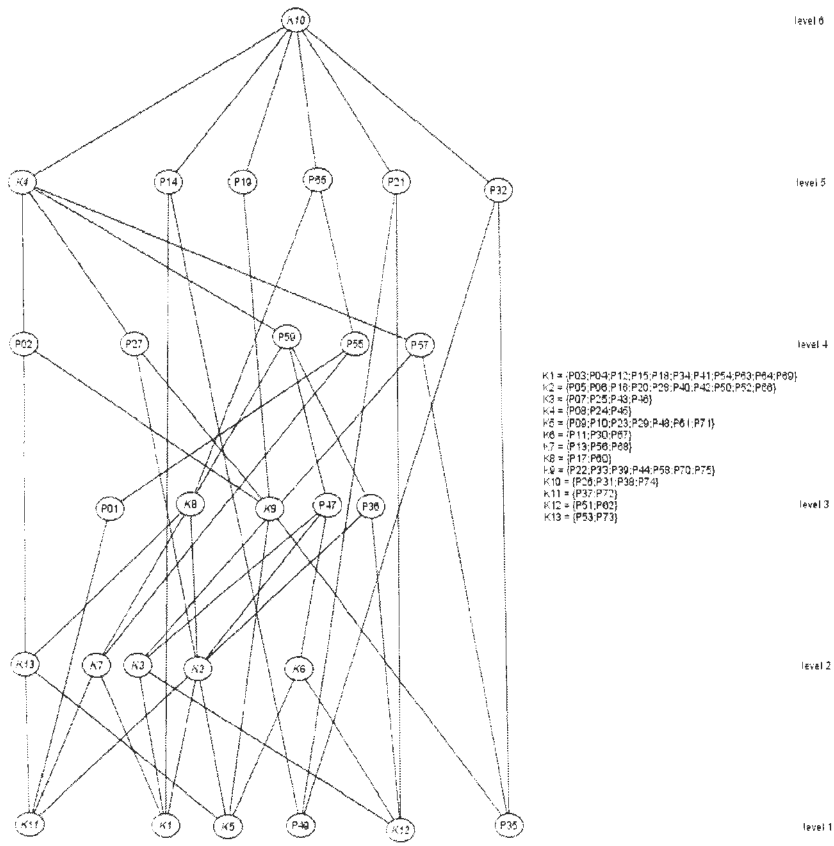

Figure 1. Hasse Diagram of 75 publications comprising pharmaceuticals evaluated by 7 environmental media

Although some important ranking information can be drawn out of the diagram, the figure calls for further data-analysis methods in order to draw more comprehensive conclusions from the data-matrix. In a recent publication the data reduction methods, Partially Ordered Scalogram Analysis with Coordinates

(POSAC) and Principle Component Analysis (PCA) were applied (Voigt et al., 2005). In this paper we follow the methodologies which comprise partial order tools like the W-Matrix and the consequentely followed reduction of attributes.

4.2 Study of the Influence of Attributes: W-Matrix

The dissimilarity matrix, also called the W-matrics Dissimilarity-matrix describes the influences of the attributes on the Hasse diagram. It is an important aspect of the data analysis by partial ordered sets. It contains the mutual comparisons of all Hasse diagrams. The larger the matrix-entries are, the greater is the difference between the Hasse diagrams. The W-matrix is calculated for all objects. The rsults are given in Table 1.

Table 1. W-Matrix entries for 75x7 data-matrix

NO.	LEAVING OUT ATTRIBUTE	NUMBER OF CHANGES
1	Drinking water	301
2	Ground water	558
3	Waste water	507
4	Sediment	42
5	Soil	123
6	Sewage sludge	275
7	**Surface water**	**622**

It can be demonstrated that many differences exist between leaving out one or the other criterion. If we leave out the attribute "surface water" 622 changes in the corresponding diagram can be found. The other extreme is the criterion "sediment". Its omission leads to only 42 changes in the diagram. This means that the criterion surface water is the most important one whereas the attribute sediment is relatively unimportant for this applied data-analysis method. This fact that the omission of the attribute "surface water" leads to significant changes in the diagram is proven in Figure 2.

