Estimation of Averaged Ranks by a Local Partial Order Model[#]

Rainer Brüggemann,*,[†] Peter B. Sørensen,[‡] Dorte Lerche,[‡] and Lars Carlsen[§]

Leibniz - Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310,
D-12587 Berlin-Friedrichshagen, Germany, Department for Policy Analysis,
National Environmental Research Institute, P.O. Box 358, DK-4000 Roskilde, Denmark,
and Awareness Center, Hyldeholm 4, DK-4000 Roskilde-Veddelev, Denmark

Received September 24, 2003

This paper continues the series of publications about applications of partial ordering. The focus of this publication is the derivation of approximate analytical expressions for the averaged rank and the ranking probabilities. To derive such combinatorial formulas a local partial order is suggested as an approximation. The performance of the approximation is rather high; we therefore conclude that three very simple descriptors of the local partial order seem to be sufficient to get a rough impression of the linear order, induced by the averaged ranks and the ranking probabilities of empirical partially ordered sets. Linear order derived from the partial order, ranking probabilities, and other characteristics are considered as parts of a so-called "General Ranking Model" (GRM). Following the local partial order, the averaged rank of an object x can be estimated applying the following simple formula: $Rk_{av} = (S+1)*(N+1)/(N+1-U)$. S is the number of successors of the object x, N is the total number of objects (of the quotient set), and U is the number of objects incomparable with x. More complex formulas for the ranking probabilities are given in the text. A list of abbreviations and symbols can be found in Tables 3 and 4.

INTRODUCTION

The mathematical structure of partial order gains more and more interest. This may be documented not only by the existence of an own mathematical journal, called "Order", but also by an increasing number of publications in chemistry and environmental sciences, as disclosed in e.g. the publications of Randic,[1−3] Klein,[4−6] or Halfon.[7−10] The increasing importance of this mathematical field is further acknowledged through an extra issue of MATCH *(Comm. Math Comp. Chem.)*, which was edited by D. J. Klein and J. Brickmann[11] and by the regular workshops about partial order in chemistry and environmental sciences, which were initialized by the first author in 1998.

Partial order may be useful in ecology, to find appropriate habitat conditions or in environmental chemistry to sort chemicals. Often however the fact that no linear order is found, hampers the applications. Therefore the concept of a General Ranking Model (GRM) was introduced, which is based on the set of linear extensions (for details, see ref 12). From the set of linear extensions an averaged rank, Rk_{av}, can be calculated, as was shown by P. Winkler.[13] As the number of linear extension increases dramatically with the number of objects (approximately with N!, N being the number of objects), gathered in the set to be partially ordered, a straightforward procedure is often not applicable. Thus it was shown in ref 12, how an iterative scheme can be applied: Its central topic is the estimation of the averaged rank by a sequence of randomly created linear extensions. This paper, however investigates if the averaged rank can be estimated in a more direct manner. If this was possible, then we would have three variants to calculate the averaged rank:

1. the straightforward ("exact") calculation, examining the full set of linear extensions,
2. the iterative scheme by a sequence of order preserving maps leading to randomly created linear extensions, and
3. the use of a local partial order model (LPOM).

The paper focuses on the third variant by presenting an approach, based on a local partial order model (abbr.: LPOM), constructed for each object x, taken from the ground set G. A partial order on G is denoted (G, ≤). We assume that G is a quotient set; the objects may be for example chemicals, geographical sites, etc. which are characterized by a tuple $(q_1, q_{,2}, .., q_n)$ of attributes. The equivalence class is induced by the equivalence relation: equality of tuples. For details, see ref 14. In order not to overburden the paper with mathematical sophistication we speak of objects, independently whether the objects are singletons or equivalence classes with more than one object.

METHODOLOGICAL DEVELOPMENT

The Role of Averaged Ranks. As it will be shown that the averaged rank of an object, x, is mainly controlled by the number of objects ranked below (successors of x) and objects ranked above x (predecessors) a warning remark will be useful in advance: The dependence just on the number of successors and predecessors means that an object with many successors will tend to get a higher averaged rank compared with one, which has instead many predecessors. Therefore, to let the average ranking concept senseful, the empirical partial order must be well justified. What does this mean? If two objects x, y are comparable (x > y or y > x)

* Corresponding author phone: +49 30 64181666; e-mail: brg@IGB-Berlin.de.

[†] Leibniz - Institute of Freshwater Ecology and Inland Fisheries.

[‡] National Environmental Research Institute.

[§] Awareness Center.

[#] Partially presented on the occasion of the third INDO-US-Workshop on Mathematical Chemistry, University of Duluth, August 2−August 7, 2003.

10.1021/ci034214m CCC: $27.50
Published on Web 01/24/2004

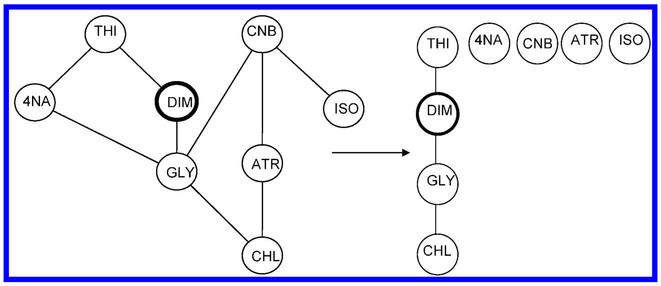


Figure 1. Constructions of a LPOM for one of the eight chemicals, namely DIM. Note, that the chemicals 4NA and DIM would have an identical LPOM, because both chemicals “see” the same number of successors, predecessors, and incomparable objects.

or incomparable (x || y, “||” is the sign for incomparability), then both expressions of the partial order should be robust, even if noise (statistical deviations) is present. If for example the statistical range of data allows in one case $x > y$, and in another case $x < y$ because for example the standard deviations are high enough, then we do not think of an empirical partial order as a well justified one. Only in the case of relevant comparable and incomparable relations the averaged rank is a reasonable construction. Therefore it is highly recommended to perform a preprocessing of data, for example by cluster analysis, see for example ref 15 or by a careful rounding procedure, see ref 16.

