Multicriteria Ordering and Ranking: Partial Orders, Ambiguities and Applied Issues
Jan W. Owsiński and Rainer Brüggemann, Editors

PYHASSE a New Software for Partially Ordered Sets: Ranking Soil against Needles

Kristina Voigt*, Rainer Brüggemann, Karl-Werner Schramm*, Manfred Kirchner***

* Helmholtz Zentrum Muenchen - German Research Center for Environmental Health (GmbH), Institute of Biomathematics and Biometry, Neuherberg, Germany, (kvoigt@helmholtz-muenchen.de, schramm@helmholtz-muenchen.de, kirchner@helmholtz-muenchen.de)

** Leibniz- Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany, (brg_home@web.de)

In complex environmental data-sets it is often necessary to compare different sets of criteria (attributes). In the similarity analysis which is part of the theory of the Hasse Diagram Technique (HDT) we intend to calculate the similarity of different posets. This similarity analysis is an important feature of the newly developed software packages called PYHASSE developed by the second author. As an example we use a data-matrix which was generated in an international project named MONARPOP (Monitoring Network in the Alpine Region for Persistent and other Organic Pollutants) in which selected chemicals in different environmental media were analysed in the years 2004 and 2005 (MONARPOP, 2008). 17 pesticides were chosen and analyzed in soil and needle samples in the alpine regions of Germany, Austria, Switzerland, Italy, and Slovenia. The samples were taken at different heights. We perform a similarity analysis of the two different targets in order to get an idea if the target type has a great impact on the ranking of the pesticides and hence on the contamination. It can be demonstrated that the different target type: soil or needles have a considerable influence on the structure of the Hasse Diagrams and hence on the intensity of the pollution load. This supports the analytical analysis of soil as well as of needle samples.

Keywords: Hasse diagram technique, partial order theory, Hasse software, PYHASSE, similarity analysis, environmental chemicals, pesticides, POPs

1. Introduction: Hasse diagrams and its software

Partial order is a discipline of Discrete Mathematics and one may consider partial order as an example of mathematics without arithmetic. A good overview can be found in a book edited by the second author (Brüggemann, 2006). The graphical representation of partial orders is laid down in so-called Hasse Diagrams. The first software was already written in the 1980s and was developed under MS-DOS. In the 1990s the software was adapted to the MS-Windows platform. The functionalities were constantly enhanced and improved in the following years (Brüggemann, 2005). The innovative tool called METEOR (Method of Evaluation by Order Theory) which attempts to resolve the incomparabilities among objects by inclusion of external knowledge is incorporated in the WHASSE program (Brüggemann, 2008; Voigt, 2006, 2007, 2008a). This software named WHASSE, written in Delphi has always been available for scientific purposes from the second author free of charge. However, new features (e.g. like the similarity analysis or the fuzzy partial order, van de Walle, 1995) made it necessary to develop a new software written in PYTHON (http://www.python.org/) being more flexible than the older software. An extensive description of PYHASSE is given by Brüggemann et al. in this issue.

In complex environmental data-sets it is often necessary to compare different sets of criteria (attributes). In the similarity analysis we intend to calculate the similarity of different posets. This similarity analysis is an important feature of the newly developed software package called PYHASSE. PYHASSE is elaborated by the second author and is still under development. Test versions are, however, available from R. Brüggemann.

2. Similarity analysis in PYHASSE

2.1 Evaluation by Hasse Diagram technique

Consider the two partially ordered sets (posets) ({a,b,c}, $<_1$) and ({a,b,c},$<_2$) (generally (P,<), P the object set), whose Hasse diagrams are shown in Fig. 1.

