

Prediction of odours of aliphatic alcohols and carbonylated compounds using fuzzy partition and self organising maps (SOM)

K. Audouze, F. Ros, M. Pintore and J. R. Chrétien*

Laboratory of Chemometrics & BioInformatics, University of Orléans, BP 6759, 45067 Orléans Cedex 2, France

A set of 114 olfactory molecules divided into fruity, ethereal and camphoraceous compounds, was submitted to an analysis by Kohonen Neural Networks, also known as Self Organising Map (SOM). The compounds are represented in a hyperspace derived from their molecular descriptors and SOM gives a useful projection of this hyperspace onto a 2D map. Owing to the complexity of the olfaction mechanism, evidenced by the fact that one compound can exhibit simultaneously different properties, SOM alone is unable to take into account the olfaction diversity of the original 114 compounds.

Then, a Fuzzy Partition method was applied on the Kohonen map previously developed. The obtained results allowed delineating different representative zones for the three odours,

expressing more closely the olfactory richness. The ability of the Hybrid System combining SOM and Fuzzy Partition to model the three odours was validated by dividing the 114 compounds into a training set and a test set, including 86 and 28 molecules, respectively. The most important olfactory characteristics were reproduced satisfactorily for the entire test set compounds.

Introduction

Recent developments in combinatorial synthesis and in High Throughput Screening (HTS) have given an impulse to research in new areas of analytical chemistry [1], or have

largely renewed the interest for the exploitation of large databases of chemical compounds in pharmaceutical or agrochemical companies. Automated data classification in large databases with help of molecular diversity analysis has become an indispensable tool in drug design when it comes to: (i) searching for new leads, (ii) delineating relatively homogeneous training sets, (iii) selecting compounds with required physico-chemical characteristics or even biological properties. The state of the art in olfaction prevents advances in the pharmaceutical field related to computer aided molecular design from being directly transposed into the olfaction one. In fact, the absence of crystallised protein and 3D data remains the major drawback to promote the most up-to-date computer aided design strategies, although large families of olfactory receptors exist, they include about a thousand subtypes [2]. The primary structures have been derived [2–9] for 30 of them and various techniques of homology modelling have been applied, but deriving acceptable 3D structures often represents a lot of work, which does not prevent great uncertainties in the estimated 3D geometrical parameters.

Furthermore, the difficulty of data mining in relatively large olfaction databases is amplified by the fact that one compound can have different odours and its activity is usually expressed only in a qualitative way. Another source of complexity derives from the fact that one receptor can recognise different chemical determinants and the same compound can be active on different receptors.

When developing automation procedures in the exploitation of molecular diversity in olfaction chemical databases, neural networks (NN) have been shown to be a powerful pattern recognition paradigm, able to analyse various types of data [10]. Self organising maps (SOM), also known as Kohonen neural networks, offers new means for data mining in databases related to biologically active compounds [11,12]. The compounds of the database are represented in a hyperspace derived from their molecular descriptors. The purpose of SOM is to give a 2D map representation of compound distribution in the original hyperspace. Compared to other techniques like principal component analysis (PCA), it presents the great advantage to obtain a projection in a single map by integrating some non-linearity in the distances between the compounds. As far as we know, analysis of the 2D map has chiefly been done with a binary partition, searching for either a delineation of active and inactive compounds or between compounds exhibiting different activity [12]. The application of SOM is usual when processing huge amounts of data in a high dimension space [13]. However it can be pointed out that there is often some confusion in the role of SOM in data processing. SOM belong to the category of unsupervised techniques which means that they do not allow the generation of classification models which can be used in prediction. They suggest only a map, which is likely to represent data distribution in the descriptor space. It is usual to combine Kohonen maps with back propagation neural networks [14], discriminant analysis [15] or other supervised techniques, but these traditional tools are not always easy to interpret.

The aim of the present paper is to combine SOM and a supervised classification technique, which would be as easily interpretable as Kohonen maps. The fuzzy set theory introduced by Zadeh in 1960 is worth investigating [16–18]. It can provide very simple and interpretable solutions to classification problems within the context of imprecise categories. Fuzzy systems share many similar philosophies with other intelligent computing paradigms, which greatly facilitates implementing adaptation. They have been widely used in the field of process control where the idea is to convert human expert knowledge into fuzzy rules. More recently, several pattern recognition methods have been proposed to automatically generate fuzzy rules from numerical data [19]. This field of research is recent but very promising and it can be related successfully to the exploitation of chemical databases. Furthermore, it should be particularly adapted to applications linked to olfaction in view of the complexity of this phenomenon.

Materials and methods

Compound selection

A set of 114 olfactory compounds was chosen among two sets of 281 and 99 compounds characterised by Schnabel *et al.* [20], and studied by Chastrette *et al.* [10], in order to develop Structure-Olfaction Relationships (SOR). This set was selected in order to avoid the trivial situation where the separation of different olfactory compounds would only be due to their obvious belonging to clearly distinct chemical classes. Furthermore, it is worth underlining that: i) the compound characterisation was performed in the same experimental conditions and given as threshold values; ii) these data were successfully used by several authors in structure-olfaction threshold relationships studies [10,21,22].

