

A Study to Generate a Weak Order from a Partially Ordered Set, Taken Biomonitoring Measurements



Rainer Bruggemann, Lars Carlsen, Bardia Panahbehagh,
and Stergios Pirtosos

1 Introduction

Several scientific disciplines are involved in the study of chemical processes that occur in water, air, terrestrial and living environments, and the effects of human activity on them. Environmental chemistry, as one of these, is not only the discipline handling substances in difficult targets, such as sludge (Jin et al. 2017), trees (Ferretti et al. 2002), or lichens (Pirtosos et al. 2006), but also having the task to support decisions in environmental systems (Pirtosos and Loppi 2008).

Due to the complexity of environmental systems a series of well-defined indicators is constructed, representing the knowledge about the system and thus supporting decisions for an appropriate management (Buonocore et al. 2018; Grönlund 2019). Hence, the start for a management and a decision based on a ranking of chemicals (our example) is the analysis of multi-indicator systems (MIS). An example of MIS is the output of several lichen biomonitoring studies concerning metal pollution in the atmospheric environment. To be more specific, the indicators are related to

R. Bruggemann (✉)

Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

e-mail: brg_home@web.de

L. Carlsen

Awareness Center, Roskilde, Denmark

e-mail: LC@AwarenessCenter.dk

B. Panahbehagh

Department of Mathematical Sciences and Computer, Kharazmi University, Tehran, Iran

e-mail: panahbehagh@khu.ac.ir

S. Pirtosos

Department of Biology, University of Crete, Heraklion, Greece

e-mail: pirtosos@biology.uoc.gr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021

R. Bruggemann et al. (eds.), *Measuring and Understanding Complex Phenomena*,

https://doi.org/10.1007/978-3-030-59683-5_5

locations, their values are derived following a procedure described by Nimis and Bargagli (1999).

Lichens are perennial, slow-growing organisms, highly dependent on the atmosphere for nutrients. The lack of a waxy cuticle and stomata allows many contaminants, which are deposited on lichens by precipitation, fog and dew, dry sedimentation and gaseous absorption, to be absorbed over the whole lichen thallus surface, indicating levels of these contaminants in the surrounding environment (Loppi et al. 1999). By biomonitoring at specifically selected sites, for example near roads (Fрати et al. 2006) or more pristine areas (Loppi and Pirintsos 2003), information is obtained about the transport and origins of pollution.

In Pirintsos et al. (2014) 11 metals/metalloids are investigated in 20 sites of an urban and industrial area based on the lichen biomonitoring data set of Demiray et al. (2012), where *Xanthoria parietina* lichen specimen have been used as a biomonitoring organism. The evaluation of the corresponding data matrix is based on the conception that the Hasse diagram technique (see below) can further be expanded and improved in the direction of (i) cumulative risk, (ii) the up-to-date formal presentation and (iii) the interpretation of results in biomonitoring studies of metal atmospheric pollution.

The analysis of biomonitoring results can be crudely characterized by two aspects: (a) attempts to support a decision, based on order relations and (b) attempts to present small scale spatial variations within a geostatistical approach. Here our focus is on the order theoretical aspects.

As the metals and metalloids are measured at m different sites the concentrations found in lichens of each site define, after transformations as recommended by Nimis and Bargagli (1999) an indicator. Hence the MIS contains m indicators, the values describing the pollution due to a single metal or metalloid.

The question arises how to derive a decision when confronted with m indicators. Here we show first a Hasse diagram, which is a visualization of the partial order, induced by the set of indicators (cf. Bruggemann and Patil 2011), then we discuss, as to how far a single ranking (a weak order (see below)) can be obtained without the need of a subjective weighting scheme of the indicators in order to aggregate them by a weighted sum. As an exact solution of the problem how to get a weak order is hardly computationally tractable, we investigate a new calculation method.

2 Material and Methods

2.1 Data Set

Eleven Metals and metalloids, i.e., Hg, Al, As, Cd, Cu, Fe, Mn, Ni, Pb, V and Zn for which their pollution has been monitored are included in the study (Pirintsos et al. 2014). For a management it is of importance which of these metals or metalloids (in the following we call them simply metals, although this is not a chemically correct

Table 1 data of 11 metals and their scores in lichens in 10 sites

metals										
Al	As	Cd	Cu	Fe	Hg	Mn	Ni	Pb	V	Zn
Sites										
st1	st2	st3	st4	st5	st6	st7	st8	st9	st10	
data matrix										
	st1	st2	st3	st4	st5	st6	st7	st8	st9	st10
Al	4	5	7	5	5	4	4	4	3	5
As	2	5	7	6	7	5	5	4	4	4
Cd	5	7	7	7	4	3	4	2	2	3
Cu	5	5	7	7	4	3	5	3	3	4
Fe	7	7	7	7	7	7	7	7	6	7
Hg	4	3	5	3	3	3	4	2	2	2
Mn	7	7	7	7	6	6	7	4	5	6
Ni	7	7	7	7	6	5	7	4	4	7
Pb	7	7	7	7	5	4	6	4	7	4
V	7	7	7	7	7	6	7	7	4	7
Zn	7	7	7	7	7	7	7	6	7	7

term) is highly concentrated. As sites we select only 10 (of 20) because of reasons which become clear in following sections. The concentrations are transformed into a scale of integers from 1 to 7, following the suggestion of Nimis and Bargagli (1999). 1 indicates a high naturality, whereas 7 express a high deviation from the natural state. The data are shown in Table 1, in the Results-section. An entry of the data matrix, $dm(I,j)$ is associated with j th site and the i th metalloid. Details can be found in Pirintsos et al. (2014).

2.2 Basic Concepts of Partial Order

Let X be a finite set of n objects, labeled by $x(i)$ ($i = 1, \dots, n$). Objects could be but not limited to

- chemical compounds (here: metals/metalloids)
- nations, characterized by for example child well-being indicators
- strategies, characterized by performance indicators
- geographical units, characterized for example by pollution, or (within a socio-economic context) by poverty indicators

Here, indeed the elements of the real example are “chemical elements”, namely $n = 11$ metals.

