MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **71** (2014) 117-142

ISSN 0340 - 6253

Average Heights in Partially Ordered Sets

Rainer Bruggemann[1*], Paola Annoni[2]

[1*] *Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany*

[2] *European Commission, DG JRC, Unit of Econometrics and Applied Statistics, Ispra, Italy*

(Received April 10, 2013)

Abstract

Ranking of chemicals is an important step in risk assessments. The hazard of chemicals exerted on human or on environments is not directly measurable, instead a set of indicators serve as proxy for this purpose. This so-called multi-indicator system is of high scientific value because it contains specific knowledge about chemicals. Often a ranking is obtained by aggregating this detailed information to get a scalar, the ranking index. Such an aggregation may be considered as subjective, at least up to certain extent. An alternative to aggregation is possible by applying simple elements of partial order theory. Instead of unique, linear orders, partial order theory can provide the so-called average heights from the concept of linear extensions. The calculation of average heights is of high interest because, once estimated, a weak order can be derived. Once the weak order is found, (tied) ranks can be determined. This is why average heights are often called average ranks. The calculation of the needed average heights is most often computationally intractable and approximations are needed. The quality of the approximations cannot be checked because of the computational restrictions in calculating exactly all needed average heights. A possible way out of this conundrum is the definition of models of partially ordered sets – model-posets - which are simple enough to derive analytical formulas for the average heights and, at the same time, provide enough structural diversity to allow for generalizations. Once model-posets are set-up, the quality of approximations can be checked and the role of different parts of the graph, representing the partially ordered set, can be studied by sensitivity analysis. We start the analysis with an example taken from environmental chemistry, where persistent organic pollutants are ranked according to three indicators, namely persistence, bioaccumulation and toxicities.

It is found that *i*) an approximation model, based on a local partial order concept, provides results in a satisfying coincidence with the exact values and *ii*) that the structure parameters can be ranked with respect to their impact on the average height value.

1. Introduction

One of the first and most important steps of chemicals' evaluation is providing a ranking. The ranking of complex systems such as chemicals usually leads to a multi-indicator system (MIS) [1] which describes all the aspects that are important for evaluation. In order to get a ranking, most often weighted sums of the indicator values are computed leading to composite indicators. Hence composite indicators are the results of a series of subjective choices, transformation, normalization, weighting scheme to name only few, which may undermine the message they convey. An alternative to the numerical combination of indicators is the application of simple elements of partial order theory. Partial order theory does not require subjective decisions, but the tradeoff is that the resulting structure, the directed acyclic graph of order relations, does not lead to a unique ranking. Therefore the general attempts are nowadays oriented toward the use of partial order theory to support conventional multi-criteria decision analysis tools, such as PROMETHEE [2] or the aforementioned composite indicators themselves. Instead of unique, linear orders, partial order theory can provide average heights [3]. Average heights are derived from the analysis of the linear extensions, as will be shortly shown. The calculation of average heights – *hav* - is of high interest because once estimated, a weak order and from them ranks can be derived (therefore in applications *hav* is often called "average rank").

The calculation of average heights is most often computationally impossible. Indeed the calculation of *hav* has the complexity measure of being #P-complete [4]. Approximations are then to be introduced in order to get estimated values [5-8]. Checking the quality of these estimates is difficult as well, because the exact solution is only available up to a certain number of objects. Our proposal is based on the following steps:

1. defining classes of simple partially ordered sets, each class called "model-poset", which are simple enough to analytically derive the exact average heights;
2. introducing a set of parameters, "structure parameters", which describe the model-poset (see below);
3. presenting formulas of the two approximations of average heights most often used, LPOM0 and LPOMext (for details of the Local Partial Order Model (LPOM) see below and [5-7].
4. performing a global sensitivity analysis to check the single and interactive influence of structure parameters on average height values;
5. assessing the quality of the estimates for in principle infinite number of posets, so far as they belong to one of the hitherto developed classes of model-posets.

In the following we refer e.g. to *hav*(LPOM0) to express that *hav* is calculated following the approximation model of LPOM0, similarly for *hav*(LPOMext) and *hav*(exact). We write *hav*(approx) to indicate that *hav* is calculated by approximations, such as LPOM0 or LPOMext. Furthermore we write *hav*(<method>, x) to indicate that one of the three models, LPOM0, LPOMext or exact is used to calculate the average height of object x. The outcomes of the study are:

1. Identifying whether or not it can be established
   hav(LPOMext) = *hav*(exact) and/or *hav*(LPOMext) = *hav*(LPOM0) and/or
   hav(LPOM0) = *hav*(exact)
   for all possible values of structure parameters or a restricted range of them.
2. Empirical approximation quality, measured by *mean*(|*hav*(exact) - *hav*(approx)|).
3. Assessment of which poset structures influence the average height and to what extent.
4. Assessment of the interaction effect, if any, among the structure parameters.

To this aim, we firstly introduce a leading example, then we explain the basics of partial order theory, the ideas concerning structure parameters and apply variance-based sensitivity method for some classes of partially ordered sets for which analytical formulas are derived.

2. Introductory example

In [9] a study of some persistent organic chemicals is described. Their multi-indicator-system comprises three indicators: persistence (P), bioaccumulation (B) and toxicity (T), quantities known as “PBT-characteristics”. The data set is shown in Table 1:

Table 1. Some persistent organic chemicals and their PBT characteristics. Details, see [9].

Chemical	abbr	P	B	T
p,p-DDT	DDT	0.8361	4.226	58.8235
p,p-DDE	DDE	0.6033	3.962	9.8039
p,p-DDD	DDD	0.5758	3.639	10.4167
Methoxychlor	MEC	0.6611	3.019	6.3291
Lindae		0.659	2.399	0.4333
	HCB	0.7518	3.448	11.3636
Chlordane	CHL	3.6792	3.771	12.8205
Heptachlor	HCl	1.9026	3.276	6.4935
Aldrin	ALD	1.3941	3.956	37.037
Dieldrin	DIE	1.4852	3.098	2.8169
Pentachlor phenol		0.612	3.045	1.006
Pentachlor nitrobenzene		0.7441	2.728	2.0492

An analysis with Hassediagram technique (HDT) was carried out [9] and the Hassediagram is shown in Figure 1.

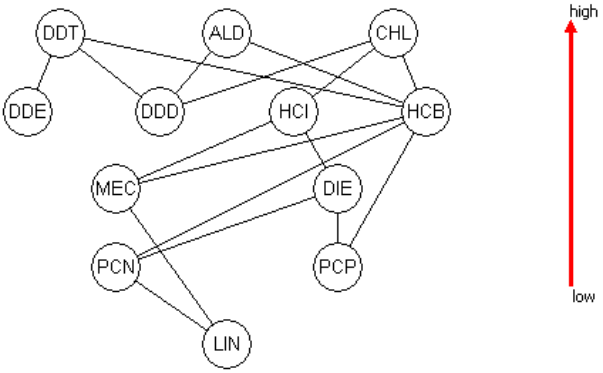

Figure 1. Hassediagram of the 12 chemicals (Table 1) (software: PyHasse, see e.g. [10-12]).