The ethereal, camphoraceous and fruity odours were defined according to a 0 to 5 scale [20]. The structures and their olfactory values are presented in table I. Forty-four compounds are very fruity (scale 4 or 5), thirty-one compounds are very ethereal and thirty-nine compounds are very camphoraceous.

Molecular descriptors

The compounds were represented by their coordinates in a 61-dimensional hyperspace derived from 61 different selected molecular descriptors. This set of 61 descriptors includes topological [23], physico-chemical and electronic descriptors [24]. In virtual screening, general descriptors have proved a good compromise, from an efficiency point of view, for data mining in large databases. The advantage of these descriptors is their ability to take into account not only the main structural features of each molecule, but also their global behaviour. Molar refractivity (MR), molar volume (MV), molecular weight (MW) and van der Waals volume (VdWV) were used as size descriptors. The shape

Table I. Set of olfactory compounds.

ID	Compound	Fruity ^a	Ethereal ^b	Camphoraceous ^a
1	1-butanol	4	0	0
2	1-pentanol	4	1	0
3	2-methyl-1-butanol	4	1	0
4	3-methyl-1-butanol	4	1	0
5	1-hexanol	4	1	0
6	3-methyl-1-pentanol	5	0	0
7	5-methyl-1-hexanol	4	0	0
8	2,2-dimethyl-1-pentanol	4	1	0
9	2-pentanol	4	1	0
10	2-heptanol	4	0	0
11	2-methyl-2-hexanol	4	0	0
12	5-methyl-2-hexanol	4	0	0
13	2-nonanol	4	0	0
14	3-heptanol	4	0	0
15	3-octanol	4	0	0
16	6-methyl-3-heptanol	4	1	0
17	5-nonanol	5	0	0
18	cyclopentanol	0	5	0
19	cycloheptanol	0	4	0
20	2,4-pentandiol	0	4	0
21	butanal	4	0	0
22	2,2-dimethylbutanal	0	4	0
23	3-methyl-2-pentanone	4	0	0
24	2-methyl-3-pentanone	0	4	0
25	3,3-dimethyl-2-butanone	0	4	0
26	2-heptanone	4	2	0
27	4-heptanone	4	2	0
28	3-methyl-2-hexanone	0	4	0
29	4-methyl-2-hexanone	0	4	0
30	5-methyl-2-hexanone	4	1	0
31	2-methyl-3-hexanone	0	4	0
32	4,4-dimethyl-2-pentanone	0	5	0
33	2,2-dimethyl-3-pentanone	0	5	0
34	2,4-dimethyl-3-pentanone	0	4	0
35	2-octanone	5	0	0
36	3-octanone	4	0	0
37	6-methyl-2-heptanone	4	0	0
38	5-methyl-3-heptanone	4	2	0
39	2-methyl-4-heptanone	4	0	0
40	3,4-dimethyl-2-hexanone	0	4	0
41	cyclobutanone	4	0	0
42	cyclopentanone	0	4	0
43	3-methylcyclohexanone	0	4	0
44	cyclononanone	0	4	0
45	cyclodecanone	0	4	0
46	cycloundecanone	0	4	0
47	methylbutanoate	0	4	0
48	ethylpropanoate	0	5	0
49	methyl-methylpropanoate	0	4	0
50	butyl-ethanoate	4	2	0
51	methyl-2-methylbutanoate	0	5	0
52	ethyl-pentanoate	0	4	0
53	propyl-butanoate	4	1	0
54	pentyl-ethanoate	5	0	0
55	2-pentyl-ethanoate	0	4	0
56	ethyl-2-methylbutanoate	0	5	0
57	ethyl-3-methylbutanoate	0	5	0
58	propyl-methylpropanoate	4	2	0

Table I. (continued).