To define an order relation among them, the relation “ \leq ” has to obey the following order axioms:

- reflexivity: the object can be compared with itself
- antisymmetry: if $x \leq y$ and $y \leq x \Rightarrow x = y$
- transitivity: if $x \leq y$ and $y \leq z \Rightarrow x \leq z$

A special realization of order relations is given by Eqs. (1, 2, and 3):

$$x(i) = (x(i, 1), x(i, 2), \dots, x(i, m)) \quad (1)$$

The quantity $x(i, j)$ is the value of the i^{th} object ($i = 1, \dots, n$) (here the i^{th} metal), the j^{th} indicator ($j = 1, \dots, m$) (here the j^{th} site) and m the number of indicators used (here $m = 10$).

Equation 1 describes a mapping $X \rightarrow IR^m$, wherein X is the set of objects (the metals) and IR^m is the set of tuples of real numbers with m components. Note that the tuples are also denoted as data profiles.

According to $m = 10$ sites, we will have a system of 10 indicators.

$$x(i1) \leq x(i2) : \iff (x(i1, 1), \dots, x(i1, m)) \leq (x(i2, 1), \dots, x(i2, m)) \quad (2)$$

Equation 2 needs clarification, as it is not yet clear under which conditions one tuple (that of $x(i1)$) is to be considered less or equal to that of $x(i2)$. The way how Eq. 2 can be given a meaning, opens the door to many variants. By Eq. 3

$$(x(i1, 1), \dots, x(i1, m)) \leq (x(i2, 1), \dots, x(i2, m)) : \iff x(i1, j) \leq x(i2, j) \\ \text{for all } j = 1, \dots, m \quad (3)$$

a special partial order is defined. Two objects, following Eq. 3 are called “comparable”, otherwise “incomparable”.

The immediate relation to the data and the corresponding indicators has two consequences:

1. Any order relation $x \leq y$ is a direct reflection of the data values of x and y . This is in contrast to many decision support systems, where an order relation cannot easily be traced back to the original data, i.e., to the data matrix.
2. The partial order methodology, based on Eq. 3 is applicable wherever a data matrix is available and where a ranking aim can be defined.

In the literature the method, based on Eq. 3 together with appropriate supporting software, is often denoted Hasse diagram technique (HDT) (Galassi et al. 1996; Grisoni et al. 2015; Halfon and Reggiani 1986; Bruggemann et al. 2001, 2008; Patil and Taillie 2004; Klein and Ivanciuc 2006; Simon et al. 2004, 2006; Helm 2003; Bruggemann and Voigt 2008, 2011, 2012; Carlsen and Bruggemann 2011, 2014a, b; Carlsen 2008a, b, 2013, 2018; Newlin and Patil 2010; Annoni et al. 2014; Sørensen

et al. 1998, 2000; Pavan and Todeschini 2004; Pudenz and Heininger 2006; Quintero et al. 2018; Restrepo and Bruggemann 2008; Restrepo et al. 2008a, b; Voigt et al. 2004a, b) with reference to the German mathematician Helmut Hasse (1967). Sets X , equipped with a partial order (and thus in this paper by the Eqs. 1, 2, and 3) are called partially ordered sets and are conveniently denoted as posets and indicated by (X, \leq) .

Two objects x, y mutually incomparable are denoted as $x \parallel y$. Within a poset (X, \leq) the number of incomparable pairs $x \parallel y$ is called U . When the orientation $x \leq y$ or $x \geq y$ is of minor interest then the mere fact of comparability is denoted by $x \perp y$.

Note that partial order methodology can also be applied, by evaluation of the space of all possible data profiles, when the indicators are discrete (cf. e.g. Fattore and Maggino 2014, as well as Maggino et al. this book).

2.3 *Hasse Diagram*

The construction of a Hasse diagram, starting from a set of partial order relations (as an outcome of Eq. 3) is frequently explained in the literature (see e.g. Bruggemann and Halfon 1997). For the sake of reader's convenience, some words about Hasse diagrams may nevertheless be useful here: The basis is the order relation $x < y$. Usually the object x will be drawn below object y ; both are vertices of a graph and presented by small circles, with the label of the object in the centre. In case $x < y$ a line is connecting x with y , called an edge, if the vertices are in a cover relation, i.e. if there is no object z for which is valid: $x < z < y$. The orientation of the order relation is just obtained from the vertical position. When two objects are not connected by a system of oriented edges the two objects are incomparable.

By this construction a Hasse diagram allows a two-fold interpretation:

1. Upwards: The numerical values of the objects are nondecreasing along a system of edges. This "vertical" oriented analysis allows a ranking of objects or subsets of X , so-called chains.
2. In contrast to (1) there is also a "horizontal" evaluation. This evaluation has its focus on not connected objects. Following the construction principles of a Hasse diagram, the objects of in the same vertical position are mutually incomparable. A set of mutually incomparable objects is called an antichain.

2.4 *Weak Order*

When the general policy of decision is to find not only the optimal option but also alternatives, then ranking is a good starting point, since suboptimal objects can be easily identified if the optimal object is not suitable (e.g., due to political or economic reasons). The task is how to get a ranking, which is at least a weak order,

if ties are accepted. Whereas a complete (i.e., total or linear) order is a set of objects, in which all elements $x, y \in X$ are mutually comparable with $x \neq y$, a weak order does not require the condition $x \neq y$, i.e., it accepts equivalent elements (or in terms of statistics: it accepts ties).

A Hasse diagram allows identifying rankings for subsets of X , without any subjectivity beyond the data matrix. It is clear that the task to get a weak order should be parameter free too. Hence, the typical procedure to aggregate the values of the m indicators into a composite indicator by a numerical procedure, where weights for each indicator and other parameters are required, is to be avoided. An important device, how to get a weak order without the need of finding additional parameters, such as weights for the indicators, out of a partially ordered set is found in the paper of Winkler (1982). The crucial term is the average height, denoted as *Hav*.

2.5 Average Height

Any poset can be represented by a set of linear order, whose elements are called linear extensions (Davey and Priestley 1990; Trotter 1992). A linear extension is a linear order, respecting all order relations within a poset. For example the set $X = \{a, b, c, d\}$ may have the following order relations:

$$a < b, a < c, a < d. \quad (4)$$

Obviously, $b \parallel c$ and $c \parallel d$, i.e. $U = 2$. Then the set of linear extensions is:

$$\{(a, b, c, d), (a, c, b, d), (a, c, d, b)\}.$$

Within the above set $\{(a, b, c, d), (a, c, b, d), (a, c, d, b)\}$ a linear extension is for example (a, b, c, d) , others are (a, c, b, d) and (a, c, d, b) . Each single linear extension indicates a complete ordered set, for example (a, b, c, d) denotes: $a < b, a < c, a < d, b < c, b < d, c < d$. All order relations of Eq. 4 are reproduced. The fact that the poset in Eq. 4 includes some incomparabilities leads to the necessity to consider the 3 linear extensions simultaneously. Within each linear extension any object x has a height that is the number of objects $\leq x$. For example, in the linear extension (a, b, c, d) object a has the height 1, b the height 2, whereas in the linear extension (a, c, d, b) object b has the height 4.