Comparing this diagram where the attribute "surface water" is left out with Figure 1 which shows the complete information base IB, one can immediately see many differences in the diagrams. First of all only one minimal object K2 (again an equivalence class) is given in Figure 2. The original diagram showed 6 minimal objects. Taking a look at the comparabilities and incomparabilities of the quotient set we have 109 incomparabilities and 81 comparabilities now.

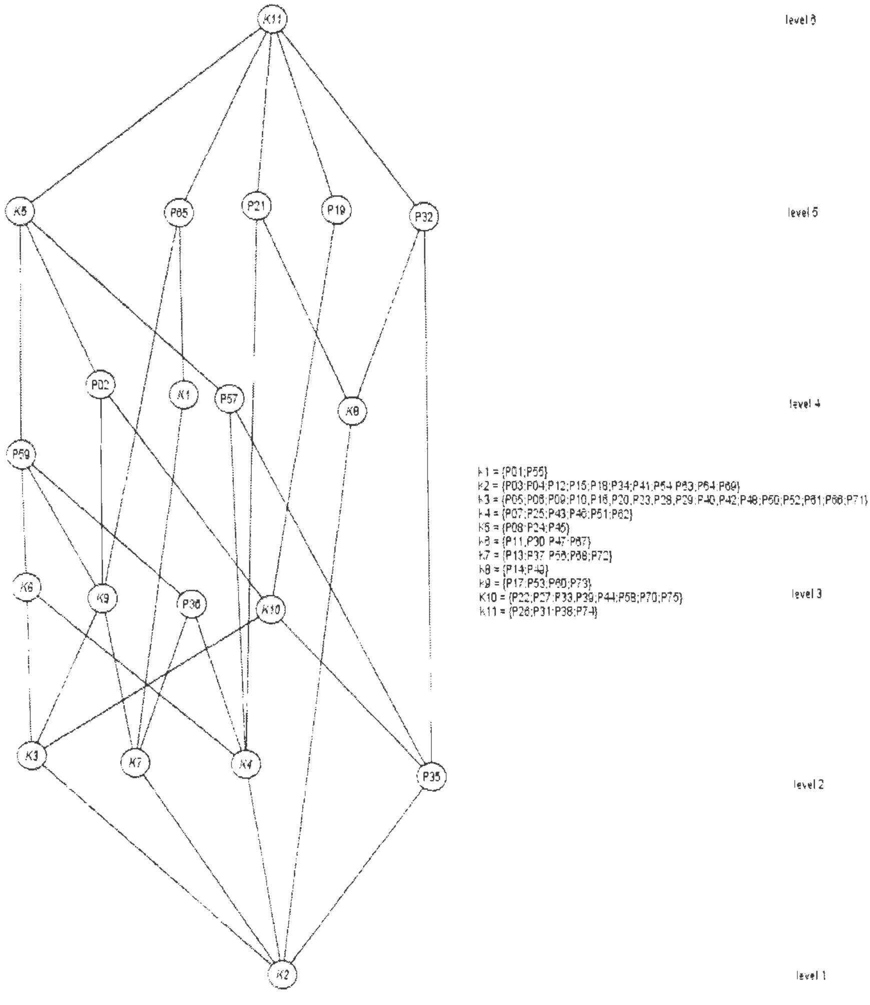

Figure 2. Hasse Diagram of 75 publications comprising pharmaceuticals evaluated by 6 environmental media (leaving out "surface water"

The attributes "soil" and "sediment" also indicate a special position in the dissimilarity matrix analysis. So it is of interest to take a look upon the publication situation for these two environmental targets. The ProRank program can easily omit attributes (as well as objects). In this case we set up a Hasse diagram for the two attributes only; hence we take a look at a 75x2 data-matrix. In Figure 3 the bar diagram of this procedure is shown. In such a diagram the attributes are shown so one can easily explain the partial order.