Their main goal in [9] was to decide the quality of different aggregation methods, all based on the technique of calculating a composite indicator (CI) from the three indicators of their multi-indicator system. It was shown with help of the average heights based on three different methods (two variants of the Local Partial Order Model (LPOM0 and LPOMext), and the exact method) that the different composite indicators show severe discrepancies with respect to the partial order derived orders. It is not claimed that partial order methods render more reliable methods, but a discrepancy should motivate a reconciliation between the weighting and the aggregation method itself. A necessary condition for performing this task is the ability to calculate -at least approximately- the average heights.

The issue discussed here and which originally motivated the study is that the very simple LPOM0 method can yield better results than the more sophisticated LPOMext. For example, in [9] it was shown that hav(LPOM0, HCB) = 0.602, hav(LPOMext, HCB) = 0.592, but hav(exact, HCB) = 0.607. Taking a look at Figure 1 one can see that Hexachlorbenzene (HCB) has a striking side-arm, whereas all other chemicals are more or less connected with the others either up or downwards. Therefore the question arose, as to how far the structure of the graph of a partial order has an impact on the quality of approximations.

We see in the analysis of more and more classes of model-posets a way to unfold the intricate relations between the order theoretical environment of an element x and the elements incomparable with x, the influence and interactions of parameters describing the environment on the average heights and the quality of the approximations.

3. Methods and Material

3.1 Basics of Partial Order

Partial order relations can be obtained in many different ways. Even, if the partial order is to be related to a data matrix we can define different partial order relations among objects. For example:

Let X be the finite set of objects, and IB the set of indicators q_i, ($i = 1, \dots, |IB|$) then we define:

$$x,y \in X: \; x \le y : \Leftrightarrow q_i(x) \le q_i(y) \text{ for all } q_i \in IB, \qquad (1)$$

Equation (1) represents a partial order P, which we denote as $P=(X, IB)$ to indicate the intricate relation between the order relation and the set of indicators. The set X together with the partial order structure is called a partially ordered set (poset). Instead of (1) for example the fuzzy concept can serve as a leading idea. Then another definition of partial order comes into play (see for instance [13]).

Some notational remarks are needed:

a) Objects, for which $x \le y$, or $x \ge y$ are called comparable, in sign: $x \perp y$.

b) Objects, for which (1) does not hold, are called incomparable. The fact that object x is incomparable with y, is denoted as $x \parallel y$.

c) Let $Y \subset X$. Then Y is a downset of (X, IB) if $x \in Y$ and $z \le x$ implies $z \in Y$. (2)

d) When $Y = \{x\}$, a principal downset is obtained, this special set is called $O(x, P)$ (order ideal of x in the partial order P). (3)

e) Let $Y \subset X$. Then Y is a upset of (X, IB) if $x \in Y$ and $z \ge x$ implies $z \in Y$. (4)

f) Let $x \in X$, $F(x, P)$: $= \{y \in X: y \ge x \text{ in } P\}$ is an upset (order filter generated by x in the partial order P). (5)

g) The set $U(x, P)$: $U(x, P)$: $= \{y \in X: y \| x \text{ in } P\}$ (6)

h) Isolated elements $Iso(X, P)$: $Iso(X, P) = \{x: x \in X$, there is no $y \in X$, such that $x > y$ or $x < y\}$. (7)

i) Chains: Let $C \subseteq X$, if all $x, y \in C$ obey (1) then C is called a chain. (8)

j) Antichain: Let $AC \subseteq X$, if for all $x, y \in AC$ is valid $x \| y$, then AC is called an antichain

k) Cover relations: Let x, y, z ∈ *X*, when x < y < z , then z "is between" x and y. If x < z without any element ∈ X, which is between x and z then z "covers" x or x is covered by z. A cover relation is denoted by x <: z.

l) An interval *I*(z,y) is the set of elements x ∈ *X* such as z ≤ x ≤ y.

m) Distance between x, y: If x⊥y, then the distance is the minimal number of covering edges between x and y. If x || y, then the distance is set to ∞.

n) Connection: It is convenient to speak of x, y ∈ *X* as being connected, when a path exists between both vertices, where the diagram of cover-relations is considered as an ordinary graph.

o) A linear extension L_i of a poset is a linear order, which preserves all order relations, found on the basis of (1). In a finite set *X* each linear extension has a bottom element. A height $h(x, L_i)$ of an object x with respect to the linear extension L_i is the number of elements in L_i within *I*[bottom element, x] (see l) in L_i.

p) An extension of a poset *P* is a poset *P1*, where all order ≤-relations are preserved. Hence, the linear extension is the special case, where the extension of a poset is a linear order.

q) A weak order is a binary relation which is reflexive, complete, and transitive. Because antisymmetry in not required, a sequence of objects may have ties, such as the weak ordered set {a, b, c, d} may be represented as a < b ≅ c < d.

3.2 Average height of x

The calculation of a weak order of a MIS can formally done in three steps::

Step 1: Calculate the partial order relations according to equation (1).

Step 2: Calculate all $h(x, L_i)$,see p) in section 3.1

Step 3: The average height hav(x) is defined as follows:

$$hav(x) = \frac{\sum_{L_i} h(x, L_i)}{LT} \qquad (9)$$

The quantity, *LT* is the number of all linear extensions of a poset.

Inserting probabilities for getting a certain height $prob(h(x,L_i)$, equation (9) can also be written as:

$$hav(x) = \sum_{r=1,\ldots,n} prob(h(x, L_i) = r) \cdot r \qquad n=|X| \qquad (10)$$

and is the starting point for deriving explicit analytical expressions for *hav*. A weak order is obtained from *hav*(x) , x ∈ *X* as follows: $x_i \le x_j$: ⇔ *hav*(x_i) ≤ *hav*(x_j) (11)

The study of appropriate approximations is of high interest in the theoretical, computationally oriented literature. One of the best exact methods is the lattice theoretical approach [14], other attempts are the method of cumulative rank frequencies [15]. A widely used approximation is that of [4], which is however time consuming (time complexity: $O(n^3 * \log(n))$ with $n=|X|$). Therefore other approximations became more and more popular, which are based on the Local Partial Order Model (LPOM) and are the core of this paper. The time complexity for LPOM is $O(n^2)$ [8]. In section 3.3 a precise characterization of both approximations, LPOM0 and LPOMext, is given.

3.3 Structure parameters

3.3.1 Partitioning X

Definition: Environment of x ∈ *X*: *env*(x) := {y ∈ *X*: y ⊥ x}

Observation 1: *X* = *U*(x) ⊕ *env*(x). I.e. By selecting an object x, the set *X* can be partitioned into *env*(x) and *U*(x). The proof can be easily performed and is suppressed here.

Example: Let'sMEC be a generating element. From Figure 1, we find:
env(MEC) = {MEC, HCL, CHL, HCB, ALD, DDT, LIN} and
U(MEC) = {DDE, DDD, DIE, PCN, PCP}.

