DETECTING SIMILARITIES OF SEMIVOLATILE ORGANIC CHEMICALS (SVOCs) IN DIFFERENT ENVIRONMENTAL AND HUMAN MATRICES

Voigt K[1]*, Bruggemann R[2], Scherb H[1] , Cok I[3], Schramm K-W[4,]

[1]Helmholtz Zentrum München – German Research Center for Environmental Health (GmbH) Institute of Computational Biology, Ingolstädter Landstr. 1, D-85764 Neuherberg, Germany, kvoigt@helmholtz-muenchen.de, scherb@helmholtz-muenchen.de
[2]Leibniz-Institute of Fresh Water Ecology and Inland Fisheries, Berlin, Germany brg_home@web.de
[3]Department of Toxicology, Faculty of Pharmacy, Gazi University, 06330, Ankara, Turkey, icetin@gazi.edu.tr
[4]Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Molecular EXposomics (MEX), Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany, schramm@helmholtz-muenchen.de
[5]TUM, Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt-, Department für Biowissenschaften, Weihenstephaner Steig 23, 85350 Freising, Germany

1. Introduction

It is well-known for ages that semivolatile organic chemicals (VOCs) do not only pollute the environment but also reach the human body and consequently act/react there. The concept of the exposome representing the totality of exposures received by a person during lifetime, encompasses all sources of toxicants and, therefore, offers scientists an approach for investigating the environmental causes of chronic diseases. In this context, it is appropriate to regard the "environment" as the body's internal chemical environment and to define "exposures" as levels of biologically active chemicals in this internal environment[1]. According to this, in order to explore the exposome, it makes sense to employ a top-down approach based upon biomonitoring (e.g. human breast milk samples) as well as a bottom-up approach that samples air, water, soil, needles, and so on.
The occurrence of Semi Volatile Organic Chemicals (SVOCs) in the Taurus Mountains in Turkey was studied recently[2,3]. Taurus Mountains were suggested because of their potential to act as a sink for organic pollutants by cold condensation and can, therefore reflect the atmospheric pollution in Turkey as well as neighbouring countries.
In this paper, we will compare the Semipermeable Membrane Devices (SPMD) (air sampling data) with human breast milk data for a selection of SVOCs. Both data sets have a reference to the lipophilic status. Furthermore, the environmental medium pine needles should be looked upon in comparative air sampling data. Human breast milk samples have already been compared with soil samples by Bruggemann et al.[4] A high degree of similarity could be demonstrated.

2. Materials and methods

2.1 Data matrix

The aim of our data evaluation approach is to find out whether there are conformities between the data sets to be compared. An appropriate data analysis method to identify such conformities and differences is the discrete mathematical Partial Order Technique (also called: Hasse diagram technique)[5]. The software package used is the PyHasse software written by Rainer Bruggemann. In this approach, we apply the basic Hasse Diagram Technique modules as well as the Similarity Analysis module for the comparison of two data matrices.
We examine the occurrence of 18 OCPs listed in Table 1 in a human medium, namely breast milk, as well as in environmental media, namely mountain pine needles and air samples. In Table 1 the 18 chemicals with their abbreviation and CAS-numbers are presented.

Table 1: 18 OCPs detected in human and environmental samples in the Taurus Mountains, Turkey

Nr.	Acronym	Name	CAS-Number
01	**AHCH**	alpha-Hexachlorcyclohexane	319-84-6
02	**BHCH**	beta-Hexachlorcyclohexane	319-85-7

03	**GHCH**	gamma-Hexachlorcyclohexane	58-89-9
04	**PECB**	Pentachlorobenzene	608-93-5
05	**HCBE**	Hexachlorobenzene	118-74-1
06	**PPDT**	p, p'-Dichlordiphenyltrichlorethane	50-29-3
07	**OPDT**	o, p'-Dichlordiphenyltrichlorethane	789-02-6
08	**PPDD**	p, p'-Dichlordiphenyldichlorethane	72-54-8
09	**OPDD**	o, p'-Dichlordiphenyldichlorethane	53-19-0
10	**PPDE**	p, p'-Dichlordiphenyldichlorethene	72-55-9
11	**OPDE**	o, p'-Dichlordiphenyldichlorethene	3424-82-6
12	**OXYC**	Oxychlordane	27304-13-8
13	**CHCE**	cis-Heptachloroepoxide	1024-57-3
14	**DIEL**	Dieldrin	60-57-1
15	**END1**	Endosulfan-1	959-98-8
16	**END2**	Endosulfan-2	33213-65-9
17	**MECH**	Methoxychlor	72-43-5
18	**MIRE**	Mirex	2385-85-5