ID	Compound	Fruity ^a	Ethereal ^b	Camphoraceous ^a
59	2-propyl-methylpropanoate	0	4	0
60	3-methylbutyl-ethanoate	4	2	0
61	butyl-butanoate	4	0	0
62	ethyl-4-methylpentanoate	4	0	0
63	2-propyl-methylbutanoate	0	4	0
64	propyl-3-methylbutanoate	5	2	0
65	2-butyl-methylpropanoate	0	4	0
66	methylpropyl-methylpropanoate	4	0	0
67	2-pentyl-butanoate	4	1	0
68	2-heptyl-ethanoate	0	4	0
69	2-butyl-2-methylbutanoate	0	4	0
70	methylpropyl-2-methylbutanoate	0	4	0
71	3-methylbutyl-methylpropanoate	4	0	0
72	pentyl-pentanoate	5	0	0
73	pentyl-2-methylbutanoate	4	2	0
74	3-methyl-2-methylbutanoate	0	4	0
75	2-ethylhexyl-ethanoate	0	4	0
76	2,2-dimethylpropan-1-ol	0	0	4
77	2-ethylbutan-1-ol	0	0	4
78	2,2-dimethylbutan-1-ol	0	0	4
79	2,3dimethylpentan-1-ol	2	0	4
80	2-ethylhexan-1-ol	2	0	4
81	2-methylpropan-2-ol	0	0	4
82	2-methylbutan-2-ol	0	0	5
83	3-methylbutan-2-ol	0	0	4
84	2-methylpentan-2-ol	1	0	4
85	3-methylpentan-2-ol	0	0	4
86	4-methylpentan-2-ol	2	0	4
87	2,3dimethylbutan-2-ol	0	0	5
88	3,3-dimethylbutan-2-ol	0	0	5
89	3-methylhexan-2-ol	1	0	5
90	2,3-dimethylpentan-2-ol	0	0	4
91	2,3-dimethylhexan-2-ol	0	0	4
92	3,4-dimethylhexan-2-ol	0	0	4
93	pentan-3-ol	0	0	4
94	hexan-3-ol	0	0	4
95	2-methylpentan-3-ol	0	0	4
96	3-methylpentan-3-ol	0	0	5
97	2-methylhexan-3-ol	1	0	4
98	3-methylhexan-3-ol	0	0	4
99	4-methylhexan-3-ol	0	0	4
100	5-methylhexan-3-ol	0	0	4
101	3-ethylpentan-3-ol	0	0	4
102	2,2-dimethylpentan-3-ol	0	0	4
103	2,3-dimethylpentan-3-ol	0	0	4
104	2,4dimethylpentan-3-ol	0	0	4
105	4-methylheptan-3-ol	1	0	4
106	4-ethyl hexan-3-ol	0	0	4
107	2,2-dimethylhexan-3-ol	1	0	4
108	2,3-dimethylhexan-3-ol	1	0	4
109	2,4-dimethylhexan-3-ol	0	0	4
110	2,5-dimethylhexan-3-ol	2	0	4
111	3,5-dimethylhexan-3-ol	0	0	4
112	2,2,4-trimethylpentan-3-ol	0	0	4
113	2-methylheptan-4-ol	2	0	4
114	3-methylheptan-4-ol	2	0	4

^a The odor intensity was derived from Schnabel *et al.* [20]

features of the molecules were characterised by the topological indices which account for the ramification degree, the oblong character, etc. The following molecular descriptors were used: 20 molecular connectivity indices (${}^0\chi$, ${}^1\chi$, ${}^2\chi$, ${}^3\chi_C$, ${}^3\chi_P$, ${}^4\chi_P$, ${}^4\chi_{PC}$, ${}^5\chi_P$, ${}^5\chi_C$, ${}^6\chi_P$, ${}^0\chi^v$, ${}^1\chi^v$, ${}^2\chi^v$, ${}^3\chi^v_C$, ${}^3\chi^v_P$, ${}^4\chi^v_P$, ${}^4\chi^v_{PC}$, ${}^5\chi^v_P$, ${}^5\chi^v_C$, ${}^6\chi^v_P$), a series of information indices (IC^0 , SIC^0 , CIC^0 , IC^1 , SIC^1 , CIC^1 , IDW), Wiener index (W), centric index (C), Balaban index (J), Gutman index (M2), Platt number (F), counts of paths of lengths 1-4, counts of vertices with 1-4 nearest neighbours. The number of N , O and S atoms in a molecule was also considered. A lipophilicity descriptor represented by the octanol/water partition coefficient ($\log P_{\text{oct/water}}$) was calculated using Hansch and Leo method [25]. Another descriptor was derived from the electronegativity of molecules (E_M^S) according to Sanderson [26].

SOM

SOM is a non-linear mapping technique which aims to give a 2D space representation of a given set of points from a multidimensional space derived from a large series of molecular descriptors [15]. Each point of this set corresponds to a node of the SOM. The co-ordinates of these nodes are represented by the weighted connections between the SOM and the input layers. Each point of the original data set occupies the nearest node in the SOM layer, in terms of Euclidean distance. Hence, the neighbouring points of the original descriptor space remain neighbours on the SOM. Training SOM consists in rearranging the Kohonen layer nodes over the hyperspace by gradually adjusting their weights. A useful option, "interpolation", allows converting the discrete coordinates of the Kohonen layer nodes into continuous coordinate values, averaging the discrete coordinates of the three nodes closest to the raw data point. A detailed description of the algorithm used in this study may be found in reference [27].

The following parameters were used in the data processing, activating the interpolate option: number of columns = 10, number of rows = 10, coefficient γ for bias calculations = 0.01, number of iterations for the training phase = 50000, coefficient β for frequency calculations = 0.001. The calculations were performed using the Neural-Works II software [28].

Fuzzy partition

The generation of fuzzy rules from numerical data, in order to classify a group of compounds, consists of two steps: (i) partitioning a working space into fuzzy subspaces; (ii) defining a fuzzy rule for each fuzzy subspace [29].