3.3.2 Characterization of the two Local Partial Order Models

1) LPOM0: All elements of *U*(x) are considered as isolated elements, i.e. they have ∞ distance to x and among each other; *env*(x) is a chain with the basic constituents, the principal downset *O*(x) and the principal upset *F*(x) (see d) and f) in section 3.1). The quantity *hav*(LPOM0,x) depends on the number of the elements of *O*(x) and *F*(x). Because $n=|X|$ is related with |*O*(x)|, |*F*(x)| and |*U*(x)| by $|F(x)|+|O(x)|+|U(x)|=n+1$, *hav*(LPOM0,x) can be either expressed in terms of |*U*(x)|, |*O*(x)| and |*F*(x)| or in terms of |*U*(x)|, |*O*(x)| and *n* (see [1,5]).

2) LPOMext: For each element y ∈ *U*(x), it is determined *U*(y) ∩ *O*(x) and *U*(y) ∩ *F*(x). Then approximately *hav*(x) can be calculated as follows:

$$hav(LPOMext, x) = |O(x)| + \sum_{y \in U(x)} \frac{|U(y) \cap O(x)|}{|U(y) \cap O(x)| + |U(y) \cap F(x)|} \qquad (12)$$

For details, see [7].

Remark: It should be clear that by evaluating equation (12) no attempt is done to take the graph-theoretical connections among the elements in $U(x)$ and of $env(x)$ into account. A major aim of this paper is to clarify the influence of possible order relations among the elements of $U(x)$. In order to study the role of approximations in detail we introduce the "structure parameters" of a Hassediagram.

3.3.3 Possibilities to obtain structure parameters of Hassediagrams

Consider a specific element $x \in X$ in PO. Then, with respect to x, there are:

- elements which are above x: $\mu_1=|F(x)-\{x\}|$,
- elements which are below x, $\mu_2=|O(x)-\{x\}|$,
- eventually elements which are isolated $\mu_3=|Iso(X)|$,
- elements which are connected with x in an ordinary graph-theoretical sense, which however are incomparable with x: μ_4.

The set of parameters μ_i is a simple example of "structure parameters" of the poset.

Remark: We see the need of a rigorous theory of structure parameters of Hassediagrams. However such a theory is beyond the focus of this paper. Instead of μ_i we need parameters describing different parts of a Hassediagram in a more detailed way. Hereto we take a pragmatic point of view and introduce the needed parameters heuristically (Figure 2).The idea is to identify the different chains (see i) in section 3,1) in a Hassediagram and to count the number of their elements.

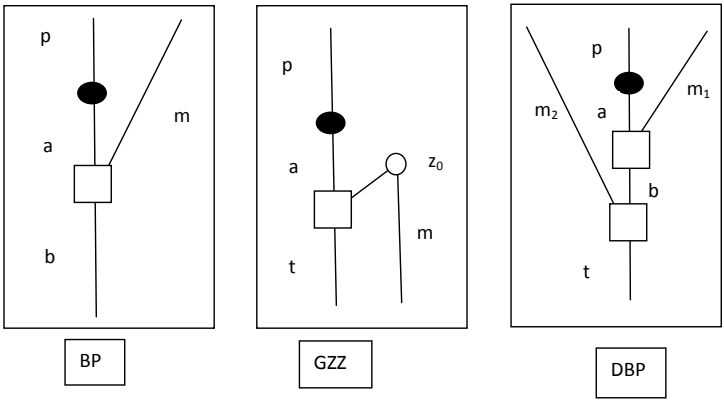

Figure 2. Schematic presentation of Hassediagrams. In all three cases: The black ellipse symbolizes the selected object x, the rectangles the "anchor points"[6], where a bifurcation appears. BP: Bifurcation poset, GZZ: Generalized zig-zag poset, (z_0 is covering the anchor point), DBP: Double bifurcation poset (see text) with lower and upper anchor points.

The meaning of the structure parameters in Figure 2 is explained in Table 2.

3.4 Model - posets

The major difficulty is to assess the role of structure parameters in influencing *hav*(x). We then take the other way round: Defining simple but typical Hassediagrams ("model-posets") and finding out the influence of structure parameters as shown in Table 2 on *hav*(x)= *f*(*structure parameters*). For example in Figure 2,"BP" the structure parameters *p, a, b* and *m* are those which can vary and influence *hav*(x).

Table 2: Heuristically introduced structure parameters of three posets. We call the point where a branch appears, the anchor point "ap"

	BP	GZZ	DBP
p	Number of elements > x	of elements > x	Number of elements > x
a	Number of elements < x and > ap	Number of elements < x and > ap	Number of elements < x and > upper ap
b	Number of elements < ap		Number of elements > lower ap and < upper ap
t		Number of elements < ap	Number of ap
m	branch, || x	Number of elements in the chain below z_0	m_1, m_2: number of elements in the two branches || x

We can speak of a type of Hassediagram(s), by defining structure parameters as was done in Figure 2 and Table 2 and varying them. That means, Monte-Carlo simulations are perfomed by changing structure parameters but keeping unaffected the type of the Hassediagram) and which provide enough different structures to meet the needs of the empirical application of posets. Each model-poset can be characterized by a set of *s* structure parameters. Once analytical formulas are at hand, a whole class of posets characterized by a *s*-tuple of parameters can be evaluated, albeit most often the complex rhs of *f*(*structure parameters*) does not allow a direct inspection. The only exception of the model posets studied here is "BP", which therefore will be used to exemplify the derivation of the rhs of the equation *hav*(x)=*f(structure parameters).*

3.5 Variance based sensitivity analysis

The term "sensitivity analysis" is used here in the context of model outputs. Our models are the analytical functions which relate *hav*(exact), *hav*(LPOM0) and *hav*(LPOMext) to the structure parameter set of certain classes of model-posets. In our case the outputs are the three

types of average heights and the goal is to assess the influence of structure parameters on their values.

Variance-based sensitivity analysis, VB-SA, consists of a set of statistical methods to rank the structure parameters according to their contribution to the output variance. The parameters are varied according to pre-defined probability distributions and some statistics are computed on the model output. The higher the contribution of a parameter to the output variance, the more relevant its role in the estimation of the model output, that is average heights in our case.

VB-SA allows for the estimation of the contribution of each single parameter to the output variance and also of its possible interactions with other parameters, thus instructing the analyst about cooperative behavior of parameters [16-18].

VB-SA is substantially based on conditional variances and expected values. Let's define a model of the form $Y=f(X_1, X_2, \ldots X_k)$, where Y is a scalar output and the X_i are the k parameters following certain probability distributions. Parameters X_i are assumed to be independent.

The main, or first order, sensitivity index for parameter X_i , called S_i, is the share of variance explained by the step-wise linear model which passes through all the points $E(Y| X_i)$. In formula:

$$S_i = \frac{V(E(Y \mid X_i)}{V(Y)} \tag{13}$$

S_i corresponds to the fraction of variance of the output that can be attributed to parameter X_i by itself. A high value of this statistics implies that the parameter is influent.

