Improved Estimation of the Ranking Probabilities in Partial Orders Using Random Linear Extensions by Approximation of the Mutual Ranking Probability

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

Institute of Chemistry, H.C. Ørsted Institute, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark, Department for Policy Analysis, National Environmental Research Institute, P.O. Box 358, DK-4000 Roskilde, Denmark, and Institute of Freshwater Ecology and Inland Fisheries, Mueggelseedamm 310, D-12587 Berlin-Friedrichshagen, Germany

Received January 27, 2003

The application of partial order theory and Hasse diagram technique in environmental science is getting increasing attention. One of the latest developments in the field of Hasse diagram technique is the use of random linear extensions to estimate ranking probabilities. In the original algorithm for estimating the ranking probability it is assumed that the order between two incomparable pair of objects can be chosen randomly. However, if the total set of linear extensions is considered there is a specific probability that one object will be larger than another, which can be far from 50%. In this study it is investigated if an approximation of the mutual ranking probability can improve the algorithm. Applying an approximation of the mutual ranking probability the estimation of the ranking probabilities are significantly improved. Using a test set of 39 partial orders with randomly chosen values the relative mean root square difference (MRSD) decrease in average from 7.9% to 2.2% and a maximum relative improvement of 90% can be found. In the most successful case the relative MRSD goes as low as 0.77%.

INTRODUCTION

The application of partial order theory and Hasse diagram technique in environmental science, including research on environmental policy for chemical substances, has received increasing attention. E.g. since 1998 annual workshops[1−4] about applications in environmental science are being held.[1] One of the latest developments in the field of Hasse diagram technique is the use of random linear extensions and the corresponding average rank.[5−7] This application is especially useful when the partial ordering is insufficient and a linear rank or a total order is needed in the decision making process. The application of the linear extensions methodology is advantageous in the sense that no weighting or mathematical correlations among the descriptors are necessary.[8] This increases the transparency of the ranking exercise relative to other methods, such as the utility function, incl. e.g. conjoint analysis. The random linear extensions methodology has already been applied in several cases.[9−11] In this study the original determination of the ranking probabilities from the random linear extensions has been significantly improved by the incorporation of an estimation of the mutual order for incomparable objects.

To illustrate how this new method can be applied to, for example, ranking chemical substances in environmental policy, a data set on persistent organic chemicals (POPs) is applied. The data set on POPs is taken from a former study aiming to identify candidates for the POP-Protocol of the United Nations Economic Commission for Europe (UN-ECE).[12]

PARTIAL ORDERS AND LINEAR EXTENSIONS

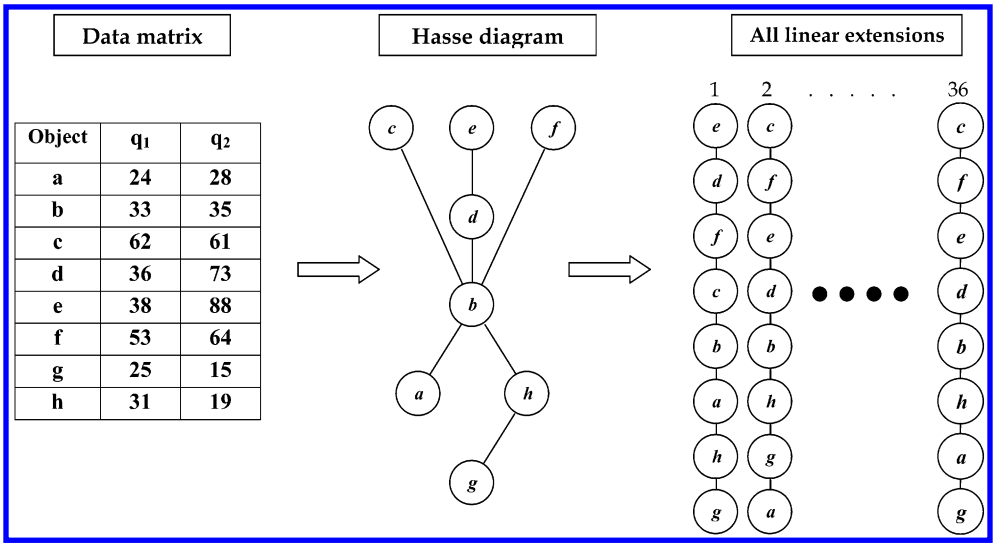
Partial Order Theory is a large mathematical discipline and several textbooks on the topic are available e.g. refs 13−17.

The principle of partial orders and linear extensions are explained in Figure 1. In a partially ordered set different descriptors, q_1, q_2, ..., q_i are used simultaneously as ranking parameters for a series of objects, **a**, **b**, **c**, ... The objects and the corresponding descriptors are referred to as the data matrix or the information base.[18]

The comparison of all descriptors for all objects leads to the partial order. If a given object, **a**, is characterized by the descriptors $q_1(\mathbf{a})$, $q_2(\mathbf{a})$, ..., $q_i(\mathbf{a})$ and another object, **b**, is characterized by the descriptors $q_1(\mathbf{b})$, $q_2(\mathbf{b})$, ..., $q_i(\mathbf{b})$, then **a** and **b** can be compared through comparison of the individual descriptors ($q_1(\mathbf{a})$ with $q_1(\mathbf{b})$ etc.). Object **a** is then ranked higher than **b** (**a** > **b**) if all descriptors for **a** are higher than or equal to the corresponding descriptors for **b** ($q_i(\mathbf{a}) \geq q_i(\mathbf{b})$ for all i). However, in the special case that all descriptors for **a** are equal to the corresponding descriptors for **b**, $q_i(\mathbf{a}) = q_i(\mathbf{b})$ for all i, the two objects will have identical rank and will be considered as equivalent, **a** ∼ **b**. In Figure 1 object **a** is ranked lower than **b** because both descriptors for **b** (33, 35) are higher than the descriptors for **a** (24, 28). In the case that two objects cannot be given a mutual order relative to each other they are said to be incomparable. These two objects are thus not ordered with respect to each other.

* Corresponding author phone: +45 46 30 18 51; e-mail: dbl@dmu.dk.
† University of Copenhagen.
‡ National Environmental Research Institute.
§ Institute of Freshwater Ecology and Inland Fisheries.

10.1021/ci0300036 CCC: $25.00
Published on Web 08/23/2003



Object	q_1	q_2
a	24	28
b	33	35
c	62	61
d	36	73
e	38	88
f	53	64
g	25	15
h	31	19

Figure 1. The data matrix, the Hasse diagram, and all linear extensions for 8 objects with two descriptors with randomly chosen numbers (sample no. 3 in Table 7).