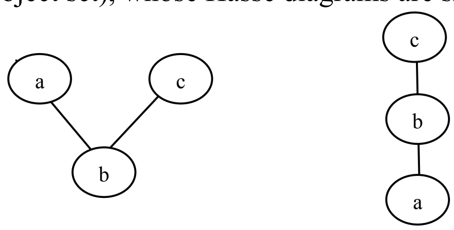

Figure 1. Two Hasse diagrams to demonstrate the similarity analysis

There is one contradiction, namely a ≤ b in the Hasse diagram (rhs) and b ≤ a in the other Hasse diagram (lhs). There is one coincidence, namely c ≥ b. A matrix-like scheme gives a comprehensive view:

	a	b	c
a	-	><	||<
b	<>	-	<<
c	||>	>>	-

Beside the diagonal which is not of interest, the entries are found by comparing the row defining element with the column defining one. The first symbol belongs to the lhs-, the second to the rhs-Hasse diagram. Most important are the entries like >> or <<, which are counting the 'isotone' character of both partial orders and the entries like ><, <> which contribute to the 'antitone' character, i.e. to the conflicts between the two partial orders. Following Rademaker (this book): two posets are in conflict or "contradict each other" (are antitone) on two objects x, y ∈ P ,"if we have $x <_1 y$ and $y <_2 x$ or $y <_1 x$ and $x <_2 y$". There are still more combinations to look upon: <||, >||, ||>,||< , =||, ||= or || || are considered as indifferent, combinations like > =, < =, = <, = > are called weak isotone. Finally the entry of type = = contributes to equivalence relations. There is an inherent symmetry, so that it is sufficient to examine only the upper or lower triangular part of the matrix.

In the example above and respecting the symmetry we count: 1 isotone, 1 antitone and 0 weak isotone, 1 indifferent and 0 equivalence relations, thus we characterize the similarity by a tuple (#isotone, #antitone, #weak isotone, #indifferent, #equivalences), where # is the symbol for counting the relations, like. One may also be interested in which object is mainly contributing to conflicts, i.e. to the antitone character of the two partial orders. In the case above the elements a and b contribute once, whereas c does not have any counter current entry.

2.2. Background and description of selected data sets

In our further presentation we will give an example applying the software tool for environmental decision support PYHASSE based on the theory of partially ordered sets briefly described above. This environmental decision support tool is now applied on the ranking of environmental pollutants found in soil and needle samples in the Alps. The increasing worldwide contamination of environmental compartments with thousands of industrial and natural chemical compounds is one of the key environmental problems facing humanity (Schwarzenbach, 2006). Although most of these compounds are present at low concentrations, many of them raise considerable toxicological concerns, particularly when present as components of complex mixtures. We frame the concerns primarily from an environmental-protection perspective with a focus on soil and forest ecosystems in the Alps, but without neglecting the human health issues. Protecting natural waters and soils against chemical pollution protects soil and aquatic life and thus, directly or indirectly, human health. Hence, any measures taken to prevent the chemical pollution of surface and groundwater resources as well as soils will not only improve ecosystem health, but will also benefit both the production of clean water and safe food for human consumption (Schwarzenbach, 2006).

Especially the so-called POPs (Persistent Organic Pollutants) pose an immense threat to the environment and human health. The Stockholm Convention on Persistent Organic Pollutants is a global treaty to protect human health and the environment from chemicals that remain intact in the environment for long periods, become widely distributed geographically and accumulate in the fatty tissue of humans and wildlife. Exposure to Persistent Organic Pollutants (POPs) can lead to serious health effects including certain cancers, birth defects, dysfunctional immune and reproductive systems, greater susceptibility to disease and even diminished intelligence. Given their long range transport, no one governing acting alone can protect its citizens or its environment from POPs. In response, the Stockholm Convention, which was adopted in 2001 and entered into force 2004, requires Parties to take measures to eliminate or reduce the release of POPs into the environment. The Convention is administered by the United Nations Environment Programme and based in Geneva, Switzerland
(http://chm.pops.int/Convention/tabid/54/language/en-US/Default.aspx).

Persistent Organic Pollutants (POPs) have been monitored and analyzed within the framework of an EU-project entitled MONARPOP (Monitoring Network in the Alpine Region for Persistent and other Organic Pollutants) during the years 2004 and 2005. MONARPOP monitors POPs and other organic pollutants with respect to their long-range transport to remote alpine regions, prevalent source directions, loads within the alpine range, including regional differences, variation with altitude, present stocks, bound in forests of the alpine region, possible biological effects of the detected loads. The project further aims to provide information to decision makers. In the future, the evidence gathered during MONARPOP may help to assess the success of the Stockholm POP convention (MONARPOP, 2008).