Local Partial Order Model (LPOM). A partial order of N objects can be represented by a Hasse diagram. The theoretical base of the so-called “Hasse diagram technique” (HDT) or Partial Order Theory (POT) is explained in several papers, for example see ref 17. We omit therefore in this short communication further details of this method (see also textbooks[18,19]). In this study the application of a local partial order model is the essential step in order to estimate the averaged rank. The basic feature of the local partial order model (LPOM) is that the Hasse diagram (HD) representing the partial order of the quotient set (consisting of N objects (or classes)) and which is called HD_{tot} is substituted by N graphs, called HD_{loc}. As each object leads to its own HD_{loc}, it is sensible to write $HD_{loc}(x)$ in order to refer to the local Hasse diagram found for object x. A LPOM for any object x is obtained counting (i) all successors and (ii) all predecessors of x (in a graphical representation both sets may arranged in any linear extension) and all objects “u_i”, incomparable with x (i.e. all elements u_i || x). The incomparable objects u_i are considered as isolated objects within $HD_{loc}(x)$.

In a recent publication by Lerche et al.[22] 12 high production volume chemicals (HPVC) were discussed. The HD_{tot} representing eight of these HPVC is displayed in Figure 1 (only their identifiers are shown here for illustration as the individual chemicals are not of interest here). Figure 1 further shows how a HD_{loc} for one of the chemicals, DIM, can be created from the HD_{tot}.

The US-Model. Both, in the HD_{tot} and in the HD_{loc} two statements are evident:

The maximum and minimum possible ranks of x, $Rk_{max}(x)$ and $Rk_{min}(x)$, will be as follows:

$$Rk_{max}(x) = S + 1 + U \tag{1}$$

$$Rk_{min}(x) = S + 1 \tag{2}$$

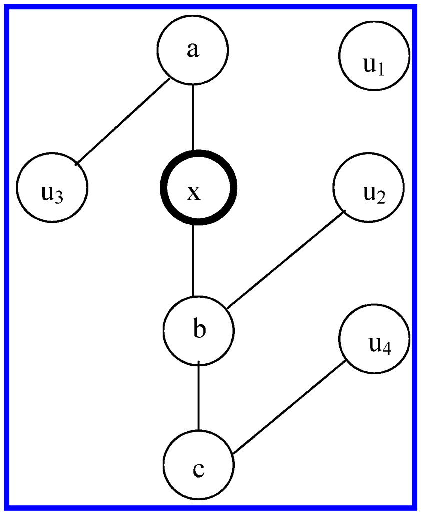


Figure 2. Four incomparable objects to the object x, however differently related to x. For example, an extension (by an order preserving map) such that u_2 will be comparable with x would have 3 positions (above b, x, and a), whereas a similar extension for u_4 would have 4 positions (above c, b, x, and a). The order ideal O(x) encompasses the elements x, b, and c, whereas the order filter F(x) has a and x as its only elements, the **S-x-P** chain: $c < b < x < a$ (see also the text).

S is the number of successors of x (note that the ground set G of the partially ordered set is assumed to be a quotient set), and U is the number of objects incomparable to x. Therefore there is a closed interval $I^{closed} := [Rk_{min}(x), Rk_{max}(x)]$ (also called a “ranking window”) which covers the possible values of the averaged rank. A straightforward idea might be to select the arithmetic mean of Rk_{min} and Rk_{max} as an approximation of the averaged rank of x, $Rk_{av}(x)$.

This approximation within the LPOM model leads to an expression for the true Rk_{av}, which we call $Rk_{av(0).}$

$$Rk_{av(0)} = (Rk_{max} + Rk_{min})/2 = S + 1 + U/2 \tag{3}$$

As $Rk_{av(0)}$ only depends on U and S, we call this approximation, based on LPOM, the US-model.

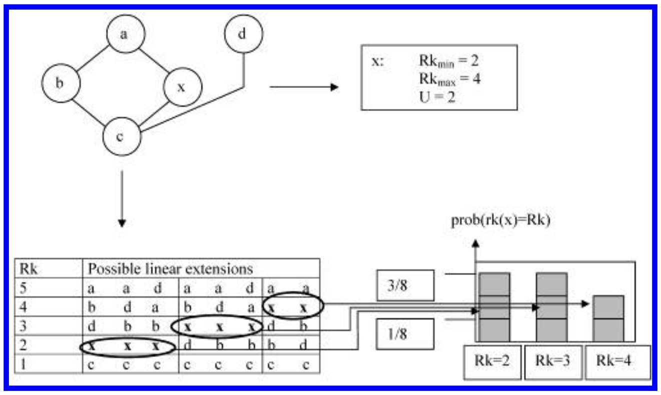
To derive a combinatorial formula for the averaged rank as done above four questions arise:

(1) Is the averaged rank of an object x independent of how the Hasse diagram of the order ideal O(x), i.e., the subposet of successors of x, looks like? (For order ideals, O(x), and the analogous dual construction of a subposet of predecessors of x, called order filter F(x), see ref 18 and Figure 2.)

(2) Objects, which are incomparable to x, may be related to x in several ways as illustrated in Figure 2. They may be isolated objects (u_1 in Figure 2) or incomparable to x but above some successors of x (u_2 and u_4 in Figure 2), and they can also be incomparable to x but below some predecessors (u_3 in Figure 2). To what extent will the different graph theoretical relations of objects, incomparable to x, influence the averaged rank of x?