Table 1. Ranking Probabilities for the Partial Order in Figure 1[a]

	rank							
object	8	7	6	5	4	3	2	1
a	0	0	0	0	0	0.333	0.333	0.333
b	0	0	0	0	1	0	0	0
c	0.25	0.25	0.25	0.25	0	0	0	0
d	0	0.167	0.333	0.5	0	0	0	0
e	0.5	0.333	0.167	0	0	0	0	0
f	0.25	0.25	0.25	0.25	0	0	0	0
g	0	0	0	0	0	0	0.333	0.667
h	0	0	0	0	0	0.667	0.333	0

[a] Sample no. 3 in Table 7.

Often several objects are incomparable and this is why the total set of objects is said to be partially ordered. In Figure 1 the objects **a** (24, 28) and **h** (31, 19) are incomparable because 24 < 31 and 28 > 19 making it impossible to order **a** with respect to **h**. As the name "incomparable" is given to the objects that cannot be given a mutual order, the name "comparable" is given to the objects for which a mutual order can be given.

The partially ordered set can be visualized as a directed acyclic graph, where each object is represented by a circle. If two objects are comparable, a line is drawn between the corresponding circles. The higher ranked of the two is given a point with a higher vertical position. Since such a graph often will contain lines for each pair of objects it might contain trivial relations (e.g. if $\mathbf{a} \leq \mathbf{b}$ and $\mathbf{b} \leq \mathbf{c}$ then $\mathbf{a} \leq \mathbf{c}$). The graph can thus be greatly simplified by only drawing next neighbor connections. A directed acyclic graph of cover-relations is thus constructed. This kind of graphical representation of ordered sets is referred to as a Hasse diagram.[19] The principle of the Hasse diagram is illustrated in Figure 1. The Hasse diagram is useful for a series of purposes. It facilitates the identification of levels and isolated groups of objects, so-called hierarchies. Based on the Hasse diagram the correlation among and the influence on the ordering from the individual descriptors can also be recognized e.g. ref 20. However, in this paper the role of the Hasse diagram is just to display the order relations.

A linear extension is a projection of the partial order into a total order that complies with all the relations in the partial order. It is an order preserving mapping of the partial order. The linear extension or the linear order can be considered as a specific outcome of the partial order. In each of the linear extensions the incomparable objects are thus given an order. E.g. if a linear extension is made from the Hasse diagram in Figure 1, **b** must be larger than **a**, (**b** > **a**), in the linear extensions since **b** is larger than **a** in the Hasse diagram. On the other hand, **a** can be both larger than or smaller than **h**, (**a** > **h**) or (**a** < **h**) because these objects are incomparable in the Hasse diagram. Given a partial order it is mostly possible to find several linear extensions. For the partial order in Figure 1 it is possible to find 36 linear extensions. The total set of linear extensions for a partial order can be found by a combinatorial exercise, where the incomparable objects systematically are given an order with respect to each other.

Table 2. Objects above and below **g** and **h** in the Partial Order in Figure 2[a]

objects > g	objects < g	objects > h	objects < h
a		a	j
k		d	
c			
d			
e			
i			
b			

[a] Sample no. 25 in Table 7.

Table 3. Possible Ranking Position for Object **g** and **h** in the Partial Order in Figure 2

rank	**g**	**h**
12	no	no
11	no	no
10	no	yes
9	no	yes
8	no	yes
7	no	yes
6	no	yes
5	yes	yes
4	yes	yes
3	yes	yes
2	yes	yes
1	yes	no

Table 4. First 18 Possible Combinations of Ranking Positions for the Objects **g** and **h** in the Partial Order in Figure 2

ranking position of **g**	mutual order	ranking position of **h**
5	<	10
5	<	9
5	<	8
5	<	7
5	<	6
5	>	4
5	>	3
5	>	2
4	<	10
4	<	9
4	<	8
4	<	7
4	<	6
4	<	5
4	>	3
4	>	2
3	<	10
3	<	9
.	.	.
.	.	.
.	.	.

If all linear extensions are identified for a partial order it is possible to count the frequencies, n_{jk}, of the occurrence of a specific object, j, on a specific rank, k. By dividing the frequencies, n_{jk}, by the total number of linear extensions the ranking probabilities, $p_r(j,k)$, are obtained. The ranking probabilities for the partial order in Figure 1 are calculated in Table 1. Additionally, from the linear extensions the mutual ranking probabilities, p_{mr}, for two incomparable objects can be found in terms of the probability for one object to be above or below another object (Table 5). Finally, considering ranking probabilities for all objects on all ranking positions the average rank for an object can be calculated.[21,22] The average rank, the minimum and maximum rank, and the ranking probabilities p_r as well as the mutual ranking probabilities p_{mr} are outcomes of the computer-program WHASSE, whose basic properties are described in ref 18.

RANDOM LINEAR EXTENSIONS

It is possible to apply computer programs, like WHASSE, which can identify all possible linear extensions in partial orders.[18] However, the actual number of possible linear extensions is closely related to the characteristics of the partial order. Most important is the number of incomparable pairs of objects. The number of linear extensions increases dramatically as the number of incomparable pairs increases. In the case that all objects are incomparable (the upper limit) the number of linear extensions depends on the faculty of the number of objects.

For a systematic deviation of the number of linear extensions for a partial order a recurrence relation can be used, which is discussed for example by Atkinson and Chang 1986[23] and Atkinson 1989[24] (See also ref 25 for a generalization). In addition Stanley, 1986,[26] gives a useful formula. However this formula refers only to a disjoint sum of partial orders (for details see ref 20). Further, when the number of objects approach infinity almost all partial orders have 3 asymptotic layers. From this fact they deduce an estimation of the number of linear extensions to be $(n/2)!\ [(n/4)!]^2$.

In Table 7 39 partial orders with randomly chosen values are given. The number of objects varies from 8 to 14. For the partial orders with 8 objects the number of linear extensions vary from 30 to 10080. However, for the partial orders with 14 objects the number of linear extensions goes up and varies from 20513 to 290334240. This dramatic increase in the number of linear extensions makes it practically impossible to identify all linear extensions for data sets with more than 15−20 objects (where approxi-

Table 5. Probability that the Objects in the Column to the Left Are Ranked Higher than the Objects in the Row to the Right[a]

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**
a	X	1	0.894	1	1	1	1	1	1	1	0.568	0.989
b		X	0	0.078	0	0.742	1	0.355	0.184	0.719	0	0.556
c			X	0.452	0.609	1	1	0.778	0.620	0.950	0	1
d				X	0.645	1	1	1	0.663	1	0.189	0.896
e					X	1	1	0.717	0.537	0.929	0	0.854
f						X	0.629	0.176	0	0.513	0	0.357
g							X	0.106	0	0.401	0	0.258
h								X	0.325	1	0.044	0.671
i									X	0.901	0.110	0.800
j										X	0.006	0.361
k											X	1
l												X

[a] For the partial order in Figure 2.