These listed organochlorine chemicals forming the set *P*, are known to pose a serious threat to the environment and consequently to human health. Especially the endocrine disruptive potential of some of these chemicals should initiate action worldwide. In a review concerning the history of the discovery of the widespread toxicity of chlorinated hydrocarbons by Rosner and Markowitz[6], the authors conclude the enormous lag between identification of danger and ultimate regulation of these chemicals, which is still an enormous public challenge.

2.2 Application of Partial Order Technique Using PyHasse software

The easiest way to apply partial orders is to draw a Hasse diagram, which is based on the cover relations and which needs some additional heuristics in some cases, which however do not affect the partially ordered set *P*. The cover relations are deduced from the chemicals' profiles, i.e. the ordered sets of indicator values. The indicators are found from concentrations in human breast milk samples, from those of air borne materials, and of pine samples. The following data evaluation steps are performed:

a) Analyze why an object x is located in the Hasse diagram at a certain position and analyze this position in terms of the profile of x. This close interaction between the location of an object within a algebraic graph (in fact a Hasse diagram is a directed, acyclic, transitively reduced graph) and its profile can be formalized, see Wolski, 2004. However, here this formalization is not in the focus of our paper. In this context the identification of an object x being a

- maximal element (no other elements are covering x)
- minimal element (no other elements are covered by x)
- isolated element (x is at the same time a minimal and a maximal element)

is of main interest.

b) The next step is to consider x as an element of a chain. A suitable long chain (i.e. having many elements of *P*) is of interest, because a chain indicates that indicators are simultaneously not decreasing, when e.g. started at the bottom of the chain and proceed upwards ("vertical analysis").

c) Another step is to see x as an element of an antichain ("horizontal analysis"). Why is x not comparable with say y. What are the reasons causing this incomparability? Is the incomparability caused by only one pair of indicators or by several pairs and if yes, what is common for these indicators?

d) Objects may be considered to be elements of subsets of *P*, which are not related to each other by order relations or by few ones only. Is that the case, we speak of separated subsets.

2.3 PyHasse software: Similarity

In complex 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 (partially ordered sets). In the similarity analysis we calculate the proximity of different posets based on the same ground set. The outcome of a partial order for two objects a, b may be $a < b$, $a > b$, $a \parallel b$, $a = b$. When two partial orders $(G, \leq_1)$ and $(G, \leq_2)$ are to be compared,

then the combinations are counted such as: a $<_1$ b, a $<_2$ b, or a $<_1$ b, a $>_2$ b, or a $||_1$ b, a $<_2$ b for which we use the shorthand notation >>, ><, || etc. Most important are the entries like >> or <<, which are counting the 'isotone' character of both partial orders (ISO) and the entries like ><, <> which contribute to the 'antitone' character, i.e. to the conflicts between the two partial orders (ANTI). Two posets are in conflict or "contradict each other" (are antitone) on two objects x, y ∈ G," 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 (IND), combinations like > =, < =, = <, = > are called weak isotone (WISO). Finally the entry of type = = contributes to equivalence relations (IDE). The appropriately normalized contributions of ISO, ANTI, IND, WISO and IDE are the final result. It is left open how this similarity profile is mapped onto a single number, for example onto the Tanimoto index, because in different applications the importance of WISO, IND and IDE may vary.
Further descriptions and examples of the application of the similarity on an environmental health data set is given by [3,4] Voigt et al. and Bruggemann et al.. Whereas the similarity analysis is based on partial orders, the newly developed CombiSimilarity has its basis on the comparisons of the data matrices. Both variants lead to a distribution of quantities ISO, ANTI, etc., which can be further examined by appropriate statistical tests, such as the "t-test". The idea is, to eliminate the bare effect of a greater number of indicators, describing nevertheless the same context. This CombiSimilarity method is explained in detail by Bruggemann et al. [4].