Sampling sites are located in Austria, Germany, Italy, Slovenia and Switzerland. The network consists of 33 standard sites and 7 altitude profiles. A standard site is located at 1400m ± 150 a.s.l. An altitude profile is a regularly spaced series of 4–5 plots at increasing height. Soil from mountainous forests is analysed for its pollutant content. The humus layer consists of organic matter and accumulates POPs both from direct deposition and from litter fall. Conifer needles are frequently used as bioindicators. (Bioindication is a way to monitor environmental conditions, especially pollution, through their impact on living organisms) (MONARPOP, 2008).

2.3. Similarity analysis of soil samples versus needles samples

17 pesticides were chosen and analyzed in soil and needle samples in Germany, Austria, Switzerland, Italy, and Slovenia. The samples were taken at different heights. In this approach we did not regard different heights of samples but different target types (needles and soil). Our aim is to get an idea if the target type has a great impact on the ranking of the pesticides and hence on the contamination.

Table 1. Pesticides analyzed (trivial name, acronym)

Common Name	Trivial Name	Acronym
Alpha Hexachlorocyclohexane	alpha-HCH	AHCH
Beta Hexachlorocyclohexane	beta-HCH	BHCH
Gamma Hexachlorocyclohexane	gamma-HCH	GHCH
Delta Hexachlorocyclohexane	delta-HCH	DHCH
Epsilon Hexachlorocyclohexane	epsilon-HCH	EHCH
p,p’ Dichloro Diphenyl Trichloroethane	p,p’-DDT	PPDT
o,p’ Dichloro Diphenyl Trichloroethane	o,p’-DDT	OPDT
p,p’-Dichloro Diphenyl Dichloroethane	p,p’-DDD	PPDD
o,p’-Dichloro Diphenyl Dichloroethane	o,p’-DDD	OPDD
p,p’-Dichloro Diphenyl Dichloroethane	p,p’-DDE	PPDE
o,p’-Dichloro Diphenyl Dichloroethane	o,p’-DDE	OPDE
Aldrin	Aldrin	ALDR
Dieldrin	Dieldrin	DIEL
Endrin	Endrin	ENDR
Mirex	Mirex	MIRE
Hexachlorobenzene	HCB	HCBE
Heptachlor	Heptachlor	HECL

Table 1 lists the objects to be analyzed. Most of the chosen chemicals belong to the group of Persistent Organic Pollutants (POPs). The initial twelve POPs chemicals, known as 'poisons without passports', pose particular hazards because of their common characteristics. They are toxic to humans and wildlife. They are persistent and resist breaking down. POPs are semivolatile and mobile, travelling great distances on wind and water currents and are widely distributed throughout the environment. Through global distillation, they travel from temperate and tropical regions to be deposited in the colder regions of the poles (IPEN, 2005) and as demonstrated in the MONARPOP project in the Alps.

The ranking method by Hasse Diagram Technique (HDT) applied here in the first place identifies the comparabilities and incomparabilities among chemicals and demonstrates which chemicals are better or worse than the others.

Two different data-sets are analyzed: 17 pollutants in 81 soil samples in all five countries (17 x 81) and in 92 needle samples (17 x 92). The question to answer is the following: Are the two different resulting Hasse Diagrams similar and to which degree? The Hasse Diagrams are given in Fig. 2a (17x81) and Fig. 2b (17x92).