A so-called total sensitivity index, ST_i ,is also introduced which takes into account the influence of X_i together with all its interactions with other parameters:

$$ST_i = 1 - \frac{V(E(Y \mid X_{\sim i})}{V(Y)} \tag{14}$$

where $X_{\sim i}$ denotes all parameters but X_i. The total index ST_i corresponds to direct effect of X_i, as already accounted for by S_i, plus the indirect effect which X_i has on $V(Y)$ through its interactions with other parameters, if any. Main effects are always lower or equal to total effects and the quantity:

$$v_i = \frac{ST_i - S_i}{ST_i} \tag{15}$$

is the measure of the intensity of interaction effects between parameter X_i and the other model parameters. To obtain S_i and ST_i the estimation of k-dimensional integrals is needed, with k = number of parameters. Model's values Y have to be evaluated at different points within the parameter space, whose dimensionality is equal to the number k of input factors. To explore the k-dimensional parameter space random or quasi-random numbers are generally adopted. Quasi-random numbers are specifically designed to generate samples from the space of input factors as uniformly as possible [19]. Convergence curves of sensitivity index estimates are usually computed in order to check for convergence of the estimates. Quasi-random numbers are here used for the estimation of sensitivity indexes. For computational details see [20,21].

4. Results

4.1 Results for the bifurcation poset (BP)

Structure parameters:

For a representation of this type of model-poset and its structure parameters refer to Figure 2 "BP". The structure parameters are: a: number of elements in the chain below x and the anchor point, b: the number of elements below the anchor point, p: the number of elements above x and m: the number of elements in the side branch connected with env(x) at the anchor point. The elements of the branch are incomparable with x, however: $y_i, y_j \in U(x) \Rightarrow y_i \perp y_j$ ($i,j = 1, \ldots, m$). $m = |U(x)|$

Analytical results:

The basic analytical formula is derived from the starting equation (10).

Step 1: The number of all linear extensions of BP: Following [22] this number can be easily determined in dependence of the structure parameters:

$$LT(BP) = \binom{(a + p + 1 + m)}{m} \tag{16}$$

Step 2: While constructing the linear extensions, elements of U(x) can get positions between the anchor point as well as below or above x. Let us enumerate the elements of U(x) by $0,...,j,...,m$. For the elements $0, \ldots, j$ a position below x, for the elements $j+1,...,m$ a position above x is found, while constructing the linear extensions. Because the elements of U(x) are themselves forming a chain, no additional factor is needed for their eventual permutations. Furthermore we neglect for the moment the "tail" b and the anchor point, so that the elements

of $U(x)$ are merged into the chain given by the elements of part a, x itself and the elements of part p (see Figure 2). Following [22] we obtain the frequency $freq(x,h=a+1+j)$ of linear extensions with the height $a+1+j$:

$$freq(x, h = a + 1 + j) = \binom{a+j}{j} \cdot \binom{m-j+p}{m-j} \tag{17}$$

From equation (17) we can say that the number of linear extensions, where x gets the height $a+1+j$, depends not only on the number of linear extensions where elements $y \in U(x)$ are inserted below x but also how many realizations are possible for y to be positioned above x.

Step 3: If the "tail" ($b+1$) is taken into regard, the final expression for $hav(x)$ is given by equation (18):

$$hav(exact, x) = \frac{\sum_{j=0}^{m} \binom{a+j}{j} \cdot \binom{m-j+p}{m-j} \cdot (a+b+2+j)}{\binom{m+a+p+1}{m}} \tag{18}$$

The term:

$$prob(x, h(x, L_i) = (a+b+j+2)) = \frac{\binom{a+j}{j} \cdot \binom{m-j+p}{m-j}}{\binom{m+a+p+1}{m}}$$

is the probability for having height ($a+b+j+2$) for the considered element x, taken all linear extensions L_i ($i = 1, ..., LT$) into account (for the sake of simplicity denoted as: $prob(x,h)$).

Step 4: Equation (18) can also be written as follows:

$$hav(exact, x) = (a+b+2) \cdot \sum_{j=0}^{m} prob(x, h) + \frac{\sum_{j=0}^{m} \binom{a+j}{j} \cdot j \cdot \binom{m-j+p}{m-j}}{LT(BP)} \tag{19}$$

The sum in the first term in equation (19) includes all probabilities, hence:

$$\sum_{j=0}^{m} \binom{a+j}{j} \cdot \binom{m-j+p}{m-j} = \binom{m+a+p+1}{m} \tag{20}$$

Equation (20) can be considered as a binomial identity and is very useful in all analytical derivations of $hav(x)$.

In the second term in equation (19) the presence of factor *j* does not allow for any immediate simplification. Some further steps are needed:

Step 4: A new summation index *k*=*j*-1 is introduced, then the numerator in the second part of equation (19) (we call this part *B*) reads as follows:

$$B := \sum_{k=0}^{m-1} \left(\frac{(a+k+1)!}{a!\cdot k!} \right) \cdot \binom{m-k-1+p}{m-k-1}$$

Step 5: By the identity

$$\binom{x}{m-1} = \frac{m}{x-m+1} \cdot \binom{x}{m} \tag{21}$$

and intermediate substitutions we arrive finally at:

*hav(*exact,x*) = (a+b+2) + m*(a+1)/(a+p+2)* (22)

As a side result we add the following binomial identity, which so far we did not find in the corresponding combinatorial literature:

$$\sum_{j=0}^{m} \binom{a+j}{j} \cdot j \cdot \binom{m-j+p}{m-j} = (1+a) \cdot \binom{a+m+p+1}{m-1} \tag{23}$$

We want to compare equation (22) with the simplest approximation, LPOM0, therefore we derive first the *hav*(LPOM0) in terms of the structure parameters:

*hav(*LPOM0,x*) = (a+b+2)+ m*(a+b+2)/(a+b+p+3)* (24)

and secondly we see that there is a discrepancy in comparison to the exact formula. The reason is the structure parameter *b*. Indeed, if the elements of *U*(x) are considered as isolated, they can be merged everywhere in *env*(x), while constructing the linear extensions. Now we compare equation (22) with the more involved approximation, LPOMext. In terms of the structure parameters we arrive at:

hav(LPOMext,x) = *(a+b+2) + m*(a+1)/(a+p+2)* (25)

Hence for the BP-model we find: *hav*(LPOMext) = *hav*(exact).

It may be of interest, how the part (*a*+*b*+2) can be explained. This term appears in every equation (22, 24, 25). The term (*a*+*b*) is the part of the downset *O*(x); the anchor point and object x itself count 1 in the final rank. Because the *m* elements of *U*(x) merge to different amount j into the downset *O*(x) the equation 18 additionally includes the term *j*.

Summarizing, and answering one of the questions mentioned above about LPOM0, LPOMext, and exact method: *hav*(exact) = *hav*(LPOMext) ≠ *hav*(LPOM0)

Approximation quality:

From the analytical results it can be concluded that the larger b, the larger the deviation of *hav*(LPOM0) from *hav*(exact).

Variance based sensitivity:

The parameter set in the case of the bifurcation poset is $\{X_1 = p, X_2 = a, X_3 = b, X_4 = m\}$.