3. Comparisons of Human / Environmental and Environmental / Environmental Samples

We now apply the basic Hasse diagram technique to the following data matrices:
18 OCPs in 44 human breast milk samples / 18 OCPs in 8 air samples (SPMD), 1 year
18 OCPs in 7 pine needle samples / 18 OCPs in 8 air samples (SPMD), 6 monsths beginning in May

In our current study we evaluate the same number of chemicals, namely 18 pesticides, in breast milk samples as well as in pine needles samples and air monitoring samples aiming to find out as to how far concentration profiles between the environmental samples and human samples and environmental pine needles samples and environmental air samples can be considered as similar.
In Figure 1 the Hasse diagrams of 18 OCPs in 44 human breast milk samples and 18 chemicals in 8 SPMD samples over the period of one year sampling are shown.

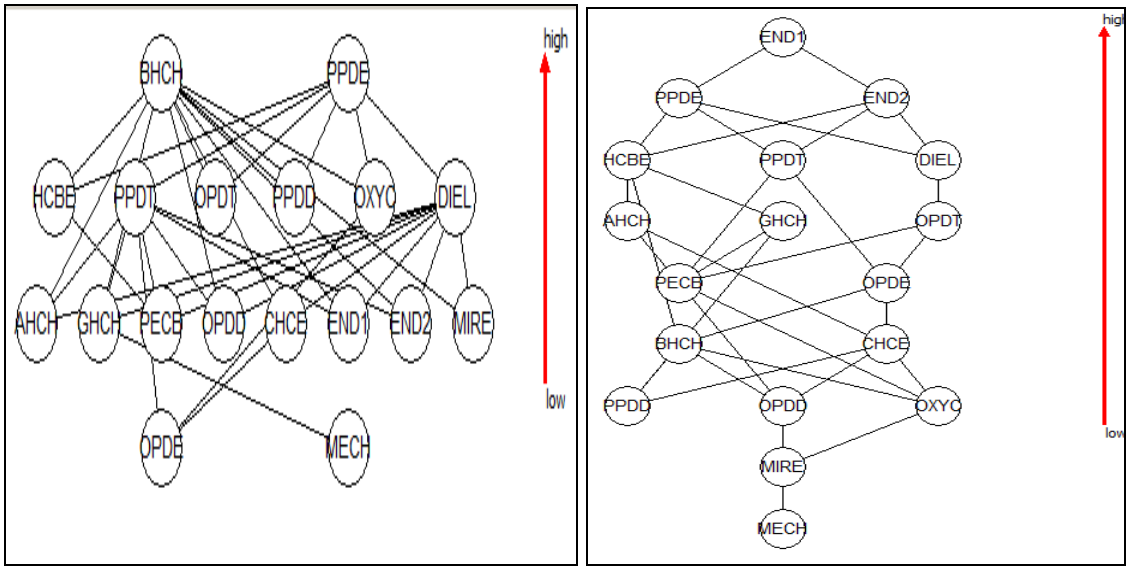

Figure 1: 18 OCPs detected in 44 human and in 8 air samples in the Taurus Mountains, Turkey

Comparing these two diagrams only few similarities can be detected, e.g. MECH is a minimal object in both cases. PPDE is in a high position in both diagrams. The position of END1 is remarkable. Whereas it is a minimal object in the left diagram, representing the pollution of breast milk samples, it is a maximal element in the air sampling approach. A further evaluation step is performed, namely the Similarity Analysis. This will be

performed for the two data sets in Figure 1 but also for the two environmental data sets (18 OCPs in 7 pine needle samples / 18 OCPs in 8 air samples (SPMD), 6 months beginning in May).

