Assessment of Strawberry Aroma through SPME/GC and ANN Methods. Classification and Discrimination of Varieties

LOUISE URRUTY,[†] JEAN-LUC GIRAUDEL,[‡] SOVAN LEK,[‡]
PHILIPPE ROUDEILLAC,[§] AND MICHEL MONTURY*,[†]

Equipe Périgourdine de Chimie Appliquée (EPCA), Laboratoire de Physico et Toxico Chimie des
Systèmes Naturels, Université Bordeaux 1, CNRS (UMR 5472), BP 1043, 24001 Périgueux Cedex,
France, Centre d'Ecologie des Systèmes Aquatiques Continentaux (CESAC), CNRS (UMR 5576),
118 rue de Narbone, Université Paul Sabatier, 31062 Toulouse Cedex, France, and
Centre Interrégional de Recherche et d'Expérimentation de la Fraise (CIREF), Lanxade,
24130 Prigonrieux, France

To provide an efficient and running analytical tool to strawberry plant breeders who have to characterize and compare the aromatic properties of new cultivars to those already known, a HS−SPME/GC−MS analysis method has been coupled with a statistical treatment method issued from the current development of artificial neuron networks (ANN), and more specifically, the unsupervised learning systems called Kohonen self-organizing maps (SOMs). So, 70 strawberry samples harvested at CIREF from 17 known varieties have been extracted by using a DVB/Carboxen/PDMS SPME fiber according to the headspace procedure, and then chromatographed. A panel of 23 characteristic aromatic constituents has been selected according to published results relative to strawberry aroma. The complex resulting matrix, collecting the relative abundance of the 23 selected constituents for each sample, has been input into the SOM software adapted and optimized from the Kohonen approach described by one of the authors. After a period of training, the self-organized system affords a map of virtual strawberries to which real samples are compared and plotted in the best matching unit (BMU) of the map. The efficiency for discriminating the real samples according to their variety is dependent on the number of units selected to define the map. In this case, a 24-unit map allowed the complete discrimination of the 17 selected varieties. Moreover, to test the validity of this approach, two additional samples were blind-analyzed and the results were computed according to the same procedure. At the end of this treatment, both samples were plotted into the same unit as those of the same variety used for training the map.

KEYWORDS: Strawberry variety; aroma; SPME analysis; Kohonen SOM; computational method; discrimination; plant breeding

INTRODUCTION

To develop new strawberry varieties with optimized aroma properties, plant breeders are presently concerned with objective analytical methods for describing aromas of valuable varieties, and comparing aroma qualities between newly produced and old reference varieties. Presently, sensory analysis methods are the only efficient analytical tool, but they can be performed only at the latest stage of the selection, and are then technically laborious and expensive for routine use. For the last forty years, many chromatographic methods have been proposed to describe the chemical composition of the strawberry aroma (*1, 2*), and have demonstrated that more than 360 constituents may be found

in this type of aroma, according to the nature of the analyzed variety and the sample preparation process used (*3−6*). Relationships between aroma constituents and specific sensory characters of varieties have rarely been observed, except for the "musk" odor of wild strawberries due to the presence of relatively high amounts of a typical organic constituent (methyl anthranilate) (*7*). Most of the character impact odorants described in the literature (*8*) and associated to the different varieties studied, are provided by several molecules (i.e., fruity associated to several chemically different esters), and do not allow analysts to propose a chemical signature for the aroma of a given variety. Moreover, the description of such an aroma is made much more complex by other parameters such as space and time. For example, the nature and the composition of the aroma associated with a defined variety will depend on agronomic and climatic conditions, will be different from one season to another, and will even vary upon the time of harvest

* To whom correspondence should be addressed [tel 33 (0)5 53 35 24 29; fax 33 (0)5 53 02 58 80; e-mail: m.montury@epca.u-bordeaux.fr].
† Université Bordeaux 1.
‡ Université Paul Sabatier.
§ Centre Interrégional de Recherche et d'Expérimentation de la Fraise.

during the same season (*4*). For these reasons, which are related to the biological nature of the fruit, it appeared clear that a tool able to give adequate measures of fruit aromas for plant breeders had to be rapid, precise, and easy to use, like a high-resolution photograph. It should work like olfactive receptors, according to a direct and non-invasive process, with a minimum of intermediaries. It should cost as little as possible to allow the operator to collect a great number of measures in order to take into account as many parameter variations as possible. And to give to this tool real efficiency, it must be coupled with powerful statistical treatment software, able to accumulate and integrate the huge number of values relative to the nature and the composition of the main aroma constituents found in each measure, and to reduce the multivariate data obtained and difficult to visualize into a two- or three-dimensional system simple enough for display on graphs.