Working space partition

Let us assume that the working space is a SOM space, a 2D map represented by a square with two partitions $\{A=A_1, \dots, A_M\}$ and $\{B=B_1, \dots, B_N\}$ for the first and the second axis, respectively (Fig. 1). M , N are the number of fuzzy subspaces and $M \times N$ is the number of fuzzy rules generated. If the coordinates of a generic compound C , in the SOM space, are (x, y) , a fuzzy rule is represented as follows:

$$\text{if } x \text{ is associated to } A_i \text{ and } y \text{ is associated to } B_j \Rightarrow C \text{ has a defined property } Z, \quad (1)$$

with $i = 1, M$ and $j = 1, N$.

To reduce the dependency of the compound classification on the choice of a partition with fuzzy functions, a fuzzy classification method has been proposed [30] that simultaneously uses several fuzzy partitions of different sizes in a single fuzzy rule based classification system. A coarse partition (for instance $M = 2$ and $N = 2$) leads to a generalist system but also to a system where results are too approximate. A fine partition (for instance $M = 10$ and $N = 10$) leads

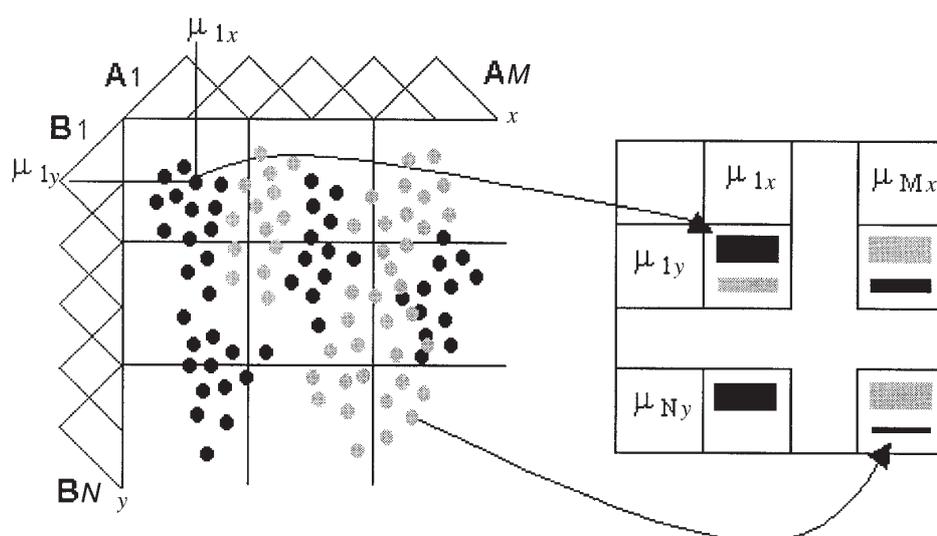


Figure 1. Working space partition of SOM. The space is divided, by fuzzy rules, into decision areas able to discriminate the compound activities. The membership functions μ , for each compound, give a value included in the interval $[0,1]$ representing the possibility for a compound to be associated to the considered activity.

to a precise system of classification but also to a non-generalist system: performances in classification are weak when patterns are unknown. Using several fuzzy partitions simultaneously allows obtaining a compromise between a generalist system, *i.e.* a coarse partition, and a specialist system, *i.e.* a fine partition. Then, the relation (1) becomes:

$$\text{if } x \text{ is associated to } A_{i,k} \text{ and } y \text{ is associated to } B_{j,k} \Rightarrow C \text{ has a defined property } Z, \quad (2)$$

with $i = 1, M; j = 1, N; k = 2, L$, where k represents the amount of fuzzy subspaces according to the partition along the axes.

Procedure of fuzzy rules generation

In a database containing $m = m_1 + m_2$ training set compounds that belong to an active class CLASS(1) and a non active class CLASS(2), the classification problem is to generate fuzzy rules that divide the pattern space into decision areas adapted to discriminate the compound activities. Given a compound C and a subspace (i,j,k) , a rule $R_{i,j,k}$ is expressed as:

$$\begin{aligned} &\text{if } x(C) \text{ is associated to } A_{i,k} \\ &\text{and } y(C) \text{ is associated to } B_{j,k} \Rightarrow C(x,y) \\ &\text{is associated to CLASS}(n) \text{ with } CF_{i,j,k}(n) \end{aligned} \quad (3)$$

where $CF_{i,j,k}(n)$ represents the certainty grade of the fuzzy rule with $n = 1, 2$.

Instead of considering a compound assigned to a class with a certain CF, it is better to get its scores in both classes. In this procedure, all the training compounds from each class are put together to determine, for each subspace (i,j,k) , a score CFS(n) according to:

$$\text{CFS}(n) = \sum_{i=1}^M \mu_{ik}(x_{ni}) \cdot \mu_{ik}(y_{ni}) \quad (4)$$

where $\mu(x)$ is the *membership function* which associates any point x with a value included in the interval $[0,1]$. This value represents the possibility for a compound to belong to the considered set.