For the sensitivity analysis, all the parameters are assumed to follow a discrete uniform distribution U_d [1,30]. Convergence curves are computed for eight different sample sizes *N* which go from 2^5 to 2^{12}. Convergence curves for main and total sensitivity indexes are shown in Figures 3-6. the four parameters and both models, *hav*(exact) and *hav*(LPOM0). We recall that in this case *hav*(exact) = *hav*(LPOMext).

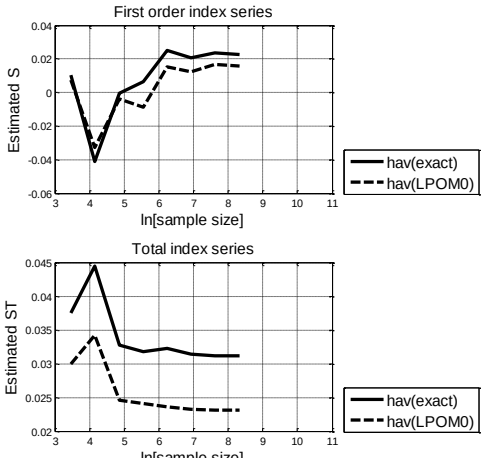

Figure 3. Bifurcation Poset: convergence curves for parameter X_1=*p*

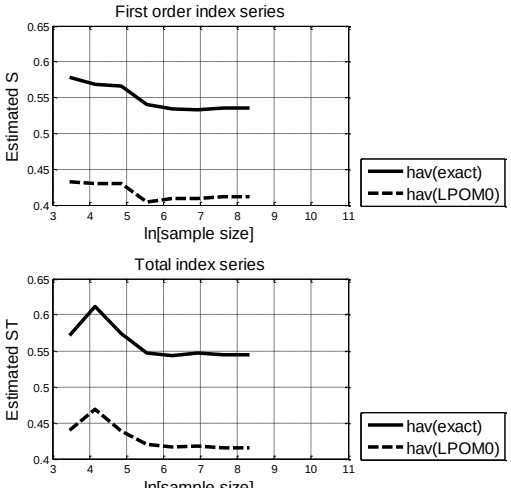

Figure 4: Bifurcation Poset: convergence curves for parameter X_2=a

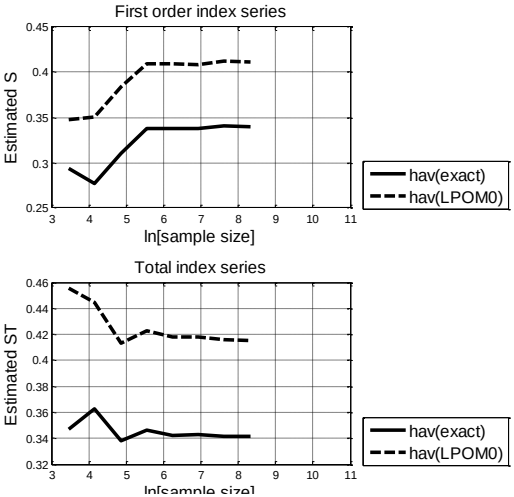

Figure 5. Bifurcation Poset: convergence curves for parameter X_3=b

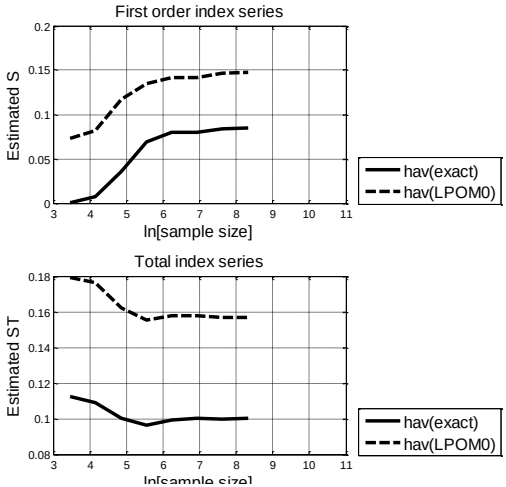

Figure 6. Bifurcation Poset: convergence curves for parameter X_4= m

At the maximum simulated sample size N=2^{12} convergence is reached for both models and all parameters. Main and total order sensitivity indexes are shown in Table 3.

Table 3. Bifurcation Poset: estimated sensitivity indexes at convergence

	hav(exact)			*hav*		
	S_i	ST_i	v_i	S_i	ST_i	v_i
X_1=p	0.0226	0.0312	0.2756	0.0160	0.0231	0.3074
X_2=a	0.5354	0.5440	0.0158	0.4122	0.4157	0.0084
X_3=b	0.3401	0.3414	0.0038	0.4114	0.4156	0.0101
X_4=m	0.0844	0.1001	0.1568	0.1474	0.1570	0.0611

Parameters can be ordered by their level of influence with respect to the output variance. For the bifurcation model sensitivity analysis shows that: $a > b > m > p$. Interactions are generally low. Values of v_i (equation 15) indicate that at most 30% of the parameter influence is due to interactions, with parameter p being the most interactive one. Parameters can also be ranked according to their estimated level of interactions: the resulting order is $p > m > a > b$.

4.2 Results for the generalized zig-zag poset (GZZ)

Structure parameters:

The structure parameters of GZZ-posets can be seen in Figure 2, "GZZ" and in Table 2.

Model posets of class GZZ provide examples where one of the subsets connected is a chain, but not all the members of the chain are comparable to the anchor point. Note, there are m+1 elements in U(x). The analytical calculations (once again starting from equation (10)), are performed by taking care that -depending on the position of z_0 in the resulting linear extension- some of the elements $\in U$(x) may be positioned above and some others below x. It turned out that it is convenient to count element z_0 separately.

Analytical results:

Although the rhs of *hav*(x) = *f(structure parameters)* could be presented, not much can be derived from them by inspection except that the expression (*t*+*a*) appears in *f*. We therefore omit most the formulas. The formulas corresponding to the approximations LPOM0 and LPOMext are:

$$hav(x) = (a+t+2)+[(a+t+2)*(m+1)]/(a+t+p+3) \quad \text{LPOM0} \quad (26)$$

$$hav(x) = (a+t+2) + (a+1)/(a+p+2) + m*(a+t+2)/(a+t+p+3) \quad \text{LPOMext} \quad (27)$$

The derivation of (26) and (27) is not difficult because the leading elements in LPOM0 and LPOMext, O(x), F(x), U(x), U(y) can be expressed in terms of the structure parameters. Once again there is the quantity *(t+a)* in common across all the expressions (26)-(27) as well as in the exact but unhandy expression for *hav*(x), meaning that all the heights in all the linear extensions will be not less than (*t*+*a*). Comparing (26) with (27) we note that: *hav(*LPOM0*)* ≠ *hav(*LPOMext*).* It seems hopeless to "boil down" equation *hav*(exact)=*f*(*structure parameter*) to an expression by which a comparison with (26) and (27) can be directly carried out. Hence it is assumed: *hav(*LPOM0*)* ≠ *hav(*exact*)* and *hav(*LPOMext*)* ≠ *hav(*exact*)*

The three methods yield different formulations for *hav*(x). We will show that there are ranges of structure parameters where the deviation of *hav*(LPOM0) from *hav*(exact) is less than that of *hav*(LPOMext).