(3) The US-model implies that the averaged rank will be in the middle of the “ranking window”. Is this implication justified?

(4) Is it possible to refine the estimation of the averaged rank by distributing each element u_i || x. over the **S-x-P** chain?



Rk	Possible linear extensions							
5	a	a	d	a	a	d	a	a
4	b	d	a	b	d	a	x	x
3	d	b	b	x	x	x	d	b
2	x	x	x	d	b	b	b	d
1	c	c	c	c	c	c	c	c

Figure 3. An example for the function prob(rk(x)=Rk), Rk ∈ {Rk_{min}, Rk_{min}+1,...,Rk_{max}}.

Question (1):

(1a) Supposed, there are no incomparable objects to the specified object x: In that case all linear extensions will contain x in the same position, only the number of linear extensions will depend on the structure of the order ideal O(x) and order filter F(x) of x in HD_{tot}.

(1b) Suppose now that there are incomparable elements, then

•there may be additionally an influence of the predecessors and of the structure of the corresponding order filter F(x), and

•there is the question of how the incomparable objects can be distributed over the linear order, which is constituted by x and its successors and predecessors, respectively.

It is still not strictly mathematically proven; however, it seems as if the first question can be answered with a "yes", i.e., that the averaged rank does not depend on the structure of the order ideal or order filter (see also "Discussion").

Question (2): All evidence shows that the assumption within LPOM, namely considering all objects incomparable with x in HD_{tot} as isolated objects in HD_{loc}(x), independently on their different relations to x (compare u_1 to u_4 in Figure 2) is the main and crucial one in order to estimate averaged ranks. A forthcoming paper will show results for tree-like Hasse diagrams.

Questions (3 and 4): To answer these questions the concept of ranking probabilities must first be introduced.

Ranking Probabilities. The set of linear extensions can be seen as another representation of a partial order (for details see ref 19). If each single linear extension had the same weight, one could see the set of linear extensions as a probability space. Counting linear extensions of a certain property and comparing this number with the total number of linear extensions, eP, gives a probability to obtain a certain property. Here it is of main interest to ask for the probability of an object to have a certain rank. In the former section it was shown that the rank of an object varies between two boundaries, which can be easily determined. Thus we may ask for the probability of the rank of an object x, rk(x) having a certain prescribed rank, Rk. Here the possible range for Rk is as follows: {Rk_{min}, Rk_{min} +1,..., Rk_{max}}. Thus, we have to analyze the discrete function prob(rk(x) = Rk). Figure 3 shows an example.

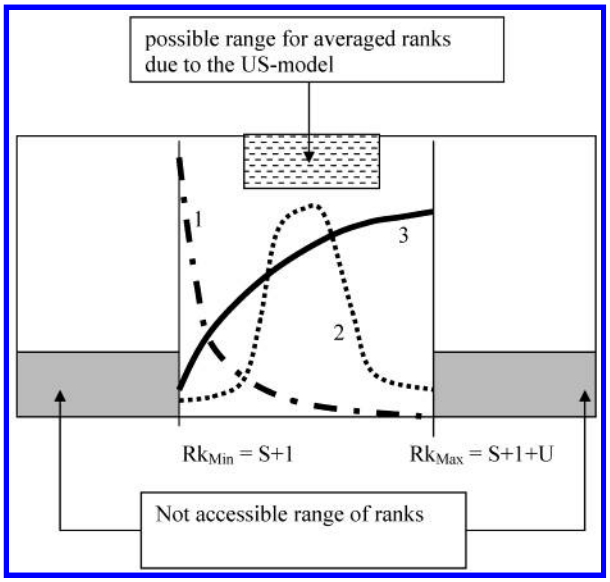


Figure 4. Possible shapes for the prob(rk(x)-Rk)-function and the allowed window for ranking values. The example shown in Figure 3 is a realization of the shape (1).

The function prob(rk(x) = Rk) may have different shapes and—in principle—even different local maxima. Therefore, knowing the interval for the ranks the expectation value for the rank depends still on the shape of prob(rk(x)=Rk). An important simplification is therefore exhibited by the Aleksandrov-Fenchel theorem which states that prob(rk(x)=Rk) found within the set of linear extensions is unimodal.[20] The prob(rk(x)=Rk)-function may be roughly sketched by three different forms, as Figure 4 shows. For the sake of readability in Figure 4 we use continuous curves for the probability density instead of discrete ranking probability values.

Now, the US-model is failing in the case of shape (1) and shape (3) (see also Figure 10 for an example of different shapes). Only in the case of shape 2 the US-model may be appropriate. Therefore, to answer question 3 the US-model has to be extended.

USN-Model. Due to the construction of LPOM there are in general some U objects, which are not comparable with x. The averaged rank will thus depend on how these U objects are distributed within the chain formed by the successors, the object (remember, this may also be a class of objects) x and predecessors (we call this chain **S-x-P**). This was not taken into account in the US-model. In the following approximation the ordering and partitioning of the single incomparable objects in relation to the single successors and the single predecessors is neglected within the LPOM (see Figure 1). All incomparable objects are considered as one single object, which can be located among either the successors or predecessors. This leaves only two possible rankings for x: rk(x) = S + 1 + U, when all the U objects are placed among the successors and rk(x) = S+1, when all the U objects are placed among the predecessors. How often one of the two ranks will be realized depends on the number of successors of x, S and the number of predecessors of x, P. An approximately averaged rank is thus defined as the weighted sum of these possible rankings as

S + 1 positions to put U objects below x. The induced ranking is rk(x) = S + 1 + U.

P + 1 positions to put U objects above x. The induced ranking is rk(x) = S + 1.