Table 6. Probability in Table 5 Estimated Using the Algorithm Developed in This Study

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**
a	X	1	0.947	1	1	1	1	1	1	1	0.800	1.000
b		X	0	0.133	0	0.730	1	0.375	0.263	0.500	0	0.444
c			X	0.500	0.500	1	1	0.737	0.625	0.828	0	1
d				X	0.651	1	1	1	0.625	1	0.270	0.765
e					X	1	1	0.625	0.500	0.737	0	0.650
f						X	0.600	0.133	0	0.313	0	0.278
g							X	0.146	0	0.313	0	0.286
h								X	0.500	1	0.136	0.616
i									X	0.716	0.208	0.659
j										X	0.059	0.444
k											X	1
l												X

Table 7. A Test Set of 39 Partial Orders and Parameters Describing the Partial Orders

sample	objects	descriptors	total comparisons	incomparable	total set of linear extensions	potentially needed number of random linear extensions[7]
1	8	2	28	14	205	14791
2	8	2	28	14	232	31386
3	8	2	28	7	36	6494
4	8	2	28	12	120	14430
5	8	2	28	7	30	5411
6	8	3	28	17	568	25614
7	8	3	28	26	10080	30304
8	8	3	28	18	1728	25974
9	8	3	28	23	4032	60608
10	8	3	28	24	6048	45455
11	10	2	45	23	4452	33460
12	10	2	45	31	12320	111117
13	10	2	45	26	4536	56813
14	10	2	45	18	1476	44374
15	10	2	45	22	1206	43507
16	10	3	45	29	24192	36364
17	10	3	45	39	345600	129879
18	10	3	45	37	107280	161269
19	10	3	45	37	138000	69146
20	10	3	45	33	25965	83638
21	12	2	66	38	84672	636260
22	12	2	66	27	4014	85179
23	12	2	66	35	81504	131265
24	12	2	66	39	41976	420629
25	12	2	66	37	38145	382295
26	12	3	66	59	18247680	653155
27	12	3	66	55	5651712	674317
28	12	3	66	60	16632000	238073
29	12	3	66	57	8553600	428625
30	12	3	66	56	5987520	642683
31	14	2	91	57	1487978	628136
32	14	2	91	37	20513	217675
33	14	2	91	39	193680	124773
34	14	2	91	52	908872	849512
35	14	2	91	40	264616	165739
36	14	3	91	66	58644768	346551
37	14	3	91	69	136509120	1466502
38	14	3	91	74	290334240	288608
39	14	3	91	76	278001360	1105494

mately two-thirds of the comparisons are incomparable). Even for very fast computers the assignment is too demanding. If, for example, it takes 1 s to find all linear extensions for 15 objects it will take approximately 350.000 years to analyze 25 objects. This is probably why the linear extension methodology has not been widely applied to establish ranking probabilities. However, since the partial order, the linear extensions, the ranking probabilities, and the average rank can be considered as a generalized model, which comprises all positive monotonic functions based on the set of descriptors, it is important to try to perform an estimation.

This was done recently when a method was suggested, which estimates the ranking probabilities and the average rank for large partial orders.[5−7] The average rank is then predicted based on a minor fraction of randomly chosen linear extensions. This extends the manageable number of objects from around 20 to 1000.

In the original algorithm when a random linear extension is created an incomparable pair is randomly selected and given a random rank. Referring to Figure 1, one of the seven incomparable pairs, either (**a**,**h**), (**a**,**g**), (**c**,**d**), (**c**,**e**), (**c**,**f**), (**d**,**f**), or (**e**,**f**) is thus randomly selected, and a random mutual order is chosen. Subsequently the remaining incomparable pairs are checked if they are still incomparable. Then one of the remaining incomparable pair is randomly chosen and a mutual order for the two objects is selected etc. until all incomparable pairs are given an order. Considering the objects in Figure 1, the first randomly selected pair of objects might be (**a**,**h**) with object **a** randomly chosen to be higher than the object **h**. In this case **a** will automatically become higher than **g** and the incomparable pairs, (**c**,**d**), (**c**,**e**), (**c**,**f**), (**d**,**f**), and (**e**,**f**), remain. The next step in this example could be to make a random decision regarding the ordering of for example **c** and **d** in relation to each other. The result of this could be to select **d** to be higher than **c**. Now the following incomparable pairs of objects remains, (**c**,**f**), (**d**,**f**), and (**e**,**f**) since **c** and **e** are automatically given an order. If finally **f** is placed between **d** and **c** the linear extension correspond to the first linear extension in Figure 1. A detailed evaluation of the methodology can be found in ref 7.

APPROXIMATION OF THE MUTUAL RANKING PROBABILITY

In the original algorithm developed to find the random linear extensions for partial orders it is assumed that the order between two incomparable pair of objects can be chosen randomly.[5] However as illustrated in Figure 2 this is a rather crude approximation. From the total set of linear extensions a specific probability that one object will be larger than another can be calculated. If for example a partial order has 36 linear extensions and one object is larger than the other

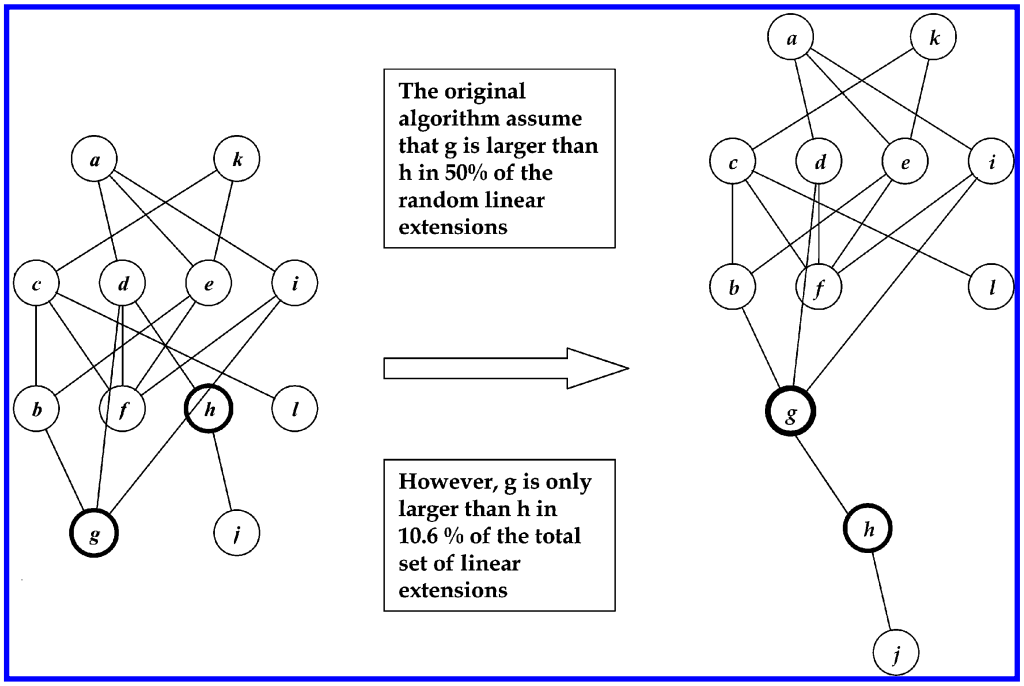


Figure 2. Selecting a mutual order for the incomparable objects **g** and **h** (sample no. 25 in Table 7).

object in 12 of the linear extensions and smaller than the other object in the remaining 24 linear extensions, then the probability that the first object will be larger than the other object is 0.333. This is referred to as the mutual ranking probability, p_{mr} (not to be confused with the ranking probability for a single object on a specific rank $p_r(j,k)$).