Approximation quality:

When a Monte Carlo simulation is performed in order to compare the three methods which each other, we find (Table 4):

Table 4. Approximation quality of GZZ-posets

poset		structure	mean(|*hav*(exact)-*hav*(LPOM0)|)	mean(|*hav*(exact)-*hav*(LPOMext)])
Generalized zig-zag	GZZ	*a* , *t* , *p* , *m*	-0.035	0.132

When the generalized zig-zag poset is considered, the LPOM0-model seems to be better than the more sophisticated LPOMext-model. In the discussion an explanation is given.

Variance-based sensitivity:

The parameter set for the GZZ poset is $\{X_1 = a, X_2 = t, X_3 = p, X_4 = m\}$.

For the sensitivity analysis the four parameters are assumed to vary in a discrete uniform distribution $U_d[1,30]$, as for the Bifurcation poset, and convergence curves are computed for sample sizes from $N=2^5$ to $N=2^{14}$. We suppress here the graphics and report only the numerical results. Convergence is always reached within the highest sample size assigned. Sensitivity indexes at convergence are listed in Table 5.

Table 5. GZZ Poset: estimated sensitivity indexes at convergence

	hav(exact)			*hav*(LPOM0)			*hav*(LPOMext)		
	S_i	ST_i	v_i	S_i	ST_i	v_i	S_i	ST_i	v_i
$X_1=a$	0.4126	0.4158	0.0077	0.4122	0.4154	0.0077	0.4183	0.4214	0.0074
$X_2=t$	0.4114	0.4146	0.0077	0.4122	0.4154	0.0077	0.4063	0.4093	0.0073
$X_3=p$	0.0189	0.0250	0.2440	0.0190	0.0250	0.2400	0.0193	0.0253	0.2372
$X_4=m$	0.1469	0.1554	0.0547	0.1463	0.1549	0.0555	0.1461	0.1546	0.0550

The three methods (exact, LPOM0 and LPOMext) yield very similar values of main and total sensitivity indexes, therefore all the sequences found for the different methods do not contradict each other. The order of the parameters is $a > \approx t > m > p$, so that *m* and *p* are detected as the least influencing ones as in the Bifurcation poset case. Also in this case the level of interactions is generally low. Parameter *p* is the most interactive one as for the previous case.

4.3 Results for the double – bifurcation poset (DBP)

Remark:

In the DBP poset model, whose graph is shown in Figure 2 and Table 2. Two sets of incomparable elements $U_1(x) \subseteq U(x)$ and $U_2(x) \subseteq U(x)$ are present. With $y_1 \in U_1(x)$ and $y_2 \in U_2(x)$ and $y_1 \parallel y_2$ a remarkable step to a generalization is done, although both subsets $U_1(x)$ and $U_2(x)$ are chains and are comparable with their upper and lower (resp.) anchor points.

Structure parameters:

The specificity of this model-poset is that two different systems of elements incomparable to x (the selected element) are considered. This class of posets is described by 6 structure parameters: m_1, m_2, t, p, a and b (see Figure 2 and Table 2).

Analytical results:

The height *hav*(x) depends on the 6 structure parameters; however no analytical information about the structure parameters can be derived. The effort of the analytical derivation of an equation for *hav*(x) is just the simplification for the Monte Carlo calculation, needed to get the variance based sensitivities. It is convenient to set $|O(x)| = T$, which is the minimum height, the element x can get.

$$T = a+b+t+3 \tag{28}$$

LPOM0:

$$hav(x) = a+b+t+3+(m_1+m_2)\cdot\frac{a+b+t+3}{a+b+t+p+4} \tag{29}$$

LPOMext:

$$hav(x) = a+b+t+3+m_1\cdot\frac{(a+1)}{(a+p+2)}+m_2\cdot\frac{a+b+2}{a+b+p+3} \tag{30}$$

In this case we then have:

hav(LPOM0) $\neq hav$(LPOMext), hav(LPOM0) $\neq hav$(exact), hav(LPOMext) $\neq hav$(exact)

We see that in (29) the term $(a+b+t)$ as well as the term (m_1+m_2) influence *hav*(x). As long these sums are equal, any variation in a, b, t on the one side and m_1, m_2 on the other side will not change the average height of x within the local partial order model LPOM0.

In LPOMext and the exact method none of these two symmetries is kept.

Approximation quality:

A Monte Carlo simulation is carried out with $2 \cdot 10^5$ runs with structure parameters randomly sampled in a Discrete Uniform $U_d[1,10]$. The robustness of the MC-simulation was controlled by checking the results for 20, 200, 2000, 20000 and 200000 runs. I.e. the mean of |*hav*(exact)-*hav*(LPOM0)| and of |*hav*(exact)-*hav*(LPOMext)| is checked for 20 to 200000 runs. Results are converging (Table 6).

Table 6. Approximation quality of Double-Bifurcation-posets

poset	class	e parameters	mean(|hav(exact)-hav(LPOM0)|)	mean(|hav(exact)-hav(LPOMext)|)
Double bifurcation poset	DBP	m_1, m_2, t, p, a, b	2.037	0.2003

Variance based sensitivity study (exact method):

The parameter set for the DB poset is $\{X_1 = m_1,\ X_2 = m_2,\ X_3 = t, X_4 = p,\ X_5 = a, X_6 = b\}$.

For the sensitivity analysis the six parameters are assumed to vary in a discrete uniform distribution $U_d[1,30]$, as for the previous cases, and convergence curves are computed for sample sizes from $N=2^5$ to $N=2^{15}$. Once again, we only report the numerical results. Convergence is always reached within the highest sample size assigned. Sensitivity indexes at convergence are listed in Table 7.

Table 7. DB Poset: estimated sensitivity indexes at convergence

	hav(exact)			*hav*(LPOM0)			*hav*(LPOMext)		
	S_i	ST_i	v_i	S_i	ST_i	v_i	S_i	ST_i	v_i
$X_1=m_1$	0.0411	0.0528	0.2216	0.1077	0.1138	0.0536	0.0476	0.0576	0.1736
$X_2=m_2$	0.0755	0.0824	0.0837	0.1083	0.1135	0.0458	0.0861	0.0921	0.0651
$X_3=t$	0.1943	0.1959	0.0082	0.2469	0.2501	0.0128	0.1948	0.1963	0.0076
$X_4=p$	0.0435	0.0542	0.1974	0.0255	0.0330	0.2273	0.0446	0.0546	0.1832
$X_5=a$	0.3868	0.3976	0.0272	0.2465	0.2501	0.0144	0.3642	0.3723	0.0218
$X_6=b$	0.2330	0.2355	0.0106	0.2464	0.2501	0.0148	0.2415	0.2439	0.0098

5. Discussion

For the sake of clarity and in order to avoid repetitive explanations we summarize here the basic concepts used for the discussion.