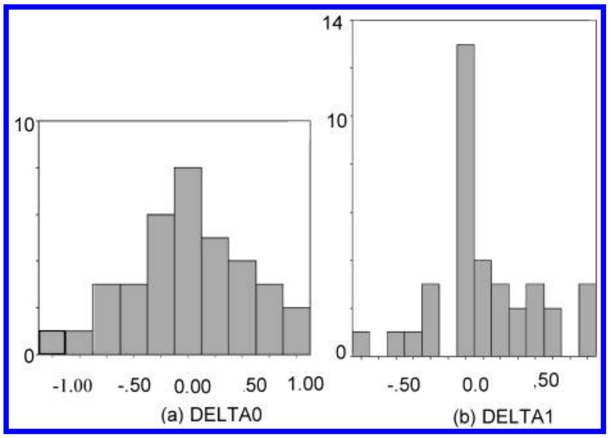


Figure 5. Performance of the US-model (a) and of the USN-model (b).

The averaged rank will therefore be

$$\mathrm{Rk}_{\mathrm{av}(1)} = [(S+1+U)*(S+1) + (S+1)*(P+1)]/[P+1+S+1] \quad (4)$$

Taking into account

$$N = S + 1 + P + U \quad (5)$$

Equation 4 can thus be rearranged into

$$\mathrm{Rk}_{\mathrm{av}(1)} = (S+1)*(N+1)/(N+1-U) \quad (6)$$

The estimation of the averaged rank now depends on U, S, and N. Therefore this model is called the USN-model.

Equation 6 implies the following:

(a) if U = 0 then $\mathrm{Rk}_{\mathrm{av}(1)} = S+1 = \mathrm{Rk}_{\mathrm{av}(0)}$.

(b) if S = P then $\mathrm{Rk}_{\mathrm{av}(1)} = \mathrm{Rk}_{\mathrm{av}(0)}$

(c) if P = 0 then $\mathrm{Rk}_{\mathrm{av}(1)} = S + 1 + U/w \geq \mathrm{Rk}_{\mathrm{av}(0)}$. As w equals $1 + 1/(S+1) \leq 2$

(d) if S =0 then $\mathrm{Rk}_{\mathrm{av}(1)} = (N+1)/(P+2) \leq \mathrm{Rk}_{\mathrm{av}(0)}$

The results (c) and (d) show that the US-model underestimates and overestimates $\mathrm{Rk}_{\mathrm{av}(1)}$ when P = 0 and S = 0, respectively.

Performance. Although the USN-model is still an approximation due to the different nature of incomparable objects and although the US-model seems to be still more approximative in its nature both models are tested. With test examples of partially ordered sets (posets), having 5−7 objects (depending on the structure of the Hasse diagram the computation of the linear extensions can be very time-consuming, when the number of objects exceeds 10) the exact values by the straightforward calculation were compared with those of both models.

For the evaluation of the performance two functions are defined:

$$\text{delta0:} = \mathrm{Rk}_{\mathrm{av\ exact}} - \mathrm{Rk}_{\mathrm{av}(0)}$$

and

$$\text{delta1:} = \mathrm{Rk}_{\mathrm{av\ exact}} - \mathrm{Rk}_{\mathrm{av}(1)}$$

We arrived at a total of 36 cases. In Figure 5a the results are shown for the US-model, and Figure 5b shows the results for the USN-model. It is striking that the USN-model is better than the US-model; however, there is a bias for higher delta1-values. Obviously the USN-model underestimates the exact values more often than overestimates them. This fact, however, may be due to the rather restricted extent of the test set of posets and of objects.

Table 1. Statistical Parameters of the Regression Analysis of Eq 7

model	n	r^2_{DF}	F	t	s
US	36	0.96	863	1.2	−0.8
USN	36	0.98	1526	0.9	0.45

In Table 1 further statistical results are shown. To check both models the linear regression

$$\mathrm{Rk}_{\mathrm{av\ exact}} = t^* \,\mathrm{Rk}_{\mathrm{av}(i)} + s$$

$$\text{Index } i{=}0 \text{ for the US-, } i{=}1 \text{ for the USN-model} \quad (7)$$

was examined. Ideally $t = 1$, $s = 0$, and r^2_{DF}, and the regression coefficient corrected for freedoms, should be 1.

Ranking Probabilities within the USN-Model. To complete the USN-model, the ranking probability of an object to get a prescribed rank Rk, prob(rk(x) = Rk) should be estimated. Of special interest are prob(rk(x) = $\mathrm{Rk}_{\mathrm{min}}$) and prob(rk(x) = $\mathrm{Rk}_{\mathrm{max}}$), respectively. As explained earlier, this task is equivalent to count those linear extensions of $\mathrm{HD}_{\mathrm{loc}}$-(x), where the rank of x is just $\mathrm{Rk}_{\mathrm{min}}$ and $\mathrm{Rk}_{\mathrm{max}}$, respectively. This is a combinatorial exercise, which is extremely simplified by knowing a general formula for the number of linear extensions: If the poset (G, ≤) consists of several disjunct posets G'_i, the number of linear extensions (eP) of the poset (G, ≤) can be calculated by applying eq 8. This equation relates the number of linear extensions of disjunct subposets (G'_i, ≤), eP_i, with the number of linear extensions of the poset (G, ≤), eP.

$G = \oplus_i G'_i$ (the disjunct union of subgroundsets, i.e., $G'_i \cap G'_j = \varnothing$ for all $i \neq j$)

$N_i = \mathrm{card}\ G'_i$ (the number of objects of each subposet)

$$eP = (\prod_i eP'_i) \cdot \frac{(\sum_i N_i)!}{\prod_i (N_i!)} \quad (8)$$

(See for example ref 21.)