The idea of Winkler (1982) is to calculate the average of all heights of all objects, denoted as $Hav(x)$. Let $L(k)$ be the k th linear extension and $h(L(k), x)$ the height of x in $L(k)$, then, after Winkler (1982)

$$Hav(x) := \left(\sum h(L(k), x) \right) / LT \quad (k = 1, \dots, LT) \tag{5}$$

where LT is the number of linear extensions derived from a specific poset.

Equation 5 could be a good starting point, when the set of linear extensions is small. Taking into mind that the number of linear extensions for an object set with n objects can be up to $n!$ the problem to generate linear extensions and store them into a memory is computationally hard (see e.g. Atkinson and Chang 1986).

The above mentioned difficulty leads to several variants:

- There is still an exact method available. It is based on the fact that the storage of some sets derived from the poset needs less memory than the storage of the linear extensions. From a methodological, mathematical point of view this method transforms the original poset into a lattice and the quantities of interest can be directly derived from this lattice (De Loof et al. 2006). However, the lattice-method is only working, when $U*n$ (U : number of incomparabilities in a set of n objects) is not too large, for details see Bruggemann and Carlsen (2011).
- Some approximations seem to have found more applications, for instance the method of Bublely and Dyer (1999), which suggests a “good” sampling of linear extensions.
- Another one has a graph – theoretical background and considers the local environment around each object within a poset. There are two variants: (1) the LPOM0 (local partial order model 0) Bruggemann et al. 2004) and (2) an extended model (LPOMext) (Bruggemann and Carlsen 2011). Although the extended variant is thought of as delivering better results than LPOM0, it turned out (Rocco and Tarantola 2014) that the more simple method (LPOM0) may be in some cases a better approximation than the extended one.

2.6 Idea for an Alternative for the Hav-Calculation

Often partial order can be considered as being composed from simpler posets, here for example, the concept of linear sum is of specific interest. It is defined as follows: Let X_1, X_2 be disjoint subsets of X with

$$X = X_1 \oplus X_2 \tag{6a}$$

$$x \in X_1, y \in X_2 \text{ implies } x > y \text{ for every } x, y \tag{6b}$$

Equation (6b) can be formulated as follows: If two sets can be found where for an element of the first set, x , and for any element of the second set, y , is valid: $x > y$; the relations among the first, and the second set, resp., are not of interest.

We may speak of “ X_1 is fully dominating X_2 ”. Equation 6b does not imply that within X_1 or X_2 the elements are mutually comparable.

Let $Hav(x, X)$ denote the average height of x , considering the set X and the settings of Eqs. 6a and 6b. Then:

$$Hav(x, X) = |X_2| + Hav(x, X_1) \quad (7)$$

where $|\dots|$ denotes the cardinality of the set. Eq. 7 is a simple conclusion found from Eq. 6b:

$$H(L(k), x \text{ in } X) = |X_2| + H(L(k), x \text{ in } X_1).$$

Thus, a calculation method can be thought of, which can be formulated as follows:

$$x \in X_1 : Hav(x, X) = Hav(x, X_1) + |X_2| \quad (8)$$

$$y \in X_2 : Hav(y, X) = Hav(y, X_2), \quad (9)$$

supposed that Eq. 6b is exactly fulfilled.

The concept of $X_1, X_2 \subset X$ with $X_1 \cap X_2 = \emptyset$ was already studied by (Restrepo and Bruggemann 2008) and lead to two quantities, the dominance of X_1 over X_2 and the separability of X_1 and X_2 (Eqs. 10 and 11).

$$\underline{Dom}(X_1, X_2) := |\{(x, y) \text{ with } x \in X_1, y \in X_2 \text{ and } x > y\}| / (|X_1| * |X_2|) \quad (10)$$

$$\underline{Sep}(X_1, X_2) := |\{(x, y) \text{ with } x \in X_1, y \in X_2 \text{ and } x \parallel y\}| / (|X_1| * |X_2|) \quad (11)$$

By a set of subsets the quantities, defined in Eqs. 10 and 11 can be conveniently denoted as matrices, dominance (Dom) and separability (Sep) matrices.

Equation (6b) demands that $Dom(X_1, X_2) = 1$ and $Sep(X_1, X_2) = 0$.

If $Dom(X_1, X_2) = 0$, then X_1, X_2 are completely separated subsets, meaning that then $x \in X_1, y \in X_2$ implies $x \parallel y$. In that case it is easily seen that we find:

$$\text{Hav}(x, X_1) < \text{Hav}(x, X) < |X_2| + \text{Hav}(x, X_1) \quad (12a)$$

$$\text{Hav}(y, X_2) < \text{Hav}(y, X) < |X_1| + \text{Hav}(y, X_2) \quad (12b)$$

because as an extremal case the subposet based on X_1 can once be completely below the subposet $(X_2, <)$ or completely above $(X_2, <)$. Hence: When $\text{Dom}(X_1, X_2) < 0.5$ then the role of the separability matrix is overwhelming (because the sum of Dom- and Sep-matrices is bounded, due to the finite number of comparabilities and incomparabilities and the Eqs. 8 and 9 fail. Therefore it is needed that the poset, to be considered, has more comparabilities than incomparabilities. This is the reason, why instead of 20 sites (the real example) only the 10 first sites were selected.

Summarizing: From a methodological point of view, we want to check, as to how far a deviation of $\text{Dom}(X_1, X_2)$ from 1 can lead to acceptable results.

3 Results

3.1 Randomly Generated Datasets

In order to test as to how far deviations of $\text{Dom}(X_1, X_2)$ from 1 lead to errors in the estimation of Hav, 22 smaller datasets (each of 10 objects) were randomly generated. For each object x of these artificial data sets the Hav-value based on the scheme given in Eq. 12 was calculated, $\text{HavDom}(x)$ and the exact value, Havexact , based on the lattice theoretical method presented by De Loof et al. (2006, 2011, 2012). The deviation was calculated:

$$\text{Eps}(x) := | \text{Havexact}(x) - \text{HavDom}(x) | \quad (13)$$

For each dataset a final value epsav was determined:

$$\text{epsav} := \sum \text{Eps}(x) / n \quad \text{with } n = |X| \quad (14)$$

the quantity epsav being the average error related to any single object. In Fig. 1 the scatterplot, together with the regression equation is shown.