In this study we suggest that an approximation of the mutual ranking probability replace the random choice. Note that it is impossible to calculate and use the true mutual ranking probability since all linear extensions are needed in order to do so. In this manner an attempt to improve the estimation of the generalized model is performed.

In ref 27, the estimation of mutual ranking probabilities is thoroughly investigated, and several approaches on how to estimate the mutual ranking probability are provided including the algorithm developed in this study. Further, the structural influence on the estimation of the mutual ranking probabilities is evaluated.[27]

In this study an approximation of the mutual ranking probability is derived by considering the number of objects above and the number of objects below two incomparable objects. E.g. in Figure 2 object **g** and **h** are incomparable. The objects above and below **g** and **h** are shown in Table 2. Now, the maximum ranking position for each of the two objects is found as the total number of objects minus the number of objects above. For **g** the maximum ranking position is 5 (the total number of objects is 12 minus the number of objects above, which is 7). The minimum ranking position is found as the number of objects below plus one. Applying the maximum and minimum positions the possible ranking positions in a random linear extension can be found. For **g** and **h** in the partial order in Figure 2 the possible ranking positions are indicated in Table 3. There are 5 possible ranking positions for **g** and 9 for **h.** From Table 3 possible mutual orders can be extracted. In total there are 41 possible combinations since the situation rank(**g**) = rank(**h**) is excluded. Table 4 shows the first 18 combinations

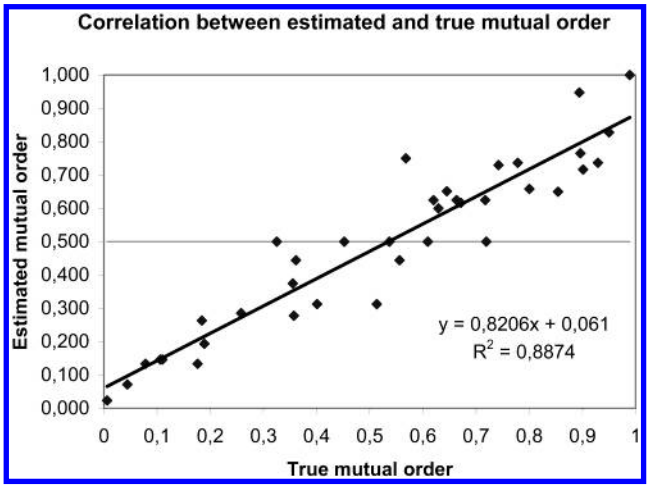


Figure 3. The correlation between the true mutual ranking probabilities and the estimated mutual ranking probabilities for the partial order presented in Figure 2.

and their mutual order. The result **g** > **h** is obtained in 6 cases, whereas **h** > **g** is obtained in 35 cases. This gives an estimated mutual rank for **g** > **h** of 14.6%.

In the same manner the mutual ranking probabilities can be estimated for all the incomparable objects. The result is shown in Table 6. In Table 5 the true mutual ranking probability is calculated using all the linear extensions. The true mutual ranking probability for **g** and **h** is found to be 10.6% for **g** > **h**, which is also indicated in Figure 2. It can thus be seen that 14.6% is more close to the real value of 10.6% than the original assumption of 50%. However, even though that in the example of **g** and **h** the mutual ranking probability gets much closer to the true mutual ranking probability it can be seen from Tables 5 and 6 that it also sometimes slightly overestimates the mutual ranking probability (e.g. for **d** and **e** or for **f** and **h**). The correlation between the true and the estimated mutual ranking probabilities is given in Figure 3. In the original version of the

Table 8. Relative Mean Root Square Distance (MRSD) and the Coefficient of Determination, R^2, and the Slope, a, and the Intersection, b (ax + b), When the True Ranking Probabilities and the Estimated Ranking Probabilities Are Compared

	new method				original method			
sample	relative MRSD (%)	R^2	a	b	relative MRSD (%)	R^2	a	b
1	2.86	0.985	0.946	0.007	5.49	0.938	0.889	0.014
2	1.70	0.984	0.976	0.003	12.25	0.950	0.921	0.010
3	1.35	0.997	0.990	0.001	2.90	0.986	0.978	0.003
4	2.63	0.986	0.976	0.003	4.05	0.977	0.950	0.006
5	3.07	0.988	0.967	0.004	3.55	0.983	0.971	0.004
6	2.97	0.963	0.926	0.009	4.85	0.940	0.891	0.013
7	1.56	0.988	1.076	−0.010	2.53	0.932	0.802	0.025
8	2.04	0.972	0.953	0.005	2.84	0.946	0.947	0.007
9	1.49	0.982	0.956	0.006	4.29	0.870	0.811	0.024
10	1.23	0.978	0.964	0.005	2.44	0.930	0.868	0.017
11	1.49	0.960	1.009	0.001	5.14	0.96	0.932	0.007
12	3.46	0.966	1.020	0.002	7.53	0.935	0.865	0.014
13	2.03	0.993	0.989	0.001	11.69	0.961	0.937	0.006
14	1.60	0.987	1.002	0.000	6.30	0.961	0.936	0.006
15	3.08	0.950	1.004	−0.004	5.74	0.954	0.906	0.009
16	1.22	0.996	1.000	0.000	2.28	0.985	0.975	0.003
17	1.26	0.962	1.047	−0.005	2.95	0.889	0.823	0.018
18	1.33	0.988	0.994	0.001	3.79	0.929	0.842	0.016
19	1.72	0.964	1.021	−0.002	2.60	0.936	0.863	0.014
20	2.23	0.972	1.005	−0.001	8.26	0.938	0.855	0.015
21	3.91	0.974	0.962	0.003	22.93	0.882	0.84	0.013
22	2.04	0.980	1.013	−0.001	9.97	0.952	0.902	0.008
23	2.54	0.969	1.033	−0.003	8.04	0.940	0.905	0.008
24	3.67	0.987	0.989	0.001	22.46	0.928	0.838	0.014
25	3.86	0.978	0.990	0.001	23.41	0.907	0.861	0.012
26	2.51	0.980	0.942	0.005	3.10	0.914	0.798	0.017
27	2.39	0.984	0.964	0.003	9.30	0.931	0.826	0.014
28	2.43	0.987	0.911	0.007	3.94	0.939	0.794	0.017
29	2.09	0.972	0.963	0.003	4.39	0.908	0.811	0.016
30	2.33	0.989	0.956	0.004	4.94	0.903	0.815	0.016
31	3.09	0.984	0.970	0.002	14.72	0.924	0.842	0.113
32	2.10	0.985	1.015	−0.001	14.95	0.947	0.903	0.007
33	0.77	0.985	1.013	−0.001	5.16	0.941	0.923	0.006
34	2.31	0.978	1.003	0.000	22.15	0.918	0.862	0.01
35	0.99	0.981	1.024	−0.002	8.38	0.956	0.914	0.006
36	1.56	0.971	1.002	0.000	7.18	0.900	0.841	0.011
37	1.39	0.991	1.022	−0.002	7.21	0.877	0.806	0.014
38	2.03	0.972	0.970	0.002	4.57	0.949	0.869	0.009
39	3.15	0.989	0.944	0.004	9.22	0.931	0.824	0.013