To apply this formula within the LPOM-concept, an extension (order preserving map) has to be performed, where the U objects are located either above (estimation of $\mathrm{Rk}_{\mathrm{min}}$) or below x (estimation of $\mathrm{Rk}_{\mathrm{max}}$), see Figure 6.It is now very simple to calculate prob(rk(x) = $\mathrm{Rk}_{\mathrm{min}}$) by taking into account that in the case of $\mathrm{Rk}_{\mathrm{min}}$ the U objects can only mix with the predecessors of x. Therefore eq 8 can be applied, where the one subposet is formed by the predecessors and the other by U isolated objects.

Subposet 1 is the set of U isolated objects. Subposet 2 is the set of predecessors, forming a chain (by assumption of LPOM). Therefore, the number of linear extensions of U objects is U! and of P objects of the predecessor chain 1. The number of linear extensions resulting from a poset where U objects mix only with the predecessors of x is thus

$$eP_P(\text{Mixture of U objects with the chain of P objects}) = U!*1 * (U+P)!/(U!*P!)$$

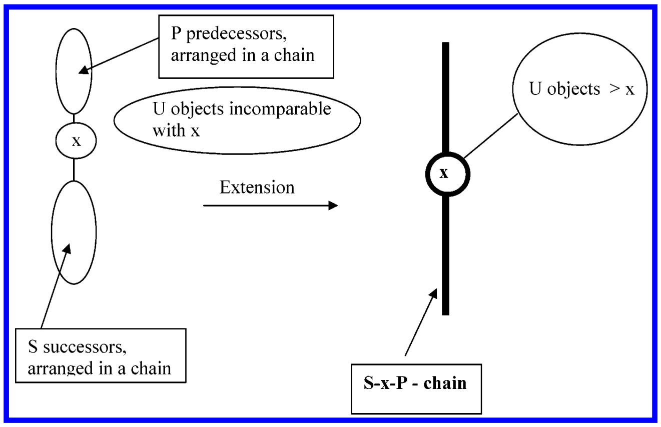

P predecessors, arranged in a chain
U objects incomparable with x
x
Extension
U objects > x
x
S successors, arranged in a chain
S-x-P - chain

Figure 6. Extension of LPOM. Note, that the linear order formed by the successors, predecessors and x is called a **S-x-P** chain.

The total number of linear extensions (mixture of U objects with the **S-x-P** chain) are

$$eP_{tot}(\text{Mixture U, N}-\text{U}) = U! * 1 * N!/[N-U)!*U!]$$

Considering the set of linear extensions as probability space means that now the probability of object x to get the rank = Rk_{min} is just

$$eP_P/eP_{tot} = U!*1 * (U+P)!/(U!*P!)/ \{ U! * 1 * N!/[N-U)!*U!]\}$$

Rearrangement leads to the final equation

$$\text{prob}(\text{rk}(x) = Rk_{min}) = [(N-U)!/N!]*[(U+P)!/P!] \quad (9)$$

and by similar arguments eq 10 can be derived taking now into account the mixing of U objects with the successors (leading to Rk_{max})

$$\text{prob}(\text{rk}(x) = Rk_{max}) = [(N-U)!/N!]*[(U+S)!/S!] \quad (10)$$

It is not difficult to obtain the probability for any rank rk(x) = S + 1 + k, with 0 ≤ k ≤ U, whereby the selection of k objects out of U isolated objects can be done by U!/[(U−k)!*k!] ways.

$$\text{prob}(\text{rk}(x) = S + 1 + k) = \binom{U}{k} \cdot \frac{\left(\frac{(U-k+P)!}{P!}\right) \cdot \left(\frac{(S+k)!}{S!}\right)}{\left(\frac{N!}{(N-U)!}\right)} \quad (11)$$

This formula encompasses the USN-model. A test with empirical posets and with a systematic study of the role of the kind, how objects u_i || x are connected with the **S-x-P**-chain is under operation.

Figure 7 shows some examples of the graph of prob(rk(x) = S+1+k) = f(k). Together with (i) the normalization,

$$\sum_{Rk} \text{prob}(\text{rk}(x) = Rk) = 1 \quad (12)$$

(ii) the Aleksandrov-Fenchel-theorem, (iii) knowledge of the averaged rank, and (iv) eventually calculating some other probabilities, applying eq 11, it should be possible to sketch now a rather realistic graph of the ranking probabilities based on the USN-model and thus for any objects taken from an empirical poset.

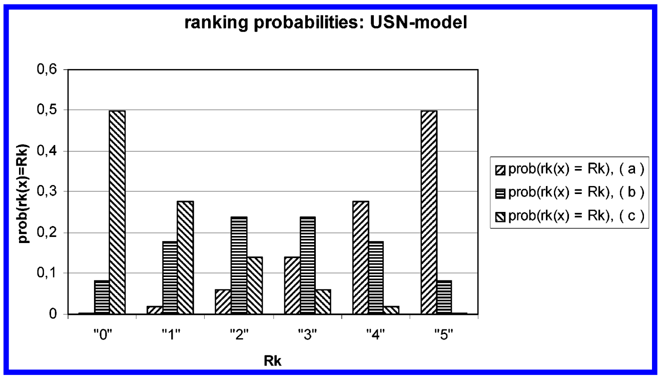


Figure 7. N = 10, U = 5, case (a): P = 0, S = 4, case (b): P = 2, S = 2, case (c): P = 4, S = 0.

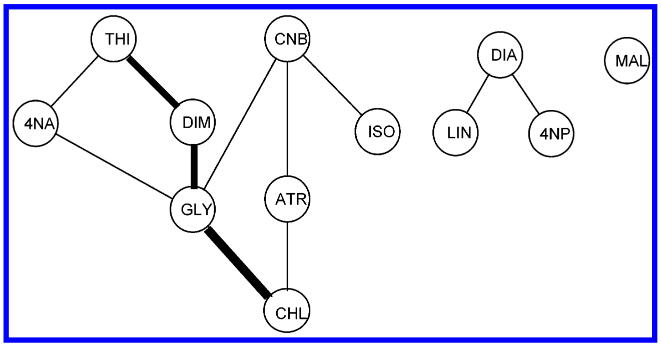


Figure 8. Hasse diagram of the 12 HPVC, discussed in ref 22. One of the longest chains is drawn as a bold line: Here an unambiguous ranking can be obtained.