The solid-phase micro-extraction (SPME), developed by Pawliszyn in the past decade, is a rapid, direct, inexpensive, and efficient technique for sampling liquid or gaseous matrixes, as reported in his two descriptive books (*9, 10*). SPME is a multi-analyte extraction technique that requires no solvents and provides linear results over a wide range of analyte concentrations in a large scope of matrixes (*11−13*). In the field of food aroma analysis, the headspace SPME (*14*) has proved to be an advanced and efficient tool for studies relative to aromas of food, beverages, fruits and fruit juices, coffees, sauces, etc, (*15−19*). In the particular field of strawberry aroma, several works have been published in the past few years (*4, 5, 8, 20*) and some of them are devoted to the need of affording aroma assessment methods to plant breeders (*21−24*). In fact, SPME/GC is a convenient technique for providing the aromatic fingerprint descriptions of each analyzed strawberry variety.

On the other hand, artificial neural networks (ANNs) have shown promise as an advanced implementation to the analysis of multivariate data, and especially in the field of chemical systems (*25−26*). SPME/GC analyses of strawberry samples generate multivariate data with many different variables, which are the aroma constituents (*n* selected components) and their relative concentrations. Each of these constituents may be regarded as constituting a different dimension and its relative concentration as indicating its coordinate on the corresponding axis. In such an *n*-dimensional space, each sample is assumed to reside at a unique position. So, a neural computation method such as the Kohonen self-organizing maps (SOMs), may perform a multivariate analysis of such a data system (*27*). It can seek "clusters" in the data, by following an "unsupervised learning" approach and then, allow the grouping of samples together on the basis of their perceived proximity in the *n*-dimensional space defined above. In terms of strawberry variety aroma, such an ANN analysis would afford a capacity for discriminating between aromas of different variety samples, and beyond this, for defining closeness between different variety aromas. Actually, these last two purposes are of major interest for plant breeding. This type of Kohonen map analysis method has already been successfully applied in the field of chemometrics, for the classification of plant seeds according to their pyrolysis mass spectra (*28*), and also to classify human blood plasma profiles on the basis of H^1 NMR spectroscopic data (*29*).

The aim of this study was to perform the SPME/GC aroma analysis of 17 strawberry varieties through 23 of their major aroma constituents (*n* = 23), selected according to previously cited publications (*8, 21*), and to evaluate the discriminating capacity of this multivariate method by using SOMs, to afford an efficient classifying tool to plant breeders who have to

Table 1. Presentation of the Selected Varieties

code	variety	crossing	origin country
A	Capron Royal	*Fragaria moschata* clone	France
B	CF129	Earliglow × Chandler	France
C	CF157	Chandler × Pajaro	France
D	CF2024	Capron royal × Belrubi	France
E	Ciflorette	Mamie × Belrubi	France
F	Cigaline	Garriguette × Earliglow	France
G	Cigoulette	Belrubi × Pajaro	France
H	Cilady	Scott × Chandler	France
I	Ciloé	Belrubi × Allstar	France
J	Cireine	Scott × Chandler	France
K	Earliglow	(Faisland × Midland) × (Redglow × Surecrop)	USA
L	Marabella	Valeta × Chandler	France
M	Marmolada	Sélection n°15 × Gorella	Italy
N	Miranda	Sélection n°5 × Santana	Italy
O	Profumata di Tortona	*Fragaria moschata* clone	Italy
P	Selene	Chandler × Selva	Italy
Q	Sengana	Markee × Sieger	Germany

establish an objective judgment of fruit aromas and clearly distinguish differences between varieties at the first step of their study. The observed results were also compared to those obtained with more conventional methods such as principal components analysis (PCA) or averaged linkage cluster analysis (ALCA) (*30−31*). Location of SPME/GC data obtained from 2 supplementary test samples, arbitrarily selected from the 17 studied varieties, onto the Kohonen map already trained with these varieties, was also performed as an efficiency test.

MATERIALS AND METHODS

Varieties. The 17 strawberry varieties proposed by CIREF for this study and the crossings from which they are issued are presented in **Table 1**. A and O are clones of a wild species *Fragaria moschata* with a high and particular flavor. K, L, M, N, P, and Q are well-known varieties cultivated in different countries around the world. B, C, E, F, G, H, I,and J are cultivars created by CIREF and are newly commercialized or presently in the breeding program. D is an interspecific crossing between Capron Royal and Belrubi obtained by CIREF.

All of the samples were cultivated under the same controlled selected conditions in the CIREF's experimental plots. The fruits were collected at full maturation, immediately frozen, and stored at −18 °C until analysis.

Chemicals. Standards of the 23 selected volatile compounds were provided by Sigma-Aldrich (Saint Quentin-Fallavier, France), and so were the solvents used. All purities were higher than 95%. Stock solutions of each standard at 1000 mg/L and a mix solution at 25 mg/L were prepared in methanol, under N_2 atmosphere.