a) "merging": The heights $h(x, L_i)$ are the final outcome of an extension process (see Section 2, points o) and p)). For each incomparable element $y \in U(x)$ it has to be checked where y can be located and how many possibilities below and above x are available. The process of assigning y to one of these locations in the poset-extensions is called a merging process.

b) For y ∈ U(x) a part of *env*(x) is accessible in order to construct poset-extensions.

c) Combinatorial effect. Let $m = |U(x)|$, then there are two permutations to be taken into account: (i) Permutations of the elements of U(x): The selection of j elements of U(x) can be done in $\binom{m}{j}$ different ways, when elements of U(x) are forming an antichain. (ii) j elements of U(x) can be combined with parts of the environment. Let r be the number of elements in the considered part of the environment. Then the leading combinatorial equation is:

$$LT(extension) = LT(U(x)) \cdot LT(part\,of\ env(x)) \cdot \frac{(m+r)!}{m!r!} \quad [22].$$

Therein the term *LT(part of env(x))* represents the combinatorial effect, see also [7].

d) "Dilution": Elements of U(x) can be merged above x and below x. When they are located above x, the height *hav*(x) is basically not affected because the height of x is influenced only by elements below x. However their influence is indirect because any element y ∈ U(x) merged above x is a missing candidate for a position below x. The effect of elements of U(x) on *hav*(x) is "diluted" by the part of the environment of x which is above x.

e) "Competition": Let pos_b be the number of positions below x and pos_a the number of positions above x, then the probability of y ∈ U(x) to be merged below x is $pos_b/(pos_b+pos_a)$. The greater the pos_a, the greater the competitive effect of the part above x with respect to the part below x.

f) According to the basic idea of model-posets *hav*(x) is always a function f of structure parameters. Sometimes f can be simplified so that no sum-sign and no combinatorial expression such as in c) eventually appear. In this case it can be said that combinatorial effects are cancelling out. In some other cases, such as in GZZ or DBP, such a simplification seems not to be possible. We assume then that combinatorial effects do not cancel out and are influencing factors in the computation of *hav*(x).

5.1 Bifurcation model-poset

Approximation quality:

The LPOM0 method necessarily deviates from the other two methods (exact and LPOMext).

Due to the assumption of elements of U(x) being isolated (they belong to an antichain) they can be merged in the whole environment of x. However, as in reality the elements of U(x) are to be located above the anchor point, they can only be located above the elements in b in any extension the elements of U(x).

LPOMext and exact methods coincide and the function *f* can be simplified. This shows that the role of combinatorial effects cancels out and confirms the quality of LPOMext for those posets belonging to the BP-class of model-posets.

Sensitivity analysis interpretation, models: *hav(*exact*)* and *hav(*LPOMext*)*

The sensitivity analysis clearly shows that the structure parameter *a* has the highest impact. The elements of the *a*-part contribute directly to *hav*(x) and the elements of *U*(x) merged within the *a*-part while constructing poset-extensions contribute to *hav*(x) as well. Structure parameter *b* also contributes to the downset of element x, but the *b*-elements are not accessible by the elements in *U*(x) while constructing poset - extensions.

The influence of parameter *m* depends on structure parameters *a* and *p*. The role of *p* is only indirect and can be interpreted as dilution.

Sensitivity analysis interpretation, model: *hav*(LPOM0)

a) main sensitivity indexes, S_i:

The *m* elements of *U*(x) are considered as isolated and a merging into all three parts of the poset, namely the parts *b*, *a* and *p* of the original poset is possible while constructing the poset-extensions. That means that the probability to be merged with the *a*-, *b*- and *p*-elements depends on *m* and on (*a*+*b*+*p*+1). The *b*-part and the *a*-part play the same role, while the *p*-part is competing for the elements of *U*(x).

b) total sensitivity indexes and interaction effects, ST_i and v_i:

Total and main sensitivity indexes are always not very different meaning that the contribution of interactions among the structure parameters is low. When the quantity v_i (equation 15) is considered more closely, it can be seen that the order of parameters due to S_i is completely reversed. The structure parameter having the highest direct impact on *hav*(x) contributes least to interactions.

In the case of LPOM0, *b* and *a* both influence *hav*(x) because they belong to the environment of x. There is no difference between the effect of *a* and *b*. Structure parameter *p* has the least influence because it only acts by competing for the elements of *U*(x) without directly increasing *hav*(x).

5.2 Generalized zig-zag model poset

Approximation quality:

The reason for the good results of LPOM0 can be explained qualitatively. Incomparable elements can be merged above the whole environment of x. The crucial assumption of the

LPOM0 model-poset can be fulfilled. As the formula for GZZ (exact) cannot be simplified to get a direct information about the status of LPOMext versus exact method, we must rely on numerical results which show that LPOMext and exact do not necessarily coincide. Obviously the role of combinatorial effects cannot be cancelled out in this case.

Sensitivity analysis interpretation, all models - *hav*(LPOM0), *hav*(LPOMext) and *hav*(exact):

As for the Bifurcation model-poset - BP - the parts which contribute to the downset of x have the highest impact while *p* has the least one. The reasons for this order of parameters are similar to those already discussed for the BP model-poset.

The remarkable similar impact of *a* and *t* can be attributed to the general assumption of $U(x) - \{z_0\}$ as isolated points. With respect to the LPOM-approximations: *a* and *t* are interchangeable in the analytical expression for average heights.

5.3 Double bifurcation model-poset

Approximation quality:

The model LPOMext is definitely better than LPOM0, which is not surprising because the accessibility of the vertices, described by the structure parameter *t* is not possible and that of the elements described by parameter *b* is only partially possible. In LPOM0 however there is made no distinction between the *b*- and *t*-elements.

Sensitivity analysis interpretation (all models):

a) main sensitivity indexes, S_i:

The order of importance of the parameters is not exactly the same among the three models to calculate *hav*. According to both S_i and ST_i for the exact model we have $a>b>t>m_2>p>m_1$. In the LPOM0 model the order of importance is $t>a\approx b>m_2\approx m_1>p$ while in the LPOMext case the order is $a>b>t>m_2>m_1>p$. The common feature of the three models is that parameters *a*, *b* and *t* are among the most important while m_1 and *p* are always the least influencing ones.

Once again the structure parameter *a* has the highest influence, because its variation causes not only a linear variation in *T* (equation 28), but also influences the number of positions for m_1- and m_2-elements to be located below x. Similarly *b* contributes additively to *T* but is also influencing the positions of the m_2 elements, below x. Finally *t* is a summand in *T*, however the m_1 and m_2-elements have no access to the corresponding positions because of preservation of order relations in constructing extensions.

b) total sensitivity indexes and interaction effects, ST_i and v_i:

Also in this case the level of interactions is generally low. The m_1 and m_2-elements do not contribute to T (equation 28) but only fractions of them, expressed by $k1$ and $k2$. Nevertheless their influence is definitely different: $S_i(m_2) > S_i(p) > S_i(m_1)$. Corresponding to the a- and b-part of the downward chain of x the effect of a variation of m_2 is higher than that of m_1.