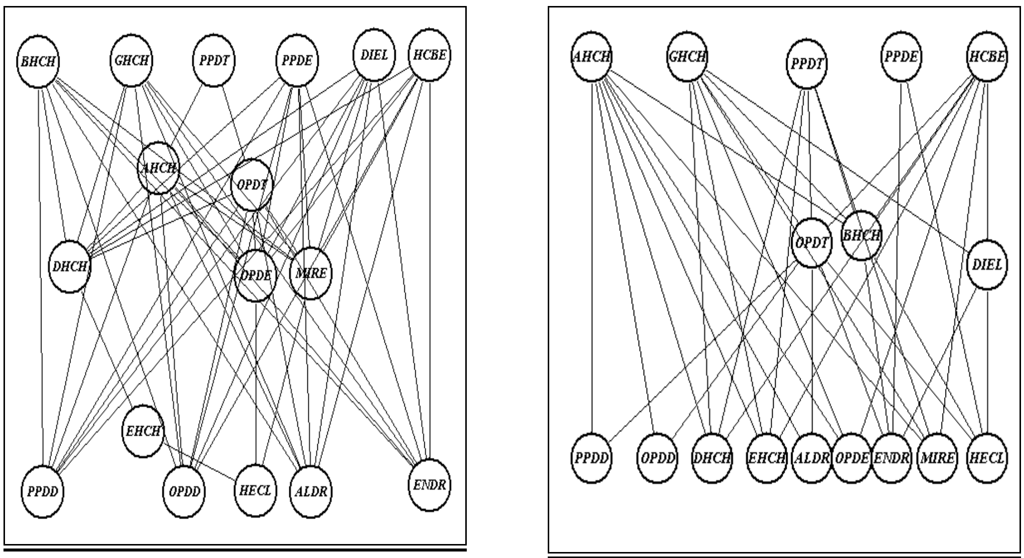

Figure 2. HD 17 pesticides x 81 soil samples (lhs) HD 17 pesticides x 92 needle samples (rhs)

Both Hasse Diagrams show quite a few differences. The diagram of the 17x81 data-matrix (lhs) has five levels whereas the HD of the 17x92 data-matrix shows only three levels. Concerning the maximal objects (on top of the diagrams) great differences are found not only in the number but also in the chemicals. Taking a look at the minimal objects (at the bottom of each diagram) we also encounter great differences in the number of minimal objects as well as in the chemicals. The differences are also given in the comparabilities and incomparabilities. It is obvious that the two HDs are not similar. However, we want to quantify this optical inspection by a tuple of characteristics of similarity as discussed before (see Section 2.1).

The similarity analysis should give some more insight into these differences. So the two data-matrices 17x81 (soil) and 17x92 (needles) are now subject to the similarity analysis.

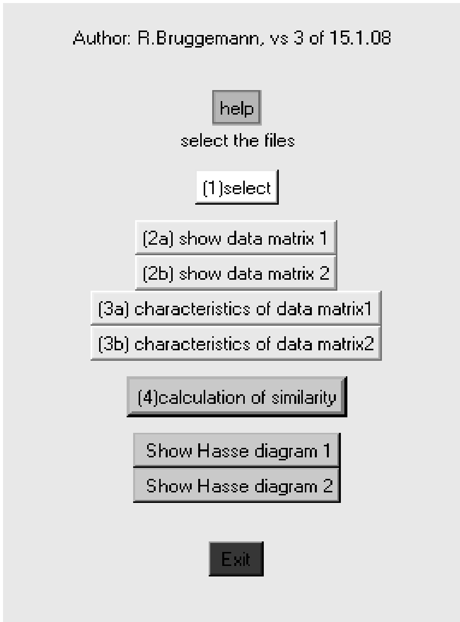

Figure 3. PyHasse similarity analysis

Fig. 3 gives a screenshot of the User interface of the Python-program "similarity" performing the similarity analysis. Two files (17x81) and (17x92) have to be uploaded. Both Hasse Diagrams (see Fig. 2 a and b) can be displayed as well as their initial data-matrices. In Fig. 4 the results of the similarity analysis are given.

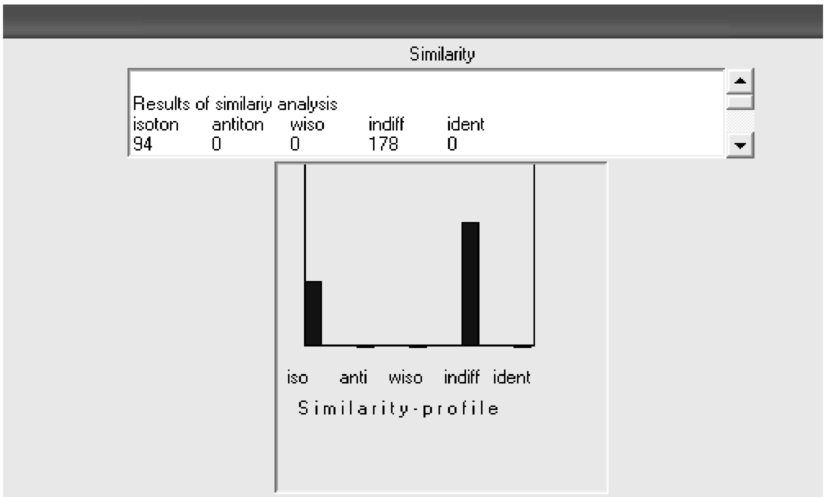

Figure 4. Results of similarity analysis of two data-sets (17x81 and 17x92) soil / needles