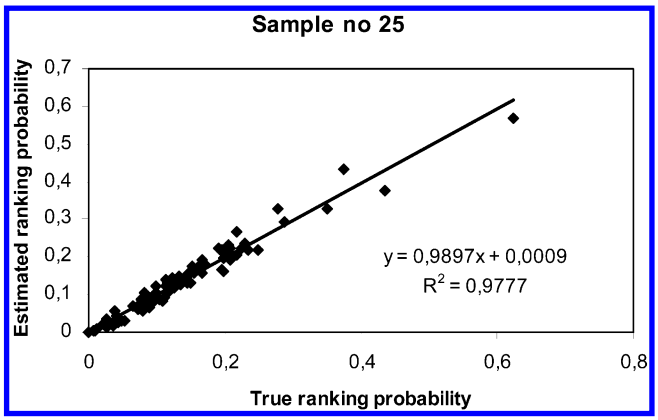


Figure 4. The correlation between the true ranking probabilities and the estimated probabilities using random linear extensions including the approximation of the mutual ranking probability for the partial order presented in Figure 2.

method the mutual order was chosen randomly ($p_{mr} = 0.5$) in all cases. This corresponds to the horizontal line in Figure 3. It can thus be seen that even though some scatter is observed the new estimation of the mutual rank gives a better correlation to the true mutual ranking probabilities than simply assuming that the mutual rank can be chosen randomly.

The algorithm finding random linear extensions and the average rank is programmed in Excel Visual Basic and can be obtained by the corresponding author.

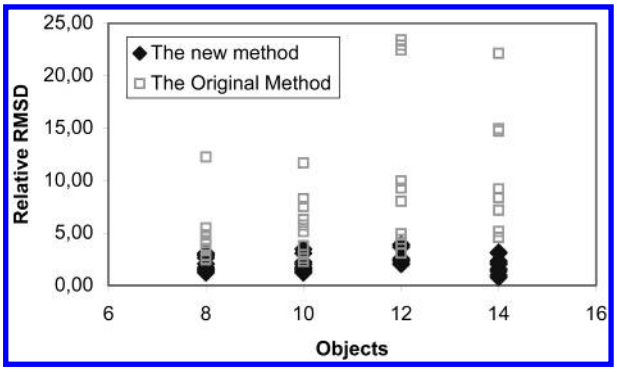


Figure 5. The new and original methods.

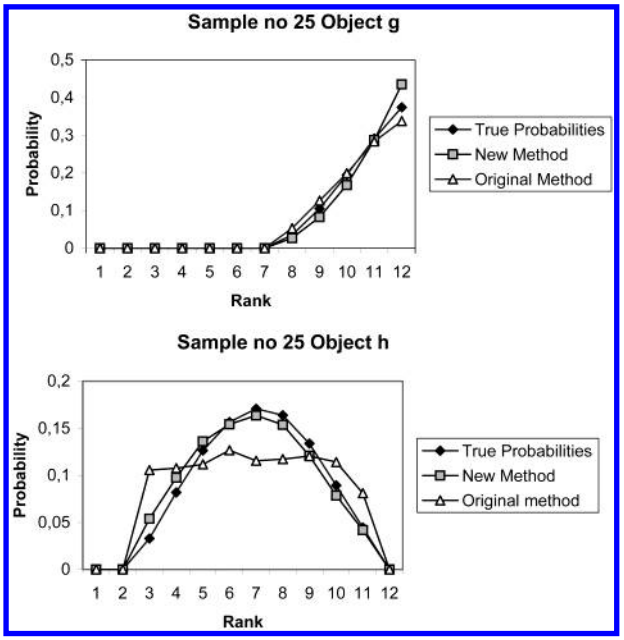


Figure 6. The true and the estimated rankling probabilities for object **h** and **g** in Figure 2.

Table 9. Top 12 POPs Found by Partial Order Theory and Random Linear Extensions[a]

rank	CAS no.	name of compound	av rank
1	*608-93-5*	*pentachlorobenzene*	*7.77*
2	*335-76-2*	*nonadecafluorodecanoic acid*	*8.16*
3	1134-04-9	2,3,4,5-tetrachloro-6-(trichloromethyl)-pyridine	8.91
4	634-66-2	tetrachloro-1,2,3,4-benzene	10.57
5	*76-21-1*	*2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9-hexadecafluorononanoic acid*	*10.62*
6	95-94-3	tetrachloro-1,2,4,5-benzene	10.72
7	*1335-87-1*	*hexachloronaphthalene*	*10.98*
8	87-68-3	1,1,2,3,4,4-hexachloro-1,3-butadiene	11.03
9	*3825-26-1*	*ammonium perfluorooctanoate*	*11.09*
10	2991-51-7	*N*-ethyl-*N*-(heptadecafluorooctyl)glycine	14.06
11	1861-40-1	benfluralin	14.92
12	5216-25-1	1-chloro-4-(trichloromethyl)benzene	15.04

[a] The italicized entries indicate the chemicals fulfilling criteria of the UN-ECE POP-Protocol.