Figure 1 confirms that the deviations epsav will be rather large, when $\text{Dom}(X_1, X_2)$ becomes small values. It is clear that the way, how the partitioning of X into two subsets X_1 and X_2 is selected, plays an important role. However, aiming at an efficient method for the calculation of Hav, the principles were:

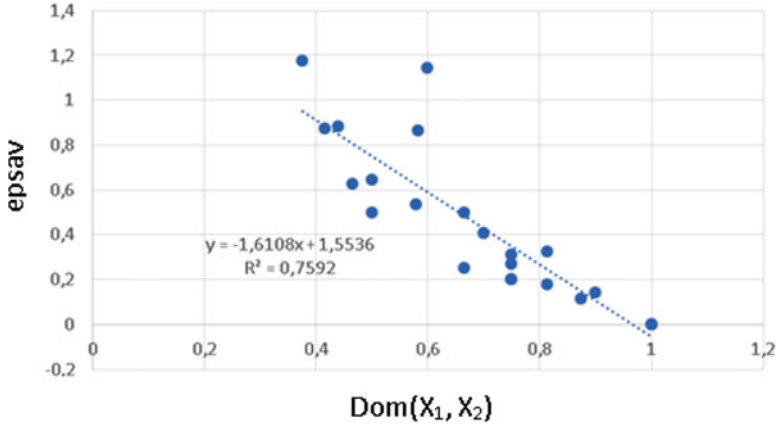


Fig. 1 Scatterplot of $epsav$ vs $Dom(X_1, X_2)$ and the regression equation with $R^2 \approx 0.76$

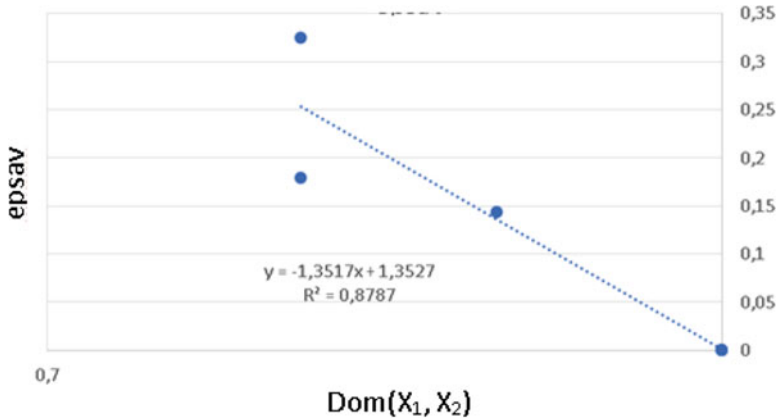


Fig. 2 $epsav$ vs $Dom(X_1, X_2)$, with $Dom(X_1, X_2) \geq 0.8$

1. To select the X_1, X_2 in that manner that they have approximately the same number of elements
2. To find a selection that maximizes $Dom(X_1, X_2)$.

The principle (1) was a priori considered as more important than the principle (2). From Fig. 1 it becomes clear that obviously in the specific considered randomly generated case (for details, see below) the deviations $epsav$ require $Dom(X_1, X_2) \geq 0.8$.

When the regression is restricted to those pairs of values $(Dom(X_1, X_2), epsav)$, where $Dom(X_1, X_2) \geq 0.8$, then the result is (more or less trivially) better, see Fig. 2.

The regression equation based on 6 pairs (the pair (1,0) is realized three times) has the striking structure:

$$epsav = a * (1 - Dom(X_1, X_2)) \text{ for } 0.8 < Dom(X_1, X_2) \leq 1 \quad (15)$$

with the coefficient of determination, $R^2 = 0.88$ and the coefficient a around 1.35. This statistical result indicates that the relevant quantity is the deviation Δ :

$$\Delta := 1 - Dom(X_1, X_2) \quad (16)$$

Hence, Eq. 15 expresses proportionality between the error $epsav$ and Δ . The crucial value 0.8 for separating relevant $Dom(X_1, X_2)$ -values from irrelevant ones, may vary from case to case and is open for future research. Furthermore, Fig. 2 shows that the average error related to single objects is less than 0.25 and the deviations from the regression line will be larger the smaller the value $Dom(X_1, X_2)$ is.

3.2 Application to Real Data Set

The estimation method needs the following steps:

1. Defining X_1, X_2 and the Hasse diagram for the full set X
2. Calculation of $Dom(X_1, X_2)$
3. Providing the data for X_1 and X_2
4. Application of the lattice theoretical method: (a) for X , (b) for X_1 , (c) for X_2
5. Performing the calculations due to Eqs. 8 and 9
6. Inspecting $epsav$ to check the quality of the results

Up to now there is no program performing all 6 steps. However, for steps (1)–(3) the program package PyHasse (see for details Bruggemann et al. 2014) was extended by the new module DomRkav. Its graphical user interface is shown in Fig. 3. The data is found in Table 1.

The corresponding Hasse diagram is shown in Fig. 4.

Some remarks concerning Fig. 4 may be useful here:

- Hg and As are minimal elements, they cause the least deviation from a natural state
- Fe and Zn are maximal elements, they are most problematic because the deviation of the natural state is very high.
- Incomparabilities, such as for Zn and Fe show that the loading of the lichens in general is high, however with some geographical differentiation.