Only isotone and indifferent relations are calculated. 94 isotone relations are given and 178 indifferent relations. This means that the different target type has a considerable impact on the structure of the Hasse Diagrams: Whereas isotone relations demonstrate a high degree of similarity of the two data-sets, the indifferent relations indicate combinations of comparabilities of the one Hasse Diagram with incomparabilities of the other one. From Fig. 2 it can easily be seen that that the "HD needles" has a great amount of incomparabilities. Therefore a pollution trend in soils (Fig. 2 lhs) with respect to one sampling site implies the same ranking in other soil sampling sites. In contrast an ordering between any two chemicals in (Fig. 2 rhs) will pretty often exhibit a reverse order in other needle sampling sites. We may hypothesize that in the target needles more processes are encountered than in soil. This means that the choice of different targets like soil and needles is more than justified.

3. Discussion and outlook

In the current data-analysis the test-sets of 17 pesticides in 81 soil samples (humus and mineral soil) and the same 17 pesticides in 92 needle samples in the

Alps are looked upon. Most of these chemicals belong to the so-called Persisted Organic Pollutants (POPs). POPs and other semivolatile organic compounds (SOCs) are ubiquitous and have been measured in snow and ice in high altitudes and latitudes in several approaches. Additionally, it is known from field and modelling studies that snowfall events and the presence of snow covering soil and vegetation influence the environmental distribution of chemicals. Therefore, it is important to understand how these chemicals interact with snow and ice and how they are transported to cold regions. The latter is especially important in the context of international regulation of chemical use to reduce transboundary pollution and to protect remote regions against chemical pollution (Stocker et al., 2007). Stocker and associates modelled the effect of snow and ice on the global environmental fate and long-range transport potential of semivolatile organic compounds. Another interesting modelling approach concerning the global fate of DDT has recently been published by Schenker et al. (2008). The global environmental fate model named CliMoChem was used to calculate concentrations of dichlorodiphenyltrichloroethane (DDT) and its degradation products in the environment. To this end, best available physico-chemical properties of DDT have been assembled, and a realistic DDT emission scenario covering the period from 1940 to 2005 has been generated. Results from the model are temporally and geographically resolved concentrations of DDT, DDE, and DDD in various environmental media. To confirm model results with measurements, the authors have developed a method for a qualitative and quantitative comparison of model and measurements. The agreement between the model and measurements is good, especially in the temporal dimension, and in the soil and air compartments. Using estimated DDT emissions for the future, the authors predict environmental concentrations in the next 50 years. The results show that, if emissions continue at a low level, concentrations will decrease by a factor of 30 in temperate regions and by a factor of 100 in the Arctic, as compared to the concentrations in the 1960s and 1970s. In the tropics, levels decrease by a factor of 5 to 10, only. Whereas environmental concentrations and estimated future emissions are at steady state after about 10 years in temperate and tropical regions, this takes over 50 years in the Arctic.

This is one of the reasons why measurements and data-analysis of these chemicals are extremely important and should be continued in the future.

In this paper we introduced a PYTHON program named PYHASSE which allows a deepened graph theoretical analysis of the directed graphs, generated by the partial order relation. For the programming language PYTHON, see for example Weigend (2006) or Lutz (2003). Using the newly developed feature of similarity analysis, we conclude that the two Hasse Diagrams cannot be considered as similar albeit they are not in conflict with each other because the main contribution of the similarity tuple is the indifference relation. This means that not only soil samples but also different targets like needles should be monitored with respect to POPs in the Alpine regions.