Then, at the end of the training phase, each subspace is associated with a certainty grade for each class, calculated as:

$$\text{CF}(n) = \text{CFS}(n) / \sum_{k=2}^L \text{CF}_{ijk}(n) \quad (5)$$

Classification procedure of the test set compounds

Each compound score is derived by combining the compound location in the fuzzy partitioned map and the certainty grade $CF_{ijk}(n)$ defined during the training phase. To take into account the real position (x_0, y_0) of the test set compound, the following expression is used to calculate the score in each class:

$$\text{Score (CLASS}(n)) = \max \{ \mu_{ik}(x_0) \cdot \mu_{jk}(y_0) \cdot \text{CF}_{i,j,k}(n) \} \quad (6)$$

with $i = 1, M, j = 1, N$ and $k = 2, L$.

In this work, for the series of 114 olfactory compounds, the generation of fuzzy rules was made using $k = \{3,4,5\}$, each map consisting of 441 points, 21 per axis. Calculations were performed using proprietary software.

Results and discussion

A Kohonen map for the 114 olfactory compounds is shown in figure 2. No special relation between structures and odors was evidenced by using a multi-parametric base with 61 descriptors. It under-lines the difficulty to express the

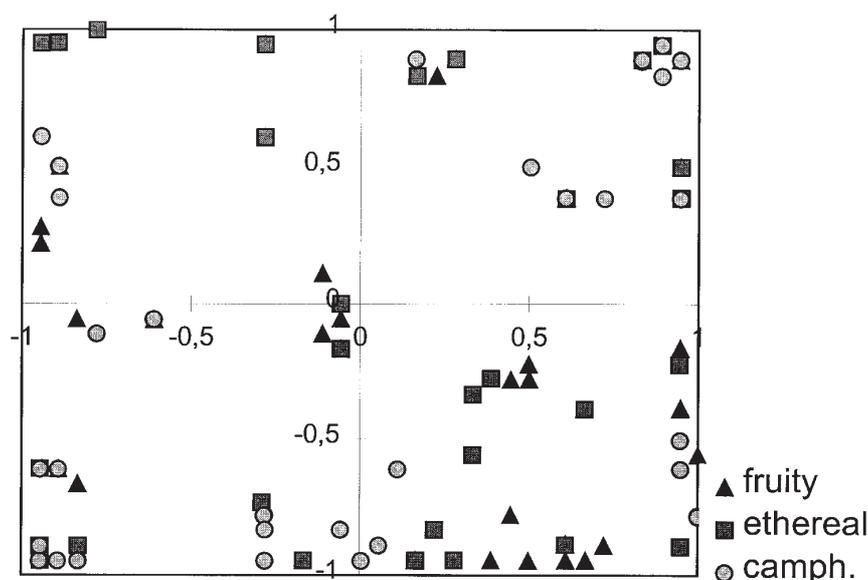


Figure 2. Kohonen map (10 × 10, interpolated map). Projection of the data set of 114 olfactory compounds. All molecular descriptors were taken into account.

diversity of possible notes related to some structure-activity relationships. The weak separations between fruity, ethereal and camphoraceous compounds show the importance of the notion of multi-odours and the complexity of the olfaction

biochemical mechanism. Fuzzy logic can help to express this olfactory richness because it allows, for each compound, to define a score, or membership degree, for the three odours.

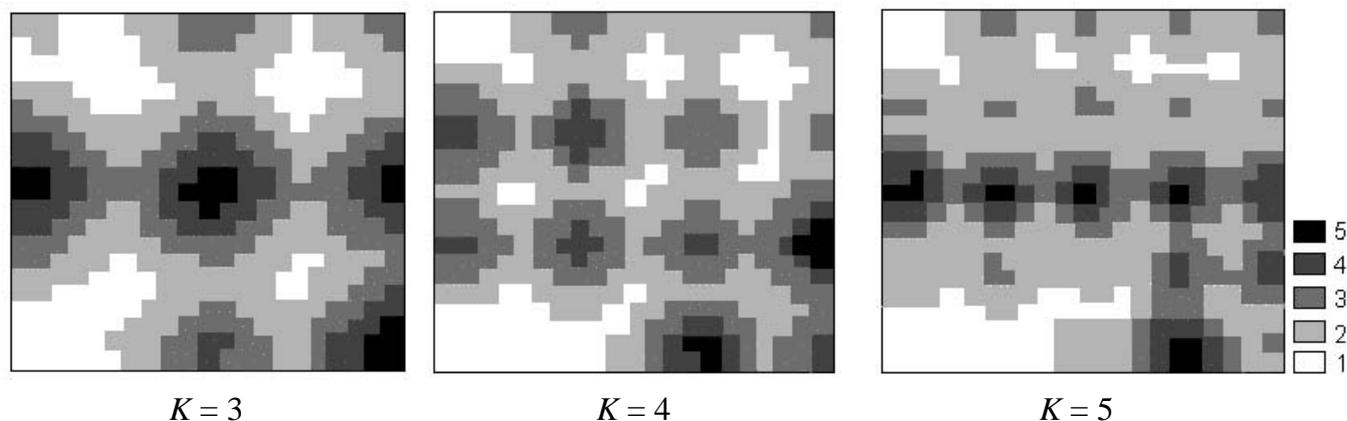


Figure 3. Fuzzy partition of fruity compounds derived from the Kohonen SOM. Three different partitions, corresponding to $K = 3$, 4 and 5, were used, generating 9, 16 and 25 subspaces, respectively.