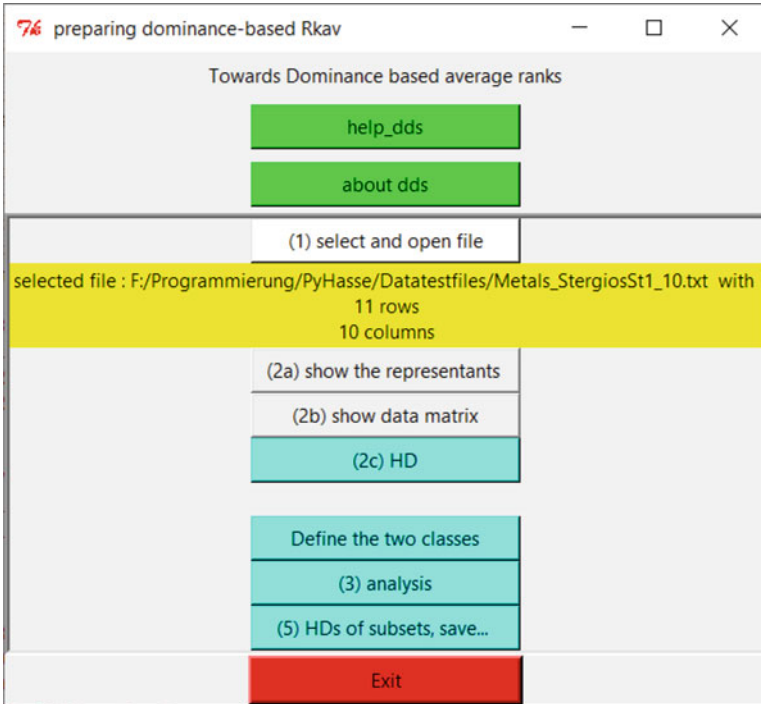
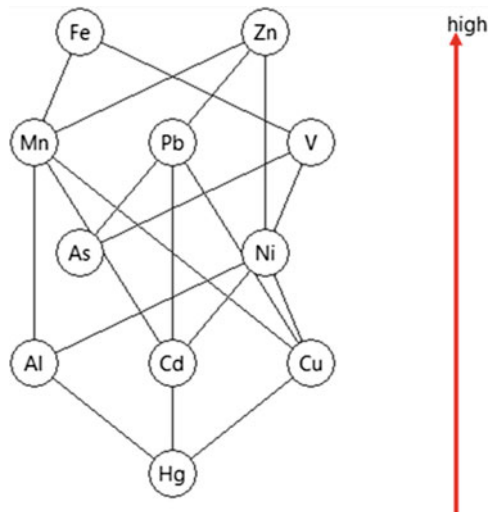


Fig. 3 Graphical user interface for the new PyHasse module DomRkav

Fig. 4 Hasse diagram of the metals, according to their deviations from their natural state in the studied stations



In order to perform the calculation scheme based on Eqs. 8 and 9 the first step is to select X_1 and X_2 , i.e. the partitioning of set X .

Step 1:

The sets X_1 and X_2 are:

set X_1 : Fe, Zn, Mn, Pb, V, As

set X_2 : Ni, Al, Cd, Cu, Hg

Step2:

The Dom-matrix is:

	X_1	X_2
X_1 :	0.361	0.733
X_2 :	0.0	0.48

Remark 1:

Although $Dom(X_1, X_2) = 0.733$ is less 0.8 the next calculation steps are documented, just for a demonstration.

Remark 2:

The partitioning selected above is not the only possible one. For example, the metalloid As is a minimal element. Why not assign As to X_2 ? Let $X_2' = X_2 \cup \{As\}$ and $X_1' = X_1 - \{As\}$. Indeed the value of $Dom(X_1', X_2') = 0.833$ is better than that of $Dom(X_1, X_2)$ and correspondingly $epsav = 0.395$. The disadvantage is that (X_2', \leq) leads due to its symmetry to a very high degree of degeneracy: $Fe \cong Zn$, $Pb \cong V$ and $As \cong Al \cong Cd \cong Cu$. Therefore we continue with the partitioning of X into X_1 and X_2 as given above.

Step 3: Calculation of the averaged ranks by the lattice-theoretical method (De Loof et al., 2006) due to X_1 and X_2 of step 1.

Figure 5 shows the Hasse diagrams of the two subsets.

Fig. 5 The two Hasse diagrams due to X_1 and X_2

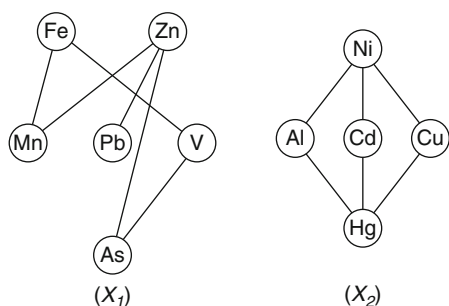


Table 2 Summarizing the results of step 4 and step 5. Whether $x \in X_1$ or $x \in X_2$ is indicated by a membership function. If $x \in X_1$ then the value in the corresponding column = 1, otherwise 0. Similarly for $x \in X_2$

Object x	Hav_{exact}	X_1	X_2	$Hav(\dots, X_1)$	$Hav(\dots, X_2)$	$HavDom$	$Eps(x)$
Al	3,581	0	1		3	3	0,581
As	4,246	1	0	1,597		6,597	2,351
Cd	3,36	0	1		3	3	0,36
Cu	3,36	0	1		3	3	0,36
Fe	10,487	1	0	5,455		10,455	0,032
Hg	1,135	0	1		1	1	0,135
Mn	7,175	1	0	2,364		7,364	0,189
Ni	6,559	0	1		5	5	1,559
Pb	7,169	1	0	2,636		7,636	0,467
V	8,693	1	0	3,576		8,576	0,117
Zn	10,236	1	0	5,273		10,273	0,037

$$\sum Eps(x) = 6.188$$

$$epsav = 0.563$$

$$Dom(X_1, X_2) = 0.733$$

Step 4: Application of Eqs. 8 and 9.

Step 5: Check for the accuracy of the results.

The remaining steps 4 and 5 are summarized in the following Table 2. The column below X_1 and X_2 is the membership function, indicating whether or not the metal belongs to X_1 or to X_2 .

The value of $epsav = 0.563$ deviates from the value obtained from Eq. 15; ($epsav_{eq.15} = 0.36$). However the value of $Dom(X_1, X_2)$ is not within the range of applicability of Eq. 15. As to be expected, the measure of deviation, $epsav$, indicates a bad approximation. Due to pretty large deviations (in terms of $epsav$ the final weak order shows two inversions:

Exact: $Hg < Cd \cong Cu < Al < As < Ni < Pb < Mn < V < Zn < Fe$

Approx.: $Hg < Cd \cong Cu \cong Al < Ni < As < Mn < Pb < V < Zn < Fe$

As it is often the case, different methods coincide, when extremal ranking positions are to be detected. This empirical finding is found here as well, i.e., Hg Cd, Cu, as well as V, Zn and Fe coincide in their positions at the beginning or the end of the ranking sequence. The other positions in a ranking sequence are usually determined by many factors. Therefore, here different methods will lead to different ranking positions. Here, indeed, some other metals change their position (As, Ni) and (Mn, Pb), when the exact, lattice theoretical method is compared with the approximation, suggested here. The reasons for the inversion Mn, Pb is that Mn “sees” four vertices order theoretically less than Mn, whereas Pb only “sees” three vertices. In the approximation however, both are minimal elements, so that for both metals the Eq. 8 gives the same summand $|X_2|$, being 4. A similar argument holds for the pair (As, Ni).