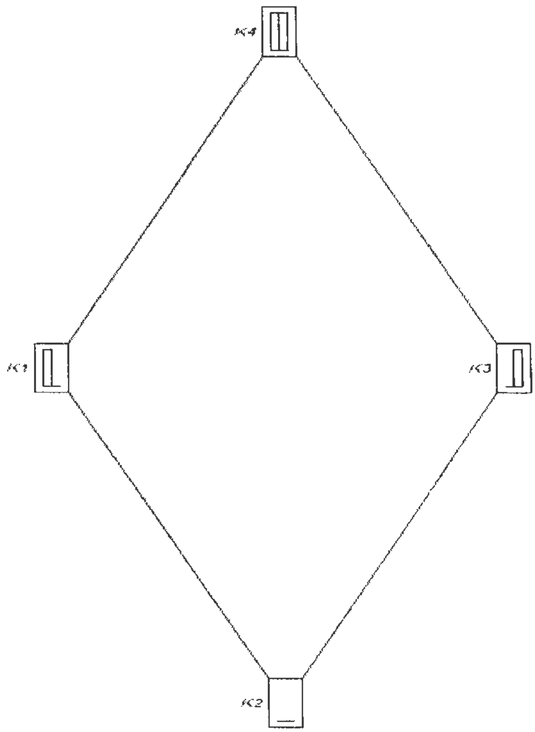

Figure 3. Hasse Diagram of 75 publications comprising pharmaceuticals evaluated by the media "sediment" and "soil"

The result is given in this very concise diagram. Only 4 objects are shown, all of them are equivalence classes. The equivalence class K4 is the maximal object P26, P31, P32, P38, P74.

It can easily be seen that this object has entries in both attributes; this means that the artibles comprised in K4 contain information on pharmaceuticals in soil and in sediment. The equivalence class K2 is the minimal object. It comprises the the following publications: P02, P03, P04, P05, P06, P07, P08, P09, P10, P11, P12, P13, P15, P16, P17, P18, P20, P22, P23, P24, P25, P27, P28, P29, P30, P33, P34, P35, P36, P37, P39, P40, P41, P42, P43, P44, P45, P46, P47, P48, P50, P51, P52, P53, P54, P56, P57, P58, P59. The equivalence class K1 comprises an entry in soil, whereas K3 has an entry in sediment. This means that they are comparable to K4 but incomparable to each other. The equivalence class K2 has no entry at all. Most of the set of 75 publications, namely 49 have neither information on the selected pharmaceuticals in soil nor in sediment. This is a very simple example which is mainly used to explain the partial ordered set background.

4. Conclusions and Outlook

In the paper the subject of the occurrence of pharmaceuticals in environmental media is envisaged. Although there is a great deal of uncertainty concerning possible detrimental effects on the aquatic ecosystems and soil ecosystems, the precautionary principle – or possibly new scientific evidence – may give rise to more stringent demands on wastewater treatment in the future (Larsen et al., 2004). In order to support the enlightenment about this subject we performed a mult-criteria evaluation analysis – called Hasse Diagram Technique – which derives from Discrete Mathematics on a data-matrix elaborated in an intensive literature study. The partial order ranking method Hasse Diagram Technique reveals among other aspects the maximal and minimal objects, that is to say the most and least important publications for the occurrence and detection of 12 pharmaceuticals in different environmental media. In a further evaluation study, the dissimilarity matrix is introduced. The media "surface water" and "soil and sediment" have a special position. The data-matrix is reduced in two ways, first the most important attribute "surface water" is left out. In a second diagram only the attributes "sediment" and "soil" are looked upon. To conclude, we can state among other aspects that the evaluation of the literature study revealed the most and least encompassing publications concerning drugs and their presence/absence in 7 menvironmental media in the recent time period. Additionally, it revealed the lack of data in the media "sediment" and "soil". Further research should be initiated in these fields. The applied Hasse Diagram Technique proved to be used as a multi-criteria decision support tool. The user can not only find out the most and least important objects. The possibility is given to leave out attributes (as demonstrated in this paper) or objects. Furthermore the possibility for weighting attributes is provided. This method is named Method of Evaluation by Order Theory (METEOR). In a recent publication Voigt et al. (2004b) demonstrated the method on the evaluation of environmental and chemicals databases with environmental parameters.