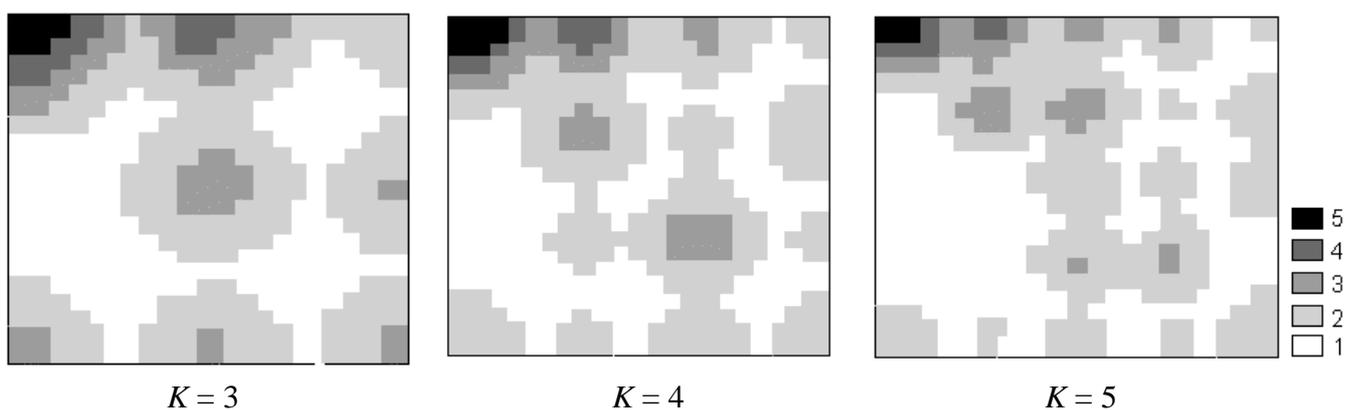


Figure 4. Fuzzy partition of ethereal compounds derived from the Kohonen SOM.

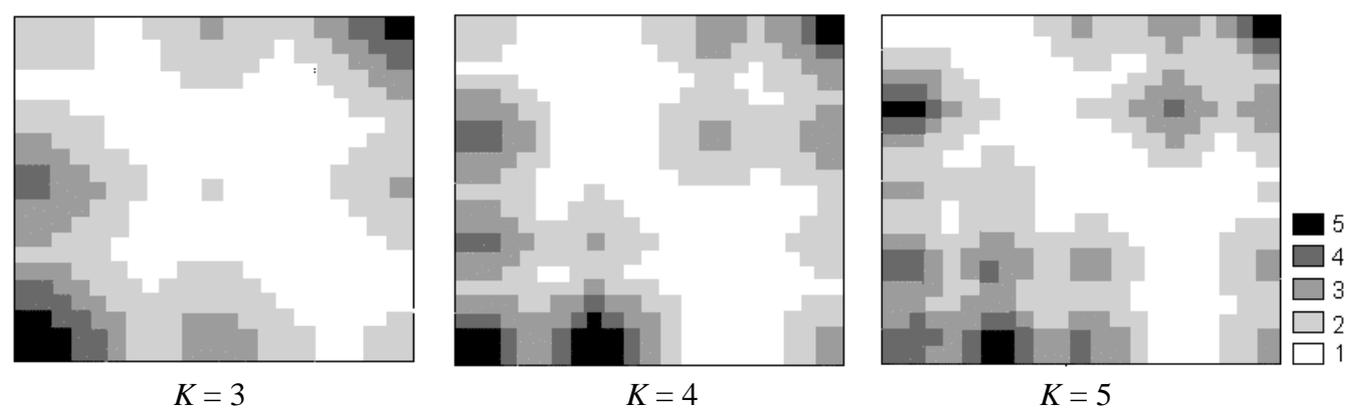


Figure 5. Fuzzy partition of camphoraceous compounds derived from the Kohonen SOM.

Figures 3, 4 and 5 show the fuzzy method scores for the fruity, ethereal and camphoraceous compounds, respectively. Increasing k values lead to more accurate compound partitions. Figure 3 shows that the fruity compounds occupy the central region of the map and the lower part of the right hand side corner. Figure 4 illustrates that the greatest part of the ethereal compounds are grouped in the upper part of the left hand side corner, whereas only a small score is present in the middle of the map. Finally, the camphoraceous compounds are concentrated above all in the extremities of the map (see Fig. 5), with the most important regions located in the upper part of the right hand side corner and in the lower part of the left corner, whereas only a very little score appears in the centre of the map.