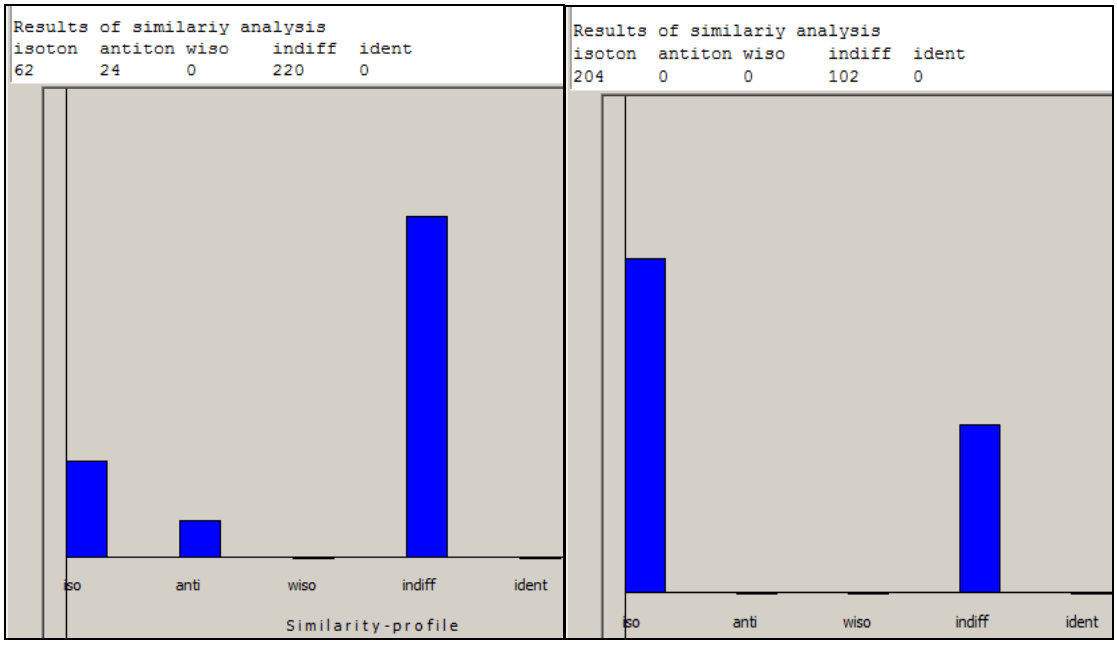

Figure 2: Similarity Graphs Human/Environmental and Environmental (needles) / Environmental (air)

In Figure 2 the different relations explained above are displayed. The isotone relations show a large degree of similarity. On the other side the indifferent relations reveal all relations with incomparabilities. Antitone relations mean that no similarity between the two data sets exists.

4. Results and discussion

From these two similarity approaches we can conclude that there exists some degree of similarities between human and environmental pollution. The similarity analysis reveals a larger degree of similarity between the two environmental data sets than between human and environmental media. Further studies must be initiated in order to underpin our findings, that OCPs pollute not only different environmental media in a similar way. Finally comparisons can be made between environmental pollution and the occurrence of these dangerous chemicals in human media, like breast milk, blood etc. Persistence, which might cause the highest degrees of similarity, will be addressed by looking at the compounds either by expert judgment, data on persistence and automated mechanisms of the Hasse-technique.

References:

1. Rappaport S.M., Smith M.T. (2010); *Science*. 330(6003) 460-461.
2. Turgut C., Atatanir L., Mazmanci B., Mazmanci M.A., Henkelmann B., Schramm K.W. (2011*); Environ Sci Pollut Res Int.* 19(2) 325-334.
3. Voigt K., Bruggemann R., Scherb H., Cok I., Mazmanci B., Mazmanci M.A., Turgut C., Schramm K.W. (2013). *Int J Environ Health Res.* 23(3) 226-246.
4. Bruggemann R., Scherb H., Schramm K.-W., Cok I., Voigt K.. (2014); *Ecotoxicology and Environmental Safety*. accepted, DOI: 10.1016/j.ecoenv.2014.03.031.
5. Bruggemann R., Patil G.P. (2011) Ranking and Prioritization for Multi-indicator Systems. Springer, Berlin.
6. Rosner D., Markowitz G. (2013); *Environ Res.* 120 126-133.