Concerning our future research we want to take a deeper look at the medium "drinking water". By applying the Method of Evaluation by Order Theory (METEOR) we can weigh attributes, in this case drinking water. So far, comparatively little study on the potential of pharmaceuticals to enter potable supplies was initiated and/ or published (Jones, 2005). Furthermore we believe that is is absolutely necessary to scrutinize the data availability on the chosen pharmaceuticals in Internet databases in order to enlarge the knowledge about the important subject We already began with the evaluation of Internet databases concerning their availability on environmental data on the chosen 12 pharmaceuticals and endocrine disruptors.

References

Brüggemann R., Bartel H.-G. (1999) A theoretical concept to rank environmentally significant chemicals. *J. Chem. Inf. Comput. Sci.*, **39**, 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**, 170-178.

Brüggemann R., Halfon E., Welzl G., Voigt K., Steinberg C. (2001) Applying the concept of partially ordered sets on the ranking of near-shore sediments by a battery of tests. *J. Chem. Inf. Comp. Sci.*, **41**, 918-925.

Brüggemann R, Welzl G. (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, Shaker-Verlag, Aachen.

Criterion (2005) WHASSE Software, http://www.criteri-on.de/StartseiteEng.html.

Halfon E., Reggiani M.G. (1986) On ranking chemicals for environmental hazard. *Environ. Sci. Technol.*, **20**, 1173-1179.

Jones O.A., Lester J.N., Voulvoulis N. (2005) Pharmaceuticals: a threat to drinking water?. *TRENDS in Biotechnology*, **23**, 163-167.

Larsen T.E., Lienert J., Josse A., Siegrist H. (2004) How to avoid pharmaceuticals in the aquatic environment. *Journal of Biotechnology*, **113**, 295-304.

Petrovic M., Gonzalez S., Barcelo D. (2003) Analysis and removal of emerging contaminants in drinking water. *Trends in Analytical Chemistry*, **22**, 685-696.

Quevauville P., Balabanis P. et al. (2005) Science-policy integration needs in support of the implementation of the EU Water Framework Directive. *Environmental Science & Policy*, **8**, 201-211.

Ternes T., Bonerz M. et al (2005) Determination of pharmaceuticals, iodinated contrast media and musk fragrances in sludge by LC/tandem MS and GS/MS. *Journal of Chromatography A.*, **1067**, 213-223.

Ternes T., Joss A., Siegrist H. (2004) Scrutinizing pharmaceuticals and personal care products in wastewater treatment. *Envl Sci. & Technol.*, **15**, 392A-399A.

Trotter W.T. (1992) *Combinatorics and Partially Ordered Sets Dimension Theory.* The John Hopkins University Press, Baltimore, Maryland, 1-307.

Voigt K., Welzl G., Brüggemann R. (2004a) Data analysis of environmental air pollutant monitoring systems in Europe. *Environmetrics*, **15**, 577-596.

Voigt K., Brüggemann R., Pudenz S. (2004b) Chemical databases evaluated by order theoretical tools. *Anal Bioanal Chem.*, **380**, 467-474.

Voigt K., Brüggemann R., Pudenz S., Scherb H. (2005 accepted) *Environmeantal Contamination with Endocrine Disruptors and Pharmaceuticals: An Envrionmetrical Evaluation Approach*, in: EnviroInfo 2005, Brno.

Jan Studzinski, Ludosław Drelichowski, Olgierd Hryniewicz
(Editors)

APPLICATIONS OF INFORMATICS IN ENVIRONMENT ENGINEERING AND MEDICINE

The purpose of the present publication is to popularize applications of informatics in environment and health engineering and protection. Runned papers are thematically chosen from the works presented during the conference *Multiaccessible Computer Systems* (*Komputerowe Systemy Wielodostępne*) that has been organized by the Systems Research Institute and University of Technology and Agriculture of Bydgoszcz for several years in Ciechocinek. Problems described in the papers concern quality management of the surface waters and the atmosphere, application of the mathematical modelling in environmental engineering, and development of computer systems in health and environmental protection. In several papers results of the research projects financed by the Polish Ministry of Science and Information Society Technologies are presented.

ISBN 83-89475-04-9
ISSN 0208-8029