Table 10. Chemical Substances and the POP-Descriptors

	CAS no.	name of chemical substance	Log K_{ow}	bio-deg[a]	P_{vapor} (Pa)	atm. half-life (d)	toxicity category: human	toxicity category: eco-tox
1	115-32-2	dicofol	5.0	2	0.0000016	3.1	2	3
2	77-47-4	1,2,3,4,5,5-hexachloro-1,3 cyclopentadiene	5.0	1	2.82	27	1	3
3	87-86-5	pentachlorophenol	5.1	1	0.0070	19	1	2
4	608-93-5	pentachlorobenzene	5.2	1	0.67	190	3	3
5	1134-04-9	2,3,4,5-tetrachloro-6-(trichloromethyl)-pyridine	5.3	2	0.011	3700	3	*b*
6	1335-87-1	hexachloro-naphthalene	7.0	1	0.00044	57	*b*	3
7	79-94-7	4,4′-1-methylethylidenebis 2,6-dibromo-phenol	7.2	1	0.0000000023	3.6	*b*	3
8	70-30-4	2,2′-methylenebis 3,4,6-trichlorophenol	7.5	1	0.000000014	4.9	2	3
9	297-78-9	isobenzan	5.2	2	0.0010	2.3	1	2
10	3825-26-1	ammonium perfluorooctanoate	6.3	2	990	21	2	*b*
11	335-76-2	nonadecafluoro-decanoic acid	8.2	2	770	21	2	*b*
12	76-21-1	2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9-hexafluoro-nonanoic acid	6.7	1	690	21	3	*b*
13	13356-08-6	hexakis-2-methyl-2-phenylpropyl fenbutatin oxide distannoxane	16	2	$2 \cdot 10^{-18}$	0.018	3	2
14	327-98-0	trichloronate	5.2	1	0.0060	0.15	1	3
15	3734-48-3	chlordene-isomer mixture	5.6	2	0.028	0.17	3	3
16	3389-71-7	hexachloronornornadiene	5.2	2	0.62	0.18	*b*	3
17	40487-42-1	pendimethalin	5.2	1	0.000088	0.35	3	1
18	82657-04-3	bifenthrin	8.2	1	0.0000046	0.36	2	1
19	1582-09-8	trifluralin	5.3	1	0.00021	0.45	2	3
20	26399-36-0	profluralin	5.6	1	0.000071	0.47	3	1
21	1861-40-1	benfluralin	5.3	1	0.00021	0.48	3	3
22	101-20-2	*N*-4-chlorophenyl-*N*′-3,4-dichlorophenyl-triclocarban urea	4.9	1	0.0000036	0.50	*b*	3
23	68334-67-8	1,1,2,3,4-pentachloro-4-(1-methylethoxy)-1,3-butadiene	4.4	1	5.5	0.87	3	*b*
24	115-29-7	endosulfan	3.8	2	0.000036	1.3	1	1
25	5216-25-1	1-chloro-4-(trichloromethyl)benzene	4.5	2	5.1	43	*b*	*b*
26	67-72-1	hexachloroethane	4.1	1	430	>100	3	3
27	4901-51-3	2,3,4,5-tetrachlorophenol	4.2	1	0.045	6.9	2	3
28	58-90-2	2,2′-methylenebis 3,4,6-tetrachlorophenol	4.5	1	0.045	39	2	3
29	25167-83-3	tetrachloro-technical phenol	4.5	1	0.045	39	2	3
30	427-45-2	fluorotris-*p*-chlorophenylstannane	4.5	1	0.67	>100	*b*	3
31	634-90-2	1,2,3,5-tetrachlorobenzene	4.6	1	4.9	54	3	3
32	634-66-2	1,2,3,4-tetrachlorobenzene	4.6	1	4.9	>100	3	3
33	95-94-3	1,2,4,5-tetrachlorobenzene	4.6	1	4.9	>100	3	3
34	87-68-3	1,1,2,3,4,4-hexachloro-1,3-butadiene	4.8	1	32	>100	2	3
35	1937-37-7	C.I. Direct Black 38	4.9	2	$4.2 \cdot 10^{-34}$	0.053	1	*b*
36	79538-32-2	tefluthrin	6.5	2	0.0097	0.63	1	*b*
37	5902-76-1	methyl(pentachlorophenolato)mercuri	4.6	1	0.00063	1.3	1	*b*
38	3447-87-8	1,4,7-methanoisosbenzofuran-5(3h)-one	4.3	2	0.0020	1.5	1	*b*
39	7682-32-8	benzimidazole-4,5,6-tribromo-2-trifluoro-methyl	4.1	1	0.0000054	22	1	*b*
40	10331-57-4	5,5′-dichloro-3,3′-dinitro-1,1′-biphenyl-2,2′-diol	4.9	1	0.00000000061	25	1	*b*
41	2338-30-9	4,5,6,7-tetrabromo 2-trifluoromethyl-benzimidazole	4.8	1	0.00000063	52	1	*b*
42	2991-51-7	*N*-ethyl-*N*-(heptadecafluorooctyl)glycine	8.0	2	0.0043	0.67	3	*b*

[a] 1 = months and 2 = more than months. [b] Data missing.

RESULTS

One way to evaluate the estimation of the ranking probability, $p_r(j,k)$, is by the use of mean root square difference (MRSD) and the relative MRSD. The MRSD is defined as

$$\mathrm{MRSD} = \sqrt{\frac{\sum (P_{\mathrm{r,true}} - P_{\mathrm{r,estimated}})^2}{N}} \qquad (1)$$

where N is the number of linear extensions and the relative MRSD defined as

$$\text{relative MRSD} = \sqrt{\frac{\sum (P_{\mathrm{r,true}} - P_{\mathrm{r,estimated}})^2 / P_{\mathrm{r,true}}}{N}} \qquad (2)$$

In Table 7 39 partial orders with randomly chosen values are applied. The number of objects varies from 8 to 14, and the number of descriptors is either two or three. Only 39 partial orders have been applied in this evaluation since finding all linear extensions for partial orders is very time-consuming, as explained earlier. Table 7 gives information on their characteristics such as the number of total comparisons, the number of incomparable object pairs, the number of linear extensions, and the number of linear extensions