GC/MS Analysis. A Varian 3400 GC coupled with a Finnigan Mat ITS 40 ion trap mass spectrometer (Thermo Finnigan, Les Ulis, France) was used. The carrier gas was helium (1 mL/min). The injector temperature was 250 °C and the split ratio was 1:60. Analytes were separated on a PTE5 30 m × 0.32 mm × 0.25 μm column (Supelco, Saint Quentin Fallavier, France) kept at 40 °C for 3 min and then ramped to 60 °C at 2 °C/min and further at 10 °C to 130 °C, held for 8 min at 130 °C and then ramped to 280 °C at 20 °C/min. The transfer line was kept at 250 °C and the detector was held at 220 °C. The MS was tuned to FC43 (perfluorotributylamine), and mass spectra were collected at the rate of 1 scan/second over the mass range (*m/z*) 50−650. All identifications were based on the comparison of retention times and mass spectra of the compounds with those of standards (**Table 2**). For each compound, quantitation was performed by replotting the full-scan data by using selective ions and measuring the corresponding peak area.

Sample Preparation. Sepals of about 100 g of frozen strawberries were dissected. Fruits were cut into several parts, and 25 g of strawberries were defrosted with 10 mL of HPLC-grade water for 1 h. Then they were blended at 8000 rpm (Ultraturrax), and centrifuged at

Strawberry Variety Aroma Analysis by SPME and ANN

J. Agric. Food Chem., Vol. 50, No. 11, 2002 **3131**

Table 2. Peak Number, Retention Time, and Mass Spectra Description of the 23 Selected Compounds

peak no.	compound[a]	RT (min)	fragments (abundance)
1	pentanone-2	2.10	58(41%); 71(35%); 86(100%)
2	methyl butanoate	2.27	74(100%); 87(32%); 103(17%)
3	ethyl butanoate	4.34	71(100%); 88(23%); 116(24%)
4	butyl acetate	4.80	56(62%); 61(100%); 116(25%)
5	butanoic acid	5.59	60(100%); 73(32%); 89(43%)
6	(E)-2-hexenal	5.24	69(84%); 83(67%); 98(12%)
7	(Z)-3-hexen-1-ol	6.03	55(41%); 67(100%); 82(35%)
8	2-methyl butanoic acid	6.09	57(45%); 74(100%); 87(27%)
9	heptanone-2	7.07	58(38%); 71(28%); 115(10%)
10	methyl hexanoate	8.40	74(100%); 87(35%); 131(10%)
11	ethyl hexanoate	12.45	99(25%); 115(11%); 145(100%)
12	hexanoic acid	13.33	60(100%); 73(47%); 117(7%)
13	DMF	15.22	69(34%); 99(10%); 142(100%)
14	DHF	15.53	69(100%); 83(45%); 128(2%)
15	linalool	16.37	71(100%); 81(25%); 93(75%)
16	octanoic acid	18.93	60(100%); 101(35%); 73(65%)
17	methyl anthranilate	21.37	92(40%); 119(75%); 151(100%)
18	eugenol	22.06	103(41%); 149(42%); 164(100%)
19	γ-decalactone	25.13	85(100%); 128(10%); 171 (24%)
20	δ-decalactone	26.16	71(100%); 99(87%); 171(46%)
21	(Z)-nerolidol	28.04	69(100%); 93(78%); 107(54%)
22	(E)-nerolidol	29.07	69(100%); 93(78%); 107(54%)
23	γ-dodecalactone	31.11	85(100%); 199(10%); 128(10%)

[a] DMF, 2,5-dimethyl-4-methoxy-3(2H)-furanone; DHF, 2,5-dimethyl-4-hydroxy-3(2H)-furanone. Underlined ions were used for quantification.

3000 *g* and 4 °C (Jouan GR412) for 15 min. The supernatant was collected in a 50-mL volumetric flask and adjusted to 50 mL with HPLC-grade water. SPME was performed immediately after sample preparation because of the observed degradation of some compounds, esters for instance, in this aqueous solution.

SPME Procedure. A manual holder (Supelco,) was used for all the experiments. Compared to all the commercially available fibers, the 2 cm, 50/30 μm divinylbenzene/carboxen on poly(dimethylsiloxane) StableFlex fiber has shown the best efficiency to extract the selected compounds. A 3-mL aliquot of the fresh strawberry preparation was transferred into a 5-mL Teflon-lined septum cap vial equipped with a glass-coated magnetic bar. To favor the transfer of the analytes from the aqueous solution to the headspace, 0.6 g of NaCl was added, and the solution was stirred at 1000 rpm at room temperature. After a 15-min equilibration time between the solution and the headspace, the fiber was exposed for 30 min in the headspace. The fiber was then withdrawn and introduced into the injector port of the GC for desorption at 250 °C during 3 min, in the splitless mode. Then the split valve was opened (ratio 1:60) but the fiber was kept for 17 min in the injector for cleaning.