Example. Considering again the 12 HPVC discussed in ref 22. Here only the results relevant for this paper should be presented: In Figure 8 the HD_{tot} of all 12 chemicals (again only their identifier are shown) is given. The partial order results from a comparison of these chemicals with respect to the four attributes

- production volume,
- acute toxicity,
- accumulation potential, and
- persistence.

The Hasse diagram is shown in Figure 8a consists of three hierarchies (not connected parts of a Hasse diagram) namely one with CHL as a minimal object, one with DIA as maximal object, and one isolated object, MAL. Beyond this an example of the longest chain is shown, where a clear comparison can be done. Often this information is too poor for decision makers, and therefore the GRM shall be applied. Figure 9 shows the straightforward calculated ranking probabilities of four chemicals, ISO, ATR, CHL, and CNB. Note that once again the ranking probabilities are shown as a continuous function (for the sake of clarity). The averaged ranks can now be determined by

$$Rk_{av\ exact} = \sum \text{prob}(\text{rk}(x){=}Rk)*Rk \quad (13)$$

One can see that only in one case, namely ATR is the US-model appropriate, whereas in all other three cases the averaged rank tends to be located near the ends of the allowed ranking intervals. In Figure 10 two chemicals CNB and CHL, respectively, are closely considered, and the ranks calculated exactly and estimated according to the US- and

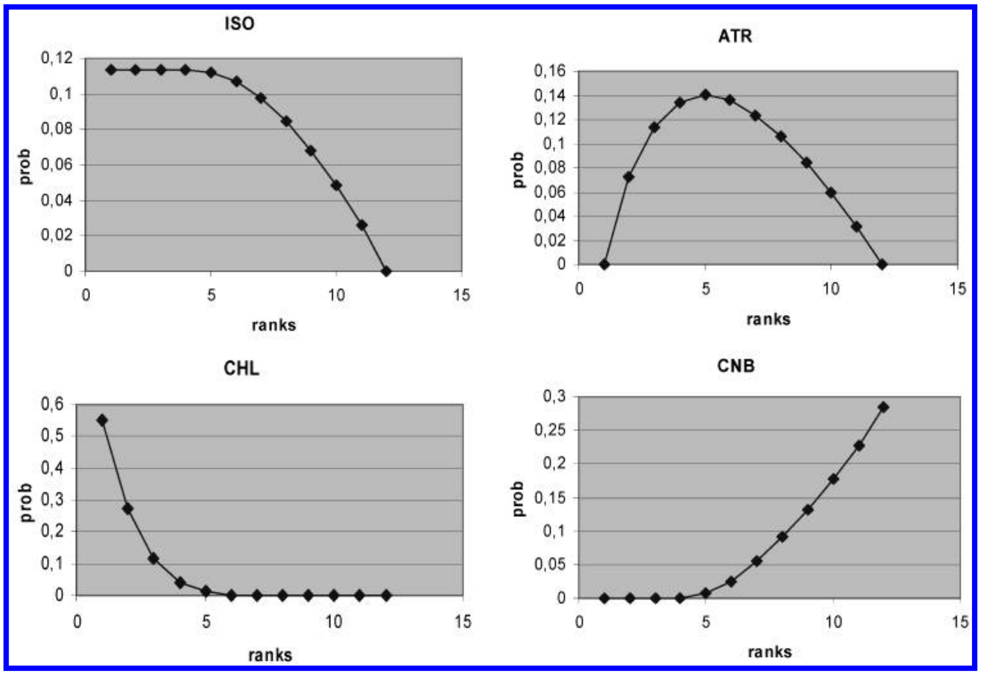


Figure 9. Ranking probabilities as a function of ranks. Note the different shapes of the functions.

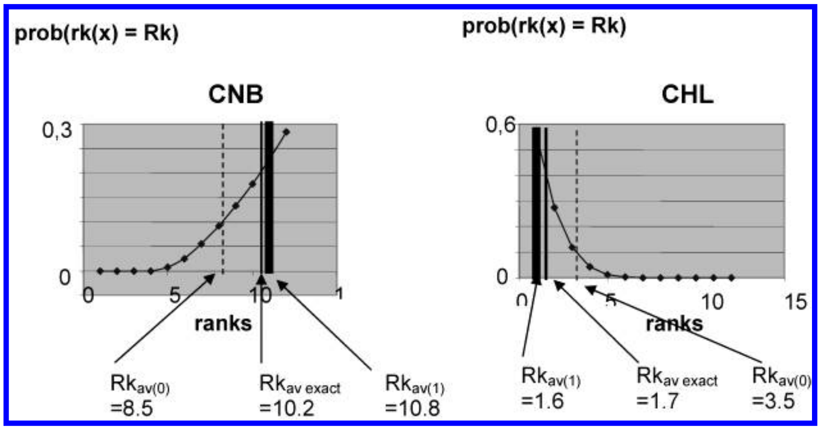


Figure 10. Ranking probabilities and averaged ranks (calculated directly by the software-program WHASSE (ref 23)), by US- and by USN-model).

USN-model, respectively, are additionally drawn as vertical lines. As one can see, within the example given here, the approximation by the USN-model works quite well.