These results show that a natural regrouping of the compounds was obtained by the procedure used, allowing define different representative zones of the map for the three olfactory properties. In fact, only for $k = 3$ and 4, it is possible to observe, in the middle of the map at the left hand side, a unique weak superposition between fruity and camphoraceous regions. For the most accurate compound partition, $k = 5$, no superposition between the different olfactory zones can be evidenced.

Validation

In order to validate the combination of SOM and Fuzzy Partition to model the three odours, 86 compounds were randomly chosen to define the training set and the remaining 28 compounds were included in the test set. After deriving SOM for the training set, the test set compounds were projected into the map and their xy coordinates were calculated. These coordinates were used as input in the Fuzzy Partitioned map and their membership degree was determined for each odour. The comparison between predicted and actual values for the test set compounds is reported in table II. The predicted values, represented by the membership degrees, are defined within a 0 to 1 range.

It must be noted that the profile of the selected raw data, based on the score on the Schnabel's scale, is strong only for one odour. There is a weak contribution with a score 1 or 2 for one third of these compounds relatively to the second odour, whereas the third odour equals zero for all of them. It is possible to observe that for most compounds of this test set the odour profile were predicted correctly. Nevertheless one exception is noticed for compound 60, the prevalent fruity score in the raw data is predicted as the second one. Furthermore, for compounds 53, 57, 62, 63, 97, a degree of contribution of the second odour at the level of the first one is predicted, instead of being weaker or negligible.

Four other divisions of the data set in training and test sets were tested, any of them including 86 and 28 compounds in the two sets, respectively. A prediction power of 81 %, in average, was got.

Table II. Comparison between predicted and actual values for the test set compounds.

ID	Predicted scores			Actual scores		
	Fruity	Ethereal	Camph.	Fruity	Ethereal	Camph.
2	0.34	0.08	0.00	4	1	0
4	0.31	0.08	0.16	4	1	0
7	0.34	0.01	0.00	4	0	0
13	0.53	0.00	0.07	4	0	0
26	0.58	0.17	0.00	4	2	0
35	0.58	0.17	0.00	5	0	0
38	0.76	0.00	0.22	4	2	0
53	0.23	0.22	0.00	4	1	0
60	0.18	0.42	0.00	4	2	0
62	0.54	0.39	0.00	4	0	0
28	0.15	0.62	0.00	0	4	0
29	0.15	0.62	0.00	0	4	0
33	0.03	0.54	0.10	0	5	0
44	0.00	0.86	0.00	0	4	0
46	0.00	0.88	0.00	0	4	0
48	0.03	0.44	0.00	0	5	0
56	0.20	0.41	0.00	0	5	0
57	0.16	0.28	0.00	0	5	0
63	0.22	0.27	0.00	0	4	0
78	0.00	0.00	0.52	0	0	4
82	0.03	0.01	0.89	0	0	5
87	0.00	0.00	0.54	0	0	5
90	0.00	0.00	0.75	0	0	4
97	0.19	0.02	0.20	1	0	4
102	0.02	0.00	0.76	0	0	4
105	0.15	0.00	0.30	1	0	4
107	0.03	0.00	0.47	1	0	4
113	0.29	0.00	0.57	2	0	4

Conclusion

Olfaction is one of the most complex biological properties to be processed, due to four different characteristic points of this biological field: (i) a huge number of receptors is involved in olfaction, (ii) knowledge related to the 3D structure of these receptors is still missing, (iii) different types of chemical compounds can affect the same receptor, (iv) one compound can exhibit simultaneously different odors, (v) one receptor can be triggered by more than one odourant.

For data mining in an olfaction data base, it is important to have more powerful methods than classical projection techniques like principal component analysis (PCA) or even more recent techniques like SOM. In fact, in this study, a Kohonen map built from a multi-parametric base with 61 molecular descriptors is not sufficient to derive a satisfactory relation between structures and odours. It does not

allow illustrating the diversity in the olfaction characteristics of the series based on 114 compounds and the ability for a same compound to exhibit different odours. Despite the interest of SOM in data mining it is chiefly suitable for binary decisions such as active or inactive, and is of limited applicability when multi-property are taken into account.

The Fuzzy Partition method helps to express this olfactory richness because it permits, for each compound, to define a membership degree for the three involved odours. It consists of dividing a map into fuzzy subspaces and assigning a fuzzy rule to each of them. The obtained results showed that for most part, the fruity, ethereal and camphoraceous compounds, despite the overlapping, occupy different regions of the map, allowing delineating different representative zones for the three odours. Ability of the combination of SOM/Fuzzy Partition to model the three odours was validated by dividing the 114 compound series into a training set and a test one, including 86 and 28 molecules, respectively. For all the test set compounds, except one, the most important odour was correctly predicted and, for many molecules, the different olfactory notes were reproduced satisfactorily by their membership degrees.