Table 11. Rank of the Chemicals in Table 10 Applying the Ranking Method Developed in this Study

rank	CAS no.	name of compound	av rank
1	608-93-5	pentachlorobenzene	7.77
2	335-76-2	nonadecafluorodecanoic acid	8.16
3	1134-04-9	2,3,4,5-tetrachloro-6-(trichloromethyl)pyridine	8.91
4	634-66-2	1,2,3,4-tetrachlorobenzene	10.57
5	76-21-1	2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9-hexadecafluorononanoic acid	10.62
6	95-94-3	1,2,4,5-tetrachlorobenzene	10.72
7	1335-87-1	hexachloronaphthalene	10.98
8	87-68-3	1,1,2,3,4,4-hexachloro-1,3-butadiene	11.03
9	3825-26-1	ammonium perfluorooctanoate	11.09
10	2991-51-7	*N*-ethyl-*N*-(heptadecafluorooctyl)glycine	14.06
11	1861-40-1	benfluralin	14.92
12	5216-25-1	1-chloro-4-(trichloromethyl)benzene	15.04
13	77-47-4	1,2,3,4,5,5-hexachloro-1,3-cyclopentadiene	15.71
14	3389-71-7	hexachloronornornadiene	16.19
15	70-30-4	2,2′-methylenebis 3,4,6-trichlorophenol	16.36
16	3734-48-3	chlordene-isomer mixture	16.42
17	634-90-2	1,2,3,5-tetrachlorobenzene	18.85
18	115-32-2	dicofol	19.11
19	67-72-1	hexachloroethane	19.91
20	427-45-2	fluorotris-*p*-chlorophenylstannane	21.01
21	13356-08-6	hexakis-2-methyl-2-phenylpropyl fenbutatin oxide distannoxane	21.49
22	79538-32-2	tefluthrin	21.89
23	297-78-9	isobenzan	22.20
24	82657-04-3	bifenthrin	23.91
25	68334-67-8	1,1,2,3,4-pentachloro-4-(1-mehylethoxy)-1,3-butadiene	24.62
26	101-20-2	*N*-4-chlorophenyl-*N*′-3,4-dichlorophenyltriclocarban urea	25.89
27	3447-87-8	1,4,7-methanoisosbenzofuran-5(3h)-one	26.39
28	79-94-7	4,4′-1-methylethylidenebis 2,6-dibromophenol	26.56
29	87-86-5	pentachlorophenol	26.82
30	26399-36-0	profluralin	27.32
31	1582-09-8	trifluralin	27.82
32	2338-30-9	4,5,6,7-tetrabromo 2-trifluoromethylbenzimidazole	28.27
33	58-90-2	2,2′-methylenebis 3,4,6-tetrachlorophenol	28.36
34	25167-83-3	tetrachloro-technical phenol	28.39
35	10331-57-4	5,5′-dichloro-3,3′-dinitro-1,1′-biphenyl-2,2′-diol	29.26
36	327-98-0	trichloronate	29.88
37	40487-42-1	pendimethalin	30.72
38	115-29-7	endosulfan	33.72
39	4901-51-3	2,3,4,5-tetrachlorophenol	35.19
40	5902-76-1	methyl(pentachlorophenolato)mercuri	35.76
41	1937-37-7	C.I. Direct Black 38	36.00
42	7682-32-8	benzimidazole-4,5,6-tribromo-2-trifluoromethyl	37.12

necessary to get an estimation with a random error below 5%.[7] For these partial orders the new methods have been applied to find the ranking probabilities, and the result is compared with the true ranking probabilities. In Table 8 the ranking probabilities found by the original method are given. These probabilities are compared with the true ranking probabilities as well. The relative MRSD decrease in average from 7.9% to 2.2%. Considering the worst relative MRSD for the original method (sample no. 25), the relative RMSD decreases from 23.4 to 3.81. In the most successful case the relative MRSD goes as low as 0.77% with the new method (sample no. 33). The largest relative improvement can be found for sample no. 34 where a 90% improvement in the relative RMSD is observed. Also for sample no. 25 which has been used as example in this paper a large relative improvement is observed (84%). The smallest relative improvement is to be found for sample no. 5 (13%). This is also the Hasse diagram with the smallest number of linear extensions. In general the number of objects seems to influence the original method in a negative manner (Figure 5). The new method, on the other hand, seams to perform well independent of the number of objects.

A more direct way to evaluate the estimation of the ranking probability is simply to display the result graphically. For sample no. 25 this is done in Figure 4. The correlation between the true and estimated ranking probability is thus easily visualized. Further the linear correlation described by $ax+b$ can be quantified by calculating the coefficient of determination, R^2. This has been done for the 39 samples in Table 8 both for the new and the original method. In the new method the average coefficient of determination, R^2, is as high as 0.98, whereas in the original method the average correlation is 0.91, which was still considered good. The slope, a, is in the new method 0.98 in average and the intersection, b, is 0.0010 in average. For the original method the slope, a, is 0.88 in average and the intersection, b, is 0.0014 in average.

The evaluation of the ranking probability can further be illustrated by a graphic comparison. In Figure 6 the ranking probability, p_r, for object **g** and **h** in the partial order in Figure 2 is shown. For object **h** it can be seen that the curve describing the ranking probabilities determined by the new method fit better to true ranking probabilities. The original method gives a curve, which is more spread out. However, for object **g** it is illustrated how the new method sometimes slightly overestimates the ranking probability on the most probable ranking position, whereas the original method tends to underestimate the highest ranking probability.

Applying other types of correlation measures, such as the linear correlation coefficient, r, they all provide the same result as was displayed by e.g. the relative RMSD or the coefficient of determination, namely that the new method always provides better results.

To illustrate how the method can be applied a data set on POPs is considered. The POPs were identified in a former study.[12] The aim of the study was through screening of databases to identify new candidates for the UN-ECE POP-Protocol. Four criteria are to be met if a chemical is to be classified as a POPs. The chemical substances have to be bioaccumulative (log K_{ow} > 5), be persistent (half-life in water > 2 month or half-life in soil or sediment > 6 months), have potential for long-range transboundary atmospheric transport (atmospheric half-life greater than 2 days and a vapor pressure < 1000 Pa), and be toxic (no cutoff values are identified in the Protocol). The specific cut off values given in the UN-ECE POP-Protocol and applied in the study are given in brackets. During the screening of the databases 12 chemicals were found to fulfill the criteria. However, 30 chemical substances, which almost fulfilled the criteria, were also identified. The 42 chemical substances and their descriptors are given in Table 10. Based on these data the chemical substances have been ranked applying the new random linear extension method developed in this study. When data are missing a median of the descriptor is applied in order not to favor or disfavor the chemical substances with missing data. The result of the ranking is given in Table 11. It is interesting to note that only 5 of the 12 chemical substances fulfilling all the criteria are among the 12 chemical substances given the highest rank in the ranking exercise (Table 9). If the data set were separated into classes and an event-space approach was applied[28] the 12 chemical substances fulfilling all five criteria would share the highest possible rank. However, the fact that other chemicals substances than those fulfilling the criteria are among the top 12 chemical substances indicates that other chemicals substances than the originally identified might be of interest for regulation.

CONCLUSION

When an approximation of the mutual ranking is used to estimate the ranking probabilities using random linear extensions the result is remarkably improved. The relative MRSD decrease in average from 7.9% to 2.2%. In the most successful case the relative MRSD is only 0.77%. The relative improvement ranges from 13 to 90%. Further the new algorithm appears to be independent of the number of objects or linear extensions. Displaying the correlation graphically the coefficient of determination, R^2, increases from 0.91 to 0.98. A graphical display of the ranking probability for an object shows that the new method gives a better fit with the true ranking probabilities.