DISCUSSION

By the two approximations, the US- and the USN-models one can avoid the time-consuming direct calculation of averaged ranks. This point is important from a practical, computational point of view, if the ground set G contains more than 25 objects. However, the insight, of how the structure of a Hasse diagram influences the outcomes of the averaged ranking, is of high interest too. It was demonstrated by the chosen example that the USN-model suffices to calculate the ranking probabilities and the averaged rank. This fact also means that the main influence is expressed by the three quantities: U, S, and P. Obviously details of the structure of a Hasse diagram do not influence the characteristics of the GRM heavily.

Besides the approximation, which is implied by supplying HD_{tot} by N local Hasse diagrams, HD_{loc}, the derivation of eq 4 seems to be very crude: Thus we come back to question 4, which we have stated in the former sections of this text: In a correct manner one should select k objects out of U objects and locate them below and the remaining U−k objects above x. The quantity k may vary in the range from 0 to U. Following eq 8, i.e., counting the linear extensions, and applying the same technique as shown in more detail, when eq 9 was derived, one can derive an exact expression for the USN-model. The result is shown in eq 14.

$$\mathrm{Rkav1a} = \frac{\sum_{k=0}^{U}\left(\frac{U!}{(U-k)!\cdot k!}\right)\cdot\frac{(S+k)!}{S!}\cdot\frac{((U+P)-k)!}{(P)!}\cdot(S+1+k)}{\sum_{k=0}^{U}\left(\frac{U!}{(U-k)!\cdot k!}\right)\cdot\frac{(S+k)!}{S!}\cdot\frac{((U+P)-k)!}{(P)!}} \qquad (14)$$

The summation index k counts the incomparable objects which are located below the specified object x, i.e., k = 0,...U. By this it is taken into account that the U incomparable objects are not at once located somewhere within the **S-x-**

Table 2. Comparison of Three Expressions for the USN-Model

equation	description
4	U objects are en bloc put into the different positions of the chain **S-x-P**
14	U single objects are put into the different positions of the chain **S-x-P** and the resulting orders are counted
15	U single objects are put into the different positions of the chain **S-x-P** and the resulting orders are **not** considered

P-chain but that there is a partitioning among the U objects. Applying the formalism of using eq 8 means that not only the U objects are distributed on several locations but also that it is important in which order this is done. This, however, is not important, if only the averaged rank is to be calculated.

Therefore one can take care for the distribution of U single objects over the chain **S-x-P** but disregard the resulting orders (which remembers somewhat on procedures in statistical thermodynamics[24]). In that case another formula can be found. It is—similar to eq 4—just a weighted sum, where now, however, not only two ranks are possible but also all ranks between S + 1 and S + 1 + U. When k objects are located within the successors of x and U−k within the predecessors, then the rank is S + 1 + k and is realized $(U!/[(U-k)!*k!])*(S+1)^{k}*(P+1)^{U-k}$ times.

$$\mathrm{Rkav1b} = \frac{\sum_{k=0}^{U} (S+1+k)\cdot\binom{U}{k}\cdot(S+1)^{k}\cdot(P+1)^{U-k}}{\sum_{k=0}^{U}\binom{U}{k}\cdot(S+1)^{k}\cdot(P+1)^{U-k}} \qquad (15)$$

Equation 15 is the adequate expression for the calculation of averaged ranks within the LPOM-concept.

The equivalence of eq 14 either to (15) or directly to (4) could up to now not be established, albeit no example could be found, where a numerical difference within these three expressions appears (testing with the software product Mathcad PLUS 6.0). Clearly it is planned to complete the formalism by proving the equivalences using the principle of induction over U or any other combinatorial proof technique. Note that the denominator of eq 15 is equal to $(S+P+2)^{U}$ due to the binomial theorem.

Summarizing: There are three equations, by which the averaged rank of an USN-model can be calculated. At least numerically the three eqs 4, 14, and 15 seem to be equivalent. Obviously, the estimation of the averaged ranks is not dependent on how the U objects are distributed within the predecessors and within the successors, respectively. This

Table 3. List of Abbreviations (Alphabetically Sorted)

abbreviation	explanation	remark
GRM	General Ranking Model	theoretical concept how to find and to characterize linear orders derived from empirical posets
HD	Hasse Diagram	
HDT	Hasse Diagram Technique	The name Hasse has its origin from the German mathematician H. Hasse, who made this kind of directed acyclic graphs popular.
HPVC	High Production Volume Chemicals	
LPOM	Local Partial Order Model	Here the most simple one is shown and discussed. Forthcoming papers will study more complex LPOMs.
posets	Partially Ordered SETS	
POT	Partial Order Theory	In comparison to HDT the concept POT pronounces the theoretical aspects. Often HDT and POT are used as synonyms.
US-model	a concept to estimate the averaged rank of an object x just by knowing the number of objects, incomparable to x (U) and the number of successors of x (S).	
USN-model	an improved concept to estimate the averaged rank of an object x by additionally taking into account the number of predecessors of x	As U, S, and P are related to N, the final equation contains U, S, and N.

Table 4. List of Symbols

symbol	explanation	remark
delta0, delta1	deviations of exact values from approximated ones	
eP	number of linear extensions	A crude upper bound is N!
G, G'_{i}	ground sets of the posets	
N, S, P, U	number of equivalence classes, number of successors-, of predecessors of x, number of objects incomparable with x	characterizing numbers of a (local) Hasse diagram
O(x), F(x)	order ideal, order filter of x	See Figure 2 for examples.
prob(rk(x)=Rk)	probability that the rank rk(x) gets the value Rk	
q_{i}	ith attribute of an object.	The starting point for HDT is to consider a specific order relation, namely the product order or component wise order.
rk(x)	rank of object x considered as random variable	
Rk_{av}, $Rk_{av(...)}$	averaged rank, different approximations	The concept LPOM can be easily extended; however, each refinement implies a manifold of different cases.
Rkav1a, Rkav1b	two expressions of averaged ranks based on more sophisticated assumptions	eqs 14 and 15
Rk_{min}, Rk_{max}	lower and upper bound of the interval of accessible ranks	Define a "ranking window".
S−x-P	a chain built from the successors of x, the object x and the predecessors of x	See for an example Figure 2.
t,s	parameters of the regression equation	equation 7
x || y	object x is incomparable with y	

fact may be interpreted as an out-averaging of the k, U−k objects located below and above the one interesting object x. The main features are summarized in Table 2.