4 Discussion

4.1 *Lichen Biomonitoring/Bioaccumulation Matrices as Multi-indicator Systems*

Undoubtedly, any data pre-processing, as done here, are of high importance not only in chemical risk assessment and management but also broader, in the decision-making process of environmental policy. However, here is not the place to discuss in depth the pre-processing, defined by Nimis and Bargagli (1999), nevertheless, we think that here some words may be helpful:

In biomonitoring techniques of air quality with native lichens, an approach to the interpretation of data of native lichens is the so-called “naturalness/alteration scales” based on thresholds identifying classes of increasing element concentrations, and obtained by the meta-analysis of a large set of bioaccumulation data. The method by Nimis and Bargagli (1999) defines seven classes of element concentrations. These classes are built up on hundreds of data points collected in Italy between the 1980s and the 1990s. The seven class scale refer to (1) very high naturalness, (2) high naturalness, (3) middle naturalness, (4) low naturalness/alteration, (5) middle alteration, (6) high alteration and (7) very high alteration based on the percentile distributions of element concentrations in lichens (Nimis et al. 2000).

Recently a paper was published, where the data pre-processing of data (is examined under the methodological background of partial order theory, see Fattore et al. (2019).

4.2 *Applicability of the Proposed Method*

The quantity $epsav$, Eq. 14 is an average value and is – as mentioned already above – related to a single object. The domain of validity for Eq. 15 is given by $0.8 \leq Dom(X_1, X_2) \leq 1.0$.

If $Dom(X_1, X_2) \rightarrow 0.8$ the deviations Eps (Eq. 13) become quickly large as Figs. 1 and 2 (randomly generated data) show. Consequently in the following paragraph we investigate reasons for large deviations of Eps .

4.3 *Reasons for Large Deviations*

First of all, a dissection of a poset (X, \leq) into two subposets (X_1, \leq) and (X_2, \leq) leads to more symmetry in the resulting graphs of the subposets (as already mentioned above). Hence, the degeneracy of Hav -values is increased. Even if the enhanced degree of ties is accepted, there can be large deviations, which result from structures like the one shown in Fig. 6.

Fig. 6 X_1 dominates fully X_2' but not X_2 . A typical situation causing deviations

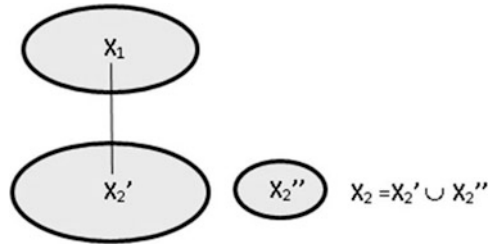
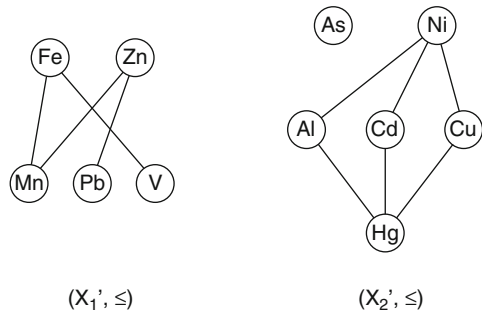


Fig. 7 A variant for partitioning of set X



Considering Fig. 6 a situation, similar to that, causing eq. 12, arises. For $x \in X_1$, $Hav(x, X) = Hav(x, X_1) + |X_2|$ is an overestimation, because by constructing the linear extensions, the elements of X_2'' can also be located above the elements of X_1 , whereas by $Hav(x, X) = Hav(x, X_1) + |X_2'|$ an underestimation follows. Based on remark 2 (see above) the two Hasse diagrams are shown, when $X_1' = X_1 - \{As\}$ and $X_2' = X_2 \cup \{As\}$ (Fig. 7)

The element Fe “sees” the same number of lower neighbours as Zn. Similarly, Pb and V have one upper neighbour. Therefore the exact method delivers $Hav_{exact}(Fe, X_1') = Hav_{exact}(Zn, X_1')$ as well as $Hav_{exact}(Pb, X_1') = Hav_{exact}(V, X_1')$. In Fig. 7, the subposet (X_2', \leq) has also symmetries, leading to: $Al \cong Cd \cong Cu$ with respect to Hav_{exact} .

4.4 Conclusive Consideration

By applying the dominance matrix and based on this, the calculation scheme seems to be attractive for an estimation method of Hav. However, the requirement of very high values of $Dom(X_1, X_2)$ seems to be too restrictive to justify to propose this method as a general approximation method. Thus, up to our actual knowledge this new procedure will not be practically feasible in comparison with exact results. When, however, a first check is wanted, for example to start from this a refinement procedure, then the scheme based on Eqs. 8 and 9 may be useful. When this line of research is to be followed, then

- A catalogue could be aimed, where structures are gathered, which typically lead to strong deviations
- As a candidate for a better approximation the method by Bublely and Dyer (1999), may be selected and modified in that manner that the weak order as a result of Eqs. 8 and 9 is a starting linear extension.

Summarizing, we hope that the present study has revealed some mathematical ideas which may be of interest and attract new research by scholars of the mathematical chemistry scene.