Considering the interaction due to the structure parameter, it is clear that the parameter group (m_2, p, m_1) must be at the top: All three parameters do not appear in T (27) and their impact on *hav*(x) depends on the values of the structure parameters a and b.

6. Conclusions

Ranking of objects, such as persistent organic pollutants, according for instance to their risk level is an important task in chemical evaluation. When multiple indicators are simultaneously observed on a set of chemicals, there are two main approaches to get a rank of chemicals: a) computing an aggregated measure for each chemical based on the observed indicator set or b) applying partial order theory to estimate the average height (average rank) of each chemical from the concept of linear extensions. This latter option is our setting in this paper. We then focus on a specific object x in a set of chemicals and provide different estimates of its average height in the partial order – *hav*(x) - induced by the multi-indicator system observed on the chemicals under analysis. Three types of estimations, called models, are discussed in this paper: "exact", "LPOM0" and "LPOMext".

The example taken from environmental chemistry [9] is the motivating starting point. There the authors could compare the results of aggregations with *hav*(exact). They have shown the pretty good coincidence of *hav*(approx) with *hav*(exact) and that there are discrepancies to the rankings due to composite indicators. In general however, *hav*(exact) is not available and then the results of our study can be applied to the analysis on larger lists of chemicals or even generalized to general sets of objects characterized by a multi-indicator system.

The graph-theoretical representation of partial order by directed, acyclic graphs motivates the concept of "structure parameters" of a poset. The concept "structure parameter" is in turn the outcome of the idea that a partial order can be considered by a system of "environment" and elements incomparable with x.

To investigate the influence of poset structure parameters on the estimates of *hav*(x) it is extremely helpful to express *hav* explicit as a function of the structure parameters.

Here, three different types of posets are considered: the Bifurcation Poset (BP), the Generalized Zig-Zag Poset (GZZ) and the Double-Bifurcation Poset (DB). These types are

posets are denoted model-posets. In fact, today there are 10 model-posets investigated [23]. For example, when the poset is described by a chain plus a set of isolated elements or by two chains then the analytical derivation of the exact value as well as LPOMext and LPOM0 lead to coinciding formulas. Combinatorial effects cancel out.

When the elements of $U(x)$ are connected with x in the ordinary graph-theoretical sense, then the consequences are:

- Need of a classification of $U(x)$ (as done in DBP model-poset) .
- The three models exact, LPOM0 and LPOMext begin to diverge, but LPOM0 is not necessarily the worst approximation, as shown by the GZZ model-poset case.

A variance-based sensitivity analysis is carried out for all the three model-posets in order to assess the structure parameter importance with respect to different *hav*(x) estimates. The sensitivity analysis supports this environmental concept.

We conclude that the concept of "environment" in application on posets seems to be useful and should be extended.

Another aspect is important to be underlined. In chemistry molecular structures can be either characterized by quantum mechanical concepts or by graph-theoretical invariants, i.e. the topological indicators. There are for example concepts describing the branching, number of paths or diameters of a molecule (just to mention some few among thousands of such topological indicators, see [24, 25]). The usage of topological indicators is of great importance for forecasting properties of the chemicals. Topological indicators are methodologically considered as predictors in a statistical model. Similarly, quantities such as *hav*(x) may be possibly described by a system of graph-theoretical invariants for directed graphs, i.e. a rigorous theory of partial order structure parameters is needed.

References

[1] R. Bruggemann, G. P. Patil, *Ranking and Prioritization for Multi–Indicator Systems – Introduction to Partial Order Applications*, Springer, New York, 2011.

[2] J. P. Brans, P. Vincke, A preference ranking organisation method: The PROMETHEE method for MCDM, *Manag. Sci.* **31** (1985) 647–656.

[3] P. Winkler, Average height in a partially ordered set, *Discr. Math.* **39** (1982) 337–341.

[4] R. Bubley, R. M. Dyer, Faster random generation of linear extensions, *Discr. Math.* **201** (1999) 81–88.

[5] R. Bruggemann, P. B. Sørensen, D. Lerche, L. Carlsen, Estimation of averaged ranks by a local partial order model, *J. Chem. Inf. Comput. Sci.* **44** (2004) 618–625.

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

[7] R. Bruggemann, L. Carlsen, An improved estimation of averaged ranks of partial orders, *MATCH Commun. Math. Comput. Chem.* **65** (2011) 383–414.

[8] K. De Loof, B. De Baets, H. De Meyer, Approximation of average ranks in posets, *MATCH Commun. Math. Comput. Chem.* **66** (2011) 219–229.

[9] Y. Sailaukhanuly, A. Zhakupbekova, F. Amutova, L. Carlsen, On the ranking of chemicals based on their PBT characteristics: Comparison of different ranking methodologies using selected POPs as an illustrative example, *Chemosphere* **90** (2013) 112–117.

[10] K. Voigt, R. Bruggemann, H. Scherb, H. Shen, K. H. Schramm, Evaluating the relationship between chemical exposure and cryptorchidism by discrete mathematical method using PyHasse software, *Envir. Model. Software* **25** (2010) 1801–1812.

[11] K. Voigt, R. Bruggemann, M. Kirchner, K. W. Schramm, Influence of altitude concerning the contamination of humus soils in the German Alps: a data evaluation approach using PyHasse, *Environ. Sci. Pollut. Res.* **17** (2010) 429–440.

[12] PyHasse software: www.pyhasse.org (last access: 15.5.2013)

[13] R. Bruggemann, A. Kerber, G. Restrepo, Ranking objects using fuzzy orders, with an application to refrigerants, *MATCH Commun. Math. Comput. Chem.* **66** (2011) 581–603.

[14] K. De Loof, H. De Meyer, B. De Baets, B. Exploiting the lattice of ideals representation of a poset, *Fund. Inf.* **71** (2006) 309–321.

[15] G. P. Patil, C. Taillie, Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization, *Environ. Ecol. Stat.* **11** (2004) 199–228.

[16] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of model output, *Rel. Engin. Sys. Safety* **52** (1996) 1–17.

[17] I. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Sim.* **55** (2001) 271–280.

[18] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Comput. Phys. Commun.* **181** (2010) 259–270.

[19] P. Bratley, B.L. Fox, Algorithm 659 implementing 2001's quasi-random sequence generator, *ACM Trans. Math. Software* **14** (1988) 88–100.

[20] I. Sobol', On quasi–Monte Carlo integrations, *Math. Comput. Sim.* **47** (1998) 103–112.

[21] S. Kucherenko, M. Rodriguez–Fernandez, C. Pantelides, N. Shah, Monte Carlo evaluation of derivative–based global sensitivity measures, *Rel. Engin. Sys. Safety* **94** (2009) 1135–1148.

[22] R. Stanley, *Enumerative Combinatorics*, Wadsworth & Brooks/Cole, Monterey, 1986.

[23] R. Bruggemann, personal communication, 2013

[24] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley–VCH, Weinheim, 2009.

[25] e-Dragon-software: http://www.vcclab.org/lab/edragon (last access 23.2.2013)