Could the LPOM-concept be improved? Obviously there will be an influence by the different relation of objects incomparable with the one specified. This point was already discussed in Figure 2. Therefore it seems promising to develop further the formalism of the LPOM-concept by introducing a classification about the incomparable objects.

Taking the practical applications of the USN-model into account, it is possible to get a linear rank, without introducing further subjective preferences by—for example—weighting the attributes, i.e., calculating a positive monotonuous function with respect to the attributes, see ref 17. Formally the weights are here supplied by the positions of any object of interest within the Hasse diagram. Therefore the final result can be manipulated by the number of successors and predecessors, which in turn depends on the preprocessing of the data matrix. One of the next topics which are of crucial importance within the GRM is to derive some standards in the preprocessing of data.

ACKNOWLEDGMENT

The first author (R.B.) thanks the National Environmental Research Institute for supporting a visiting time (October and November, 2002) as guest scientist.

REFERENCES AND NOTES

(1) Randic, M. On Comparability of Structures. *Chem. Phys. Lett.* **1978**, *55*, 547−551.

(2) Randic, M.; Wilkins, C. L. Graph theoretical ordering of structures as a basis for systematic searches for regularities in molecular data. *J. Phys. Chem.* **1979**, *83*, 1525−1540.

(3) Randic, M. In Search of Structural Invariants. *J. Math. Chem.* **1992**, *9*, 97−146.

(4) Klein, D. J. Chemical Graph-theoretic Cluster Expansions. *Int. J. Quantum Chem.* **1986**, *20*, 153−171.

(5) Klein, D. J.; Babic, D. Partial orderings in Chemistry. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 656−671.

(6) Klein, D. J. Similarity and dissimilarity in posets. *J. Math. Chem.* **1995**, *18*, 321−348.

(7) Halfon, E. Is there a best model structure? II. Comparing the Model Structures of Different Fate Models. *Ecol. Mod.* **1983**, *20*, 153−163.

(8) Halfon, E.; Reggiani, M. G. On Ranking Chemicals for Environmental Hazard. *Environ. Sci. Technol.* **1986**, *20*, 1173−1179.

(9) Halfon, E. Comparison of an Index Function and a Vectorial Approach Method for Ranking Waste Disposal Sites. *Environ. Sci. Technol.* **1989**, *23*, 600−609.

(10) Halfon, E.; Brüggemann, R. Environmental Hazard Ranking of Chemicals Spilled in the Rhine River in November 1986. *Acta Hydrochim. Hydrobiol.* **1989**, *17*, 47−60.

(11) Klein, D. J.; Brickmann, J. Partial Orderings in Chemistry. *MATCH (Comm. Math. Comp. Chem.)* **2000**, *42*, 1−290. See especially: Klein, D. J. Prolegomenon on Partial Orderings in Chemistry. *MATCH* **2000**, *42*, 7−21.

(12) Lerche, D.; Sørensen, P. B.; Brüggemann, R. Improved Estimation of the Ranking Probabilities in Partial Orders using Random Linear Extensions by Approximation of the Mutual Ranking Probability. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1471−1480.

(13) Winkler, P. Average height in a partially ordered set. *Discr. Math.* **1982**, *39*, 337−341.

(14) Brüggemann, R.; Bartel, H.-G. A Theoretical Concept to Rank Environmentally Significant Chemicals. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 211−217.

(15) Luther, B.; Brüggemann, R.; Pudenz, S. An Approach to Combine Cluster Analysis with Order Theoretical Tools in Problems of Environmental Pollution. *MATCH* **2000**, *42*, 119−143.

(16) Helm, D. Bewertung von Monitoringdaten der Umweltprobenbank des Bundes mit der Hasse-Diagramm-Technik. *UWSF − Z. Umweltchem. Ökotox.* **2003**, *15*, 85−94.

(17) Brüggemann, R.; Halfon, E.; Welzl, G.; Voigt, K.; Steinberg, C. Applying the Concept of Partially Ordered Sets on the Ranking of Near-Shore Sediments by a Battery of Tests. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 918−925.

(18) Davey, B. A.; Priestley, H. A. *Introduction to Lattices and Order*; Cambridge University Press: Cambridge, 1990; pp 1−248.

(19) Trotter, W. T. *Combinatorics and Partially Ordered Sets Dimension Theory*; The Johns Hopkins University Press: Baltimore, MD, 1992; pp 1−307.

(20) Stanley, R. P. Two Combinatorial Applications of the Aleksandrov-Fenchel Inequalities. *J. Combin. Theory (A)* **1981**, *31*, 56−65.

(21) Stanley, R. P. *Enumerative Combinatorics Volume I*; Wadsworth&Brooks/Cole: Monterey, 1986; pp 1−306.

(22) Lerche, D.; Brüggemann, R.; Sørensen, P. B.; Carlsen, L.; Nielsen, O. J. A Comparison of Partial Order Technique with three Methods of Multicriteria Analysis for Ranking of Chemical Substances. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1086−1098.

(23) Brüggemann, R.; Bücherl, C.; Pudenz, S.; Steinberg, C. Application of the concept of Partial Order on Comparative Evaluation of Environmental Chemicals. *Acta Hydrochim. Hydrobiol.* **1999**, *27*, 170−178.

(24) Kauzmann, W. *Thermodynamics and Statistics*; W. A. Benjamin, Inc.: New York, 1967; pp 1−321.

CI034214M