References

- Annoni, P., Bruggemann, R., & Carlsen, L. (2014). A multidimensional view on poverty in the European Union by partial order theory. *Journal of Applied Statistics*, 42, 535–554.
- Atkinson, M. D., & Chang, H. W. (1986). Extensions of partial orders of bounded width. *Congressus Numerantium*, 52, 21–35.
- Bruggemann, R., & Carlsen, L. (2011). An improved estimation of averaged ranks of partially orders. *MATCH Communications in Mathematical and in Computer Chemistry*, 65, 383–414.
- Bruggemann, R., & Halfon, E. (1997). Comparative analysis of nearshore contaminated sites in Lake Ontario: ranking for environmental hazard. *Journal of Environmental Science and Health*, A32(1), 277–292.
- Bruggemann, R., & Patil, G. P. (2011). *Ranking and prioritization for multi-indicator systems – Introduction to partial order applications*. New York: Springer.
- Bruggemann, R., & Voigt, K. (2008). Basic principles of hasse diagram technique in chemistry. *Combinatorial Chemistry & High Throughput Screening*, 11, 756–769.
- Bruggemann, R., & Voigt, K. (2011). A new tool to analyze partially ordered sets – Application: Ranking of polychlorinated biphenyls and alkanes/alkenes in River main, Germany. *MATCH Communications in Mathematical and in Computer Chemistry*, 66, 231–251.
- Bruggemann, R., & Voigt, K. (2012). Antichains in partial order, example: Pollution in a German region by lead, cadmium, zinc and sulfur in the herb layer. *MATCH Communications in Mathematical and in Computer Chemistry*, 67, 731–744.
- Bruggemann, 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. *The Journal for Chemical Information and Computer scientists*, 41, 918–925.
- Bruggemann, R., Sørensen, P. B., Lerche, D., & Carlsen, L. (2004). Estimation of averaged ranks by a local partial order model. *The Journal for Chemical Information and Computer scientists*, 44, 618–625.
- Bruggemann, R., Voigt, K., Restrepo, G., & Simon, U. (2008). The concept of stability fields and hot spots in ranking of environmental chemicals. *Environmental Modelling & Software*, 23, 1000–1012.
- Bruggemann, R., Carlsen, L., Voigt, K., & Wieland, R. (2014). PyHasse software for partial order analysis: Scientific background and description of selected modules. In R. Bruggemann, L. Carlsen, & J. Wittmann (Eds.), *Multi-indicator systems and modelling in partial order* (pp. 389–423). New York: Springer.
- Bublely, R., & Dyer, M. (1999). Faster random generation of linear extensions. *Discrete Mathematics*, 201, 81–88.
- Buonocore, E., Mellino, S., De Angelis, G., Liu, G., & Ulgiati, S. (2018). Life cycle assessment indicators of urban wastewater and sewage sludge treatment. *Ecological Indicators*, 94, 13–23.
- Carlsen, L. (2008a). Hierarchical partial order ranking. *Environmental Pollution*, 155, 247–253.

- Carlsen, L. (2008b). Partial ordering and prioritising polluted sites. In M. Pavan & R. Todeschini (Eds.), *Data handling in science and technology* (Vol. 27, pp. 97–109). Elsevier B.V.
- Carlsen, L. (2013). Assessing chemicals using partial order ranking methodology. *Advances in Combinational Chemistry & High Throughput Screening*, 1, 3–35.
- Carlsen, L. (2018). Happiness as a sustainability factor. The world happiness index: A posetic – based data analysis. *Sustainability Science*, 13, 549–571.
- Carlsen, L., & Bruggemann, R. (2011). Risk assessment of chemicals in the river Main (Germany): Application of selected partial order ranking tools. *Statistica & Applicazioni*, Special Issue, 2011, 125–140.
- Carlsen, L., & Bruggemann, R. (2014a). Partial order methodology: A valuable tool in chemometrics. *Journal of Chemometrics*, 28, 226–234.
- Carlsen, L., & Bruggemann, R. (2014b). An analysis of the “failed states index” by partial order methodology. *Journal of Social Structure*, 14, 1–31.
- Davey, B. A., & Priestley, H. A. (1990). *Introduction to lattices and order*. Cambridge: Cambridge University Press.
- De Loof, K., De Meyer, H., & De Baets, B. (2006). Exploiting the lattice of ideals representation of a poset. *Fundamenta Informaticae*, 71, 309–321.
- De Loof, K., De Baets, B., & De Meyer, H. (2011). Approximation of average ranks in posets. *MATCH Communications in Mathematical and in Computer Chemistry*, 66, 219–229.
- De Loof, K., Rademaker, M., Bruggemann, R., De Meyer, H., Restrepo, G., & De Baets, B. (2012). Order theoretical tools to support risk assessment of chemicals. *MATCH Communications in Mathematical and in Computer Chemistry*, 67, 213–230.
- Demiray, A. D., Yolcubal, I., Akyol, N. H., & Cobanoglu, G. (2012). Biomonitoring of airborne metals using the lichen *Xanthoria parietina* in Kocaeli Province, Turkey. *Ecological Indicators*, 18, 632–643.
- Fattore, M., & Maggino, F. (2014). Partial orders in socio-economics. A practical challenge for poset theorists or a cultural challenge for social scientists? In R. Bruggemann, L. Carlsen, & J. Wittmann (Eds.), *Multi-indicator systems and modelling in partial order* (pp. 197–214). New York: Springer.
- Fattore, M., Arcagni, A., & Maggino, F. (2019). *Optimal scoring of partially ordered data, with an application to the ranking of smart cities*. Smart Statistics for Smart Applications Book of Short Papers SIS 2019 – ISBN 9788891915108.
- Ferretti, M., Innes, J. L., Jalkanen, R., Saurer, M., Schäffer, J., Spiecker, H., & von Wilpert, K. (2002). Air pollution and environmental chemistry—what role for tree-ring studies? *Dendrochronologia*, 20, 159–174.
- Fрати, L., Caprasecca, E., Santoni, S., Gaggi, C., Guttova, A., Gaudino, S., Pati, A., Rosamilia, S., Pirintzos, S. A., & Loppi, S. (2006). Effects of NO₂ and NH₃ from road traffic on epiphytic lichens. *Environmental Pollution*, 142(1), 58–64.
- Galassi, S., Provini, A., & Halfon, E. (1996). Risk assessment for pesticides and their metabolites in water. *International Journal of Environmental Analytical Chemistry*, 65, 331–344.
- Grisoni, F., Consonni, V., Nembri, S., & Todeschini, R. (2015). How to weight Hasse matrices and reduce incomparabilities. *Chemometrics and Intelligent Laboratory Systems*, 147, 95–104.
- Grönlund, S. E. (2019). Indicators and methods to assess sustainability of wastewater sludge management in the perspective of two systems ecology models. *Ecological Indicators*, 100, 45–54.
- Halfon, E., & Reggiani, M. G. (1986). On ranking chemicals for environmental hazard. *Environmental Science & Technology*, 20, 1173–1179.
- Hasse, H. (1967). *Vorlesungen über Klassenkörpertheorie*. Marburg: Physica-Verlag.
- Helm, D. (2003). Bewertung von Monitoringdaten der Umweltprobenbank des Bundes mit der Hasse-Diagramm-Technik. *UWSF – Z Umweltchem Ökotox*, 15, 85–94.
- Jin, C., Nan, Z., Wang, H., & Jin, P. (2017). Plant growth and heavy metal bioavailability changes in a loess subsoil amended with municipal sludge compost. *Journal of Soils and Sediments*, 17, 2797–2809.