Statistical Calculation. Chemical data were characterized by large differences between the magnitude levels of observed signals according to the measured variables. To standardize these data, obtained signals, *x*, relative to each selected aromatic constituent in a given variety, were transformed into *y* values according to the relation

$$y = \log(x + 1) \tag{1}$$

and then converted into a centered reduced variable *z* as follows:

$$z = \frac{y - \bar{y}}{\sigma_y} \tag{2}$$

where *z* is the standardized value corresponding to the observed *x* value, and $\bar{y}$ and σ_y are the mean and standard deviation of the calculated *y* value, respectively.

To extract the structure of the high dimensional data matrix formed by the 23 chemical constituents analyzed in the 70 strawberry samples (S_1 to S_{70}), an unsupervised neural network method, the so-called Kohonen self-organizing map algorithm (SOM), has been used (*32*). The results obtained by this recently described method have been compared to those resulting from more traditional statistical methods
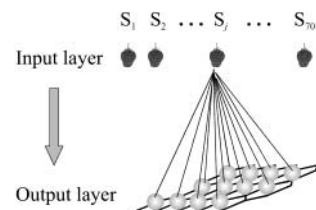
Figure 1. Representation of the nonsupervised artificial network (i.e., Kohonen self-organizing map), showing the input neurons (observed data) and the output neurons (virtual strawberries) organized on a rectangular two-dimensional grid.

such as the principal components analysis, PCA (*33*), or the average linkage cluster analysis, ALCA (*34*).

Calculations have been performed with a PC equipped with an Intel Pentium III-500 processor. For the SOM development, the MATLAB software has been used with a program file written by the authors (*35*). PCA and ALCA were carried out using the 1.2 version of R software (*36*).

Kohonen Self-Organizing Map. The SOM algorithm performs a nonlinear projection of that data set onto a rectangular grid (*r* rows and *c* columns) laid out on a hexagonal lattice with *S* hexagons ($S = r\cdot c$): the Kohonen map. The lattice type of the array can also be defined to be rectangular; hexagonal is effective for visual display (*32*). Formally, the Kohonen neural network consists of two layers: the first one (input layer) is connected to each vector of the dataset, and the second one (output layer), which is the Kohonen map, forms a two-dimensional array of nodes (**Figure 1**). Each neuron of the output layer stores a virtual strawberry with chemical components to be computed. During the training, the virtual strawberries are modified in order to approximate the probability density function of the input data. The main characteristic of the SOM classification is the conservation of the topology: close strawberries (in a chemical sense) are associated with the same node or to nearby nodes on the map.

The SOM algorithm is an unsupervised learning procedure which can be summarized as follows: (1) The virtual strawberries are initialized with random samples drawn from the input data set. (2) The virtual strawberries are updated in an iterative way: a sample unit is randomly chosen as an input unit; the Euclidean distance between this sample unit and every virtual strawberry is computed; the virtual strawberry closest to the input is selected and called "best matching unit" (BMU); and the chemical constitution of the BMU and its neighbors are moved a bit toward the constitution of the input unit.

The training was broken down into two parts previously defined by Mellsen et al. (*37*) as follows: (1) ordering phase (the 2000 first steps): when this first phase takes place, the strawberries are highly modified in a wide neighborhood of the BMU; (2) tuning phase (the number of steps was fixed at 500 times the number of neurons in the Kohonen map): during this phase, only the BMU and the virtual strawberries adjacent to the BMU are lightly modified.

At the end of training, the chemical components are known for each virtual strawberry, the BMU is determined for each strawberry, and each real strawberry is set in the corresponding hexagon of the Kohonen map. Strawberries which are neighbors on the grid are expected to represent neighboring clusters of strawberries; consequently, strawberries with a large distance from each other (according to chemical components) are expected to be distant in the feature space.

Some Kohonen maps of different sizes were trained: first, a 4-unit map was chosen, followed by maps with 8, 12, 16, 20, and 24 neurons, respectively. Increasing the map size was stopped when all the strawberries appearing in the same unit were of the same variety.

Principal Components Analysis. PCA was used to reduce the dimensionality of data and to transform interdependent variables into significant and independent component ones. In the present work, the analyses were conducted on the correlation matrix.

Cluster Analysis. The goal of cluster analysis is to subdivide a data set into different groups according to similar characteristics. For example, in our case strawberries with similar chemical components were assigned to the same clusters. The clusters were arranged in a