From a methodological point of view, the fundamental potentiality of Fuzzy Partition is to go beyond active/inactive limitation, by introducing continuity, and to express multi-property for the same compound. It is a requisite in olfaction but worth also to be developed in medicinal chemistry when not only one biological activity is considered but also potential secondary effects or various toxicities dealing with people or environment.

These preliminary results show that the proposed methods are worth investigating more deeply and testing with a large chemical database. Work is underway to exploit a large database of 3000 compounds with 80 olfactory properties.

Acknowledgements

Mrs. Roudnitska and the E. Roudnitska Foundation are gratefully acknowledged for their support and the grant to one of us (K.A). Professor M. Chastrette is acknowledged for his kind comments and encouragement for this work.

References

- Zhao, Y. Z.; van Breemen, R. B.; Nikolic, D.; Huang, C. R.; Woodbury, C. P.; Schilling, A.; Venton, D. L. *J. Med. Chem.* **1997**, *40*, 4006-4012.
- Buck, L.; Axel, R. *Cell* **1991**, *65*, 175-187.
- Fan, W.; Liu, Y.; Parimoo, S.; Weissman, S. *Genomics* **1995**, *27*, 119-123.
- Glusman, G.; Clifton, S.; Roe, B.; Lancet, D. *Genomics* **1996**, *37*, 147-160.
- Issel-Tarver, I.; Rine, J. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10897-10902.
- Nef, P.; Hermans-Borgmeyer, I.; Artieres-Pin, H.; Beasley, L.; Dionne, V.; Heinemann S. F. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8948-8952.
- Nef, S.; Allaman, I.; Fiumelli, H.; Castro, E. D.; Nef, P. *Mech. Dev.* **1996**, *55*, 65-77.
- Parmentier, M.; Libert, F.; Schurmans, S.; Schiffmann, S.; Lefort, A.; Eggerickx, D.; Ledent, C.; Mollereau, C.; Gerard, C.; Perret, J. *Nature* **1992**, *355*, 453-455.
- Schurmans, S.; Muscatelli, F.; Miot, F.; Mattei, M.; Vassart, G.; Parmentier, M. *Cytogenet. Cell. Genet.* **1993**, *63*, 200-204.
- Chastrette, M.; Cretin, D.; Aidi, C. E. J. *Chem. Inf. Sci.* **1996**, *36*, 108-113.
- Bernard, P.; Golbraikh, A.; Kireev, D.; Chrétien, J. R.; Rozhkova N. *Analisis* **1998**, *26*, 333-341.
- Kireev, D.; Chrétien, J.R.; Bernard, P.; Ros, F. *SAR QSAR Environ. Res.* **1998**, *8*, 93-107.
- Varfis, A.; Versino, C. *Neural Network World* **1992**, *2*, 813-834.
- Hecht-Nielsen, R. In *Proceedings of the IEEE. First International Conference of Neural Networks*; Condilland, M.; Butler, C., Eds.; San Diego (USA): SOS Printing, 1987; pp 19-32.
- Kohonen, T. *Self-Organizing Maps*; New York: Springer-Verlag, 1995.
- Zadeh, L. A.; Fu, K. S.; Tanaka, K.; Shimura, M. *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*; New York: Academic Press, 1975.
- Zadeh, L. A. In *Fuzzy sets and their applications to classification and clustering, Classification and Clustering*; Van Ryzin, J., Ed., New York: Academic Press: 1977, pp 251-299.
- Dubois, D.; Prade, H. *Théorie des Possibilités: Applications à la représentation des connaissances en informatique*; Paris: Masson, 1988.
- Pedrycs, W. *Pattern Recognition* **1990**, *23*, 121-146.
- Schnabel, K.-O.; Beltz, H.-D.; Ransen, C. V. *Z. Lebensm. Unters. Forsch.* **1988**, *187*, 215-223.
- Anker, L. S.; Jurs, P. C.; Edwards, P. A. *Anal. Chem.* **1990**, *62*, 2676-2684.
- Zakarya, D. *New J. Chem.* **1992**, *16*, 1039-1042.
- Dearden, J. C. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J., Eds., Dordrecht: Kluwer Academic Publishers, 1990, pp 25-29.
- Sabljić, A. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J., Eds., Dordrecht: Kluwer Academic Publishers, 1990, pp 61-82.
- Hansch, C.; Leo, A. W. *Substituent constants for correlation analysis in chemistry and biology*; New York: Wiley, 1979, pp 18-43.
- Sanderson, R. *Chemical bonds and bond energy*; New York: Academic press, 1976.
- Neural computing: A technology handbook for neuralworks professional II/Plus and Neuralworks explorer*; NeuralWare Inc ,Technical publications group.
- Neural Computing*; NeuralWare, Inc., 1995.
- Lin, Y.; Cunningham, G. J. *Intell. Fuzzy Syst.* **1994**, *2*, 243-250.
- Ishibuchi, H.; Nozaki, K.; Tanaka, H. *Fuzzy Set Syst.* **1992**, *52*, 21-32.