This result is different from a previously performed data-analysis where we took a closer look at the two different soil sample types, namely the humus soil and the mineral soil layer. There we could determine that the type of soil only showed some indifferent relations but did not have a great impact on the Hasse Diagram. This means that the pesticides are found not only at the first soil layer, the humus layer, but also in considerably comparable concentrations in the mineral soil layer which is situated underneath the humus layer (Voigt, 2008b).

Other methods like e.g. the comparison of cluster analysis results are not considered in this paper as our focus lies on partial order. Further research studies with respect to the data-analysis by partial order of the data elaborated during the MONARPOP project will follow. We intend to apply different features of the newly developed software package PYHASSE, e.g. Fuzzy Partial Order.

To conclude we can state that there is a great opportunity for data-analytical tools like the introduced software PYHASSE to contribute to sustainability and thereby minimize the risk of further environmental pollution. Dissemination of information about environmental risks, along with opportunities for action, will be an area of increasing importance (Hepting, 2007).

References

Brüggemann R., Simon U., Mey S. (2005) Estimation of averaged ranks by extended local partial order models. *MATCH – Commun. Math. Comput. Chem*, **54**, 489-518.

Brüggemann R. and Carlson L. (Eds.) (2006) *Partial Orders in Environmental Sciences and Chemistry*. Springer-Verlag, Berlin.

Brüggemann R., Voigt K., Restrepo G., Simon U. (2008) Concept of Stability Fields and Hot Spots in Ranking of Environmental Chemicals. *Environmental Modelling & Software*, **23**, 1000-1012.

Hepting D.H. (2007) Decision support for local environmental impact assessment. *Environmental Modelling & Software*, **22**, 436-441.

IPEN (2005) IPEN Body Burden Community Monitoring Handbook, http://www.oztoxics.org/cmwg/chemicals/pop%20chem.html.

Lutz M. and Ascher D. (2003) *Learning Python*. O'Reilly Publisher, Beijing.

MONARPOP (2008) Monitoring Network in the Alpine Region for Persistent and other Organic Pollutants http://www.monarpop.at/.

Schenker U., Scheringer M., Hungerbuehler K. (2008) Investigating the global fate of DDT: Model evaluation and estimation of future trends. *Environ. Sci. Technol.*, **42**, 1178-1184.

Schwarzenbach R.P., Escher B.I., Fenner K., Hofstetter T.B., Johnson A.C., von Gunten U., Wehrli B. (2006) The challenge of micropollutants in aquatic systems. *Review. Science*, **313**, 1072-1077.

Stocker J., Scheringer M., Wegmann F., Hungerbuehler K. (2007) Modeling the effect of snow and ice on the global environmental fate and long-range transport potential of semivolatile organic compounds. *Environ. Sci. Technol.*, **41**, 6192-6198.

Van de Walle B., De Baets B., Kersebaum K.C (1995) Fuzzy multi-criteria analysis of cutting techniques in a nuclear dismantling project. *Fuzzy sets and Systems* **74,**115-126.

Voigt K., Brüggemann R., Pudenz S. (2006) A multi-criteria evaluation of environmental databases using the Hasse Diagram Technique (ProRank) software. *Environmental Modelling and Software*, **21**, 1587-1597.

Voigt K. and Brüggemann R. (2007) Data-Availability of Pharmaceuticals detected in water: An evaluation study by order theory (METEOR). In: Hryniewicz O., Studzinski J., Szediw A., *Environmental Informatics and Systems Research, Vol. 2: Workshop and application papers*, Shaker Verlag, Aachen, 201-208.

Voigt K. and Brüggemann R. (2008a) Ranking of Pharmaceuticals Detected in the Environment: Aggregation and Weighting Procedures. *Combinatorial Chemistry & High Throughput Screening* (accepted).

Voigt K., Brüggemann R., Schramm K.-W., Kirchner (2008b) M. PYHASSE a New Software Tool for Partially Ordered Sets Example: Ranking of contaminated Regions in the Alps. *EnviroInfo 2008*, Lüneburg, September 10-12, 2008, Metropolis-Verlag, Aachen, 631-636.

Weigend M. (2006) *Objektorientierte Programmierung mit Python*. mitp-Verlag, Bonn.