REFERENCES AND NOTES

(1) *Order Theoretical Tools in Environmental Sciences, Proceedings of Workshop, November 16, 1998, Berlin, Germany*; Brüggemann, R., Ed.; Leibniz Institute of Freshwater Ecology and Inland Fisheries: Berlin, Germany, 1998; Heft 6, Sonderheft I, p 117.

(2) *Order Theoretical Tools in Environmental Sciences, Proceedings of the Second Workshop, October 21, 1999, Roskilde, Denmark*; NERI Technical Report No. 318, May 2000; Sørensen, P. B., Carlsen, L., Mogensen, B. B., Brüggemann, R., Luther, B., Pudenz, S., Simon, U., Halfon, E., Bittner, T., Voigt, K., Welzl, G., Rediske, F., Eds.; National Environmental Research Institute: Roskilde, Denmark, 2000.

(3) *Order Theoretical Tools in Environmental Science and Decision Systems, Proceedings of the Third Workshop, November 6−7, 2000, Berlin, Germany*; Pudenz, S., Brüggemann, R., Lühr, H.-P., Eds.; Leibniz Insitute of Freshwater Ecology and Inland Fisheries: Berlin, Germany, 2000; Heft 14, Sonderheft IV, p 222.

(4) *Order Theoretical Tools in Environmental Science, Order Theory Meets Multivariate Statistics, Proceedings of the Fourth Workshop, November 5−6, 2001 Iffeldorf, Germany*; Voigt, K., Welzl, G., Eds.; Shaker Verlag: Aachen, Germany, 2002.

(5) Sørensen, P. B.; Lerche, D.; Carlsen, L.; Brüggemann, R. Statistically Approach for Estimating the Total Set of Linear Orders − A possible way for analysing larger partial order sets. In *Order Theoretical Tools in Environmental Science and Decision Systems, Proceedings of the Third Workshop November 6−7, 2000 Berlin, Germany*; Pudenz, S., Brüggemann, R., Lühr, H.-P., Eds.; Leibniz Institute of Freshwater Ecology and Inland Fisheries: Berlin, Germany, 2000; Heft 14, Sonderheft IV.

(6) Sørensen, P. B.; Lerche, D. Quantification of the uncertainty related to the use of a limited number of random linear extensions for large partial orders. In *Order Theoretical Tools in Environmental Science, Order Theory Meets Multivariate Statistics, Proceedings of the Fourth Workshop, November 5−6, 2001 Iffeldorf, Germany;* Welzl, G., Voigt, K., Brüggemann, R., Eds.; Limnologische Station der Technischen Universität München: Germany, 2001.

(7) Lerche, D.; Sørensen, P. B. Evaluation of the ranking probabilities for partial orders based on random linear extensions. *Chemosphere*, in print May 2003.

(8) Lerche, D.; Brüggemann, R.; Sørensen, P. B.; Carlsen, L.; Nielsen, O. J. A comparison of Partial Order Technique with Three Methods of Multi-Criteria Analysis for Ranking of Chemical Substances. *J. Chem. Inf. Comput. Sci.* **2002**, 42, 1086−1098.

(9) Lerche, D.; Sørensen, P. B.; Larsen, H. S.; Carlsen, L.; Nielsen O. J. Comparison of the Combined Monitoring-based and Modelling-based Priority Setting Scheme with Partial Order Theory for Ranking of Chemical Substances. *Chemosphere* **2002**, 49, 637−649.

(10) Nielsen, T. S.; Lerche, D.; Sørensen, P. B. Ranking Contaminated Sites using Partially Ordering Method. *Environ. Toxicol. Chem.* **2003**, *22*, in print.

(11) Carlsen, L.; Lerche, D.; Sørensen, P. B. Improving the predicting power of partial order based QSARs through linear extensions. *J. Chem. Inf. Comput. Sci.* **2001**, 42, 806−811.

(12) Lerche, D.; van de Plassche, E.; Schwegler, A.; Balk, F. Selecting chemical substances for the UN-ECE POP-Protocol. *Chemosphere* **2002**, *47*, 617−630.

(13) Davey, B. A.; Priestley, H. A. *Introduction to Lattices and Order*; Cambridge Mathematical Textbooks: UK.

(14) *Algorithms and Order*; Rival, I., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989; NATO ASI Series, Series C: Mathematical and Physical Sciences − Vol. 255.

(15) Bollobas, B. *Combinatorics, Set Systems, Hypergraphs, Families of Vectors and Combinatorial probability;* Cambridge University Press: Cambridge, UK, 1986.

(16) Birkhoff, G. In *Lattice Theory*; American Mathematical Society, Ed.; Providence, RI, 1984.

(17) Trotter, W. T. *Combinatorics and Partially Ordered Sets, Dimension Theory*; The John Hopkins University Press: U.S.A., 1992.

(18) Brüggemann, R.; Halfon, E. *Theoretical Base of the Program "Hasse"*; GSF-Bericht 20/95, GSF-Forschungszentrum, Oberschleissheim, Germany, 1995.

(19) Hasse, H. *Über die Klassenzahl abelscher Zahlkörper*; Akademie Verlag: Berlin, Germany, 1952.

(20) Brüggemann, R.; Halfon, E.; Welzl, G.; Voigt, K.; Steinberg, C. E. W. 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.

(21) Graham, R. L. Linear Extensions of Partial Orders and the FKG Inequality. In *Ordered Sets*; Rival, I., Ed.; 1982; pp 213−236.

(22) Winkler, P. Average height in a partially ordered set. *Discrete Mathematics* **1982**, *39*, 337−341.

(23) Atkinson, M. D.; Chang, H. W. Extensions of partial order of bounded width. *Congressus Numerantium* **1986**, *52*, 21−35.

(24) Atkinson, M. D. The complexity of orders. In *Algorithms and Order*; Rival, I., Ed.; Dordrecht, 1989; NATO series C: Mathematical and Physical Sciences, Vol. 255, pp 195−230.

(25) Edelman, P.; Hibi, T.; Stanley, R. P. *A Recurrence for linear extensions;* 1989; Order 6, pp 15−18.
(26) Stanley, R. P. *Enumerative combinatorics*; Wadsworth&Brooks/cole: Monterey, 1986; Vol. I, 306 pp.
(27) Brüggemann, R.; Lerche, D.; Sørensen, P. B. First attempts to relate structures of Hasse diagrams with mutual probabilities. Manuscript in preparation.
(28) Sørensen, P. B.; Gyldenkærne, S.; Lerche, D.; Brüggemann, R.; Thomsen, M.; Fauzer, P.; Mogensen, B. B. New ranking probability approach applied for prioritisation using multiple criteria. Case: Pesticides and GIS. Manuscript in preparation.

CI0300036