- Klein, D. J., & Ivanciuc, T. (2006). Directed reaction graphs as posets. In R. Bruggemann & L. Carlsen (Eds.), *Partial order in environmental sciences and chemistry* (pp. 35–57). Berlin: Springer.
- Loppi, S., & Pirintsos, S. A. (2003). Epiphytic lichens as sentinels for heavy metal pollution at forest ecosystems (central Italy). *Environmental Pollution*, *121*, 327–332.
- Loppi, S., Pirintsos, S. A., & De Dominicis, V. (1999). Soil contribution to the elemental composition of epiphytic lichens (Tuscany, central Italy). *Environmental Monitoring and Assessment*, *58*, 121–131.
- Newlin, J., & Patil, G. P. (2010). Application of partial order to stream channel assessment at bridge infrastructure for mitigation management. *Environmental and Ecological Statistics*, *17*, 437–454.
- Nimis, P. L., & Bargagli, R. (1999). *Linee-guidaper l'utilizzo come bioaccumulatori di metalli in traccia*. In: Proceedings of Workshop Biomonitoraggio della qualita dell' ariasul territorio nazionale, Roma 26–27, Giugno 1998, ANPA-Serie Atti, pp. 279–287
- Nimis, P. L., Lazzarin, G., Lazzarin, A., & Skert, N. (2000). Biomonitoring of trace elements with lichens in Veneto (NE Italy). *Science of the Total Environment*, *255*, 97–111.
- Patil, G. P., & Taillie, C. (2004). Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization. *Environmental and Ecological Statistics*, *11*, 199–228.
- Pavan, M., & Todeschini, R. (2004). New indices for analysing partial ranking diagrams. *Analytica Chimica Acta*, *515*, 167–181.
- Pirintsos, S. A., & Loppi, S. (2008). Biomonitoring atmospheric pollution: The challenge of times in environmental policy on air quality. *Environmental Pollution*, *151*, 269–271.
- Pirintsos, S. A., Matsi, T., Vokou, D., Gaggi, C., & Loppi, S. (2006). Vertical distribution patterns of trace elements in an urban environment as reflected by their accumulation in lichen transplants. *Journal of Atmospheric Chemistry*, *54*, 121–131.
- Pirintsos, S., Bariotakis, M., Kalogrias, V., Katsogianni, S., & Bruggemann, R. (2014). Hasse diagram technique can further improve the interpretation of results in multielemental large-scale biomonitoring studies of atmospheric metal pollution. In R. Bruggemann, L. Carlsen, & J. Wittmann (Eds.), *Multi-indicator systems and modelling in partial order* (pp. 237–251). New York: Springer.
- Pudenz, S., & Heininger, P. (2006). Comparative evaluation and analysis of water sediment data. In R. Bruggemann & L. Carlsen (Eds.), *Partial order in environmental sciences and chemistry* (pp. 111–151). Berlin: Springer.
- Quintero, N. Y., Bruggemann, R., & Restrepo, G. (2018). Mapping posets into low dimensional spaces: The case of uranium trappers. *MATCH Communications in Mathematical and in Computer Chemistry*, *80*, 793–820.
- Restrepo, G., & Bruggemann, R. (2008). Dominance and separability in posets, their application to isoelectronic species with equal total charge. *Journal of Mathematical Chemistry*, *44*, 577–602.
- Restrepo, G., Bruggemann, R., Weckert, M., Gerstmann, S., & Frank, H. (2008a). Ranking patterns, an application to refrigerants. *MATCH Communications in Mathematical and in Computer Chemistry*, *59*, 555–584.
- Restrepo, G., Weckert, M., Bruggemann, R., Gerstmann, S., & Frank, H. (2008b). Ranking of refrigerants. *Environmental Science & Technology*, *42*, 2925–2930.
- Rocco, C. M., & Tarantola, S. (2014). Evaluating ranking robustness in multi-indicator uncertain matrices: An application based on simulation and global sensitivity analysis. In R. Bruggemann, L. Carlsen, & J. Wittmann (Eds.), *Multi-indicator systems and modelling in partial order* (pp. 275–292). New York: Springer.
- Simon, U., Bruggemann, R., & Pudenz, S. (2004). Aspects of decision support in water management – Example Berlin and Potsdam (Germany) I – Spatially differentiated evaluation. *Water Research*, *38*, 1809–1816.

- Simon, U., Bruggemann, R., Behrendt, H., Shulenberg, E., & Pudenz, S. (2006). METEOR: A step-by-step procedure to explore effects of indicator aggregation in multi criteria decision aiding – Application to water management in Berlin, Germany. *Acta Hydrochimica et Hydrobiologica*, 34, 126–136.
- Sørensen, P. B., Mogensen, B. B., Gyldenkaerne, S., & Rasmussen, A. G. (1998). Pesticide leaching assessment method for ranking both single substances and scenarios of multiple substance use. *Chemosphere*, 36, 2251–2276.
- Sørensen, P. B., Mogensen, B. B., Carlsen, L., & Thomsen, M. (2000). The influence on partial order ranking from input parameter uncertainty – Definition of a robustness parameter. *Chemosphere*, 41, 595–600.
- Trotter, W. T. (1992). *Combinatorics and partially ordered sets, dimension theory*. Baltimore: The Johns Hopkins University Press.
- Voigt, K., Welzl, G., & Bruggemann, R. (2004a). Data analysis of environmental air pollutant monitoring systems in Europe. *Environmetrics*, 15, 577–596.
- Voigt, K., Bruggemann, R., & Pudenz, S. (2004b). Chemical databases evaluated by order theoretical tools. *Analytical and Bioanalytical Chemistry*, 380, 467–474.
- Winkler, P. (1982). Average height in a partially ordered set. *Discrete Mathematics*, 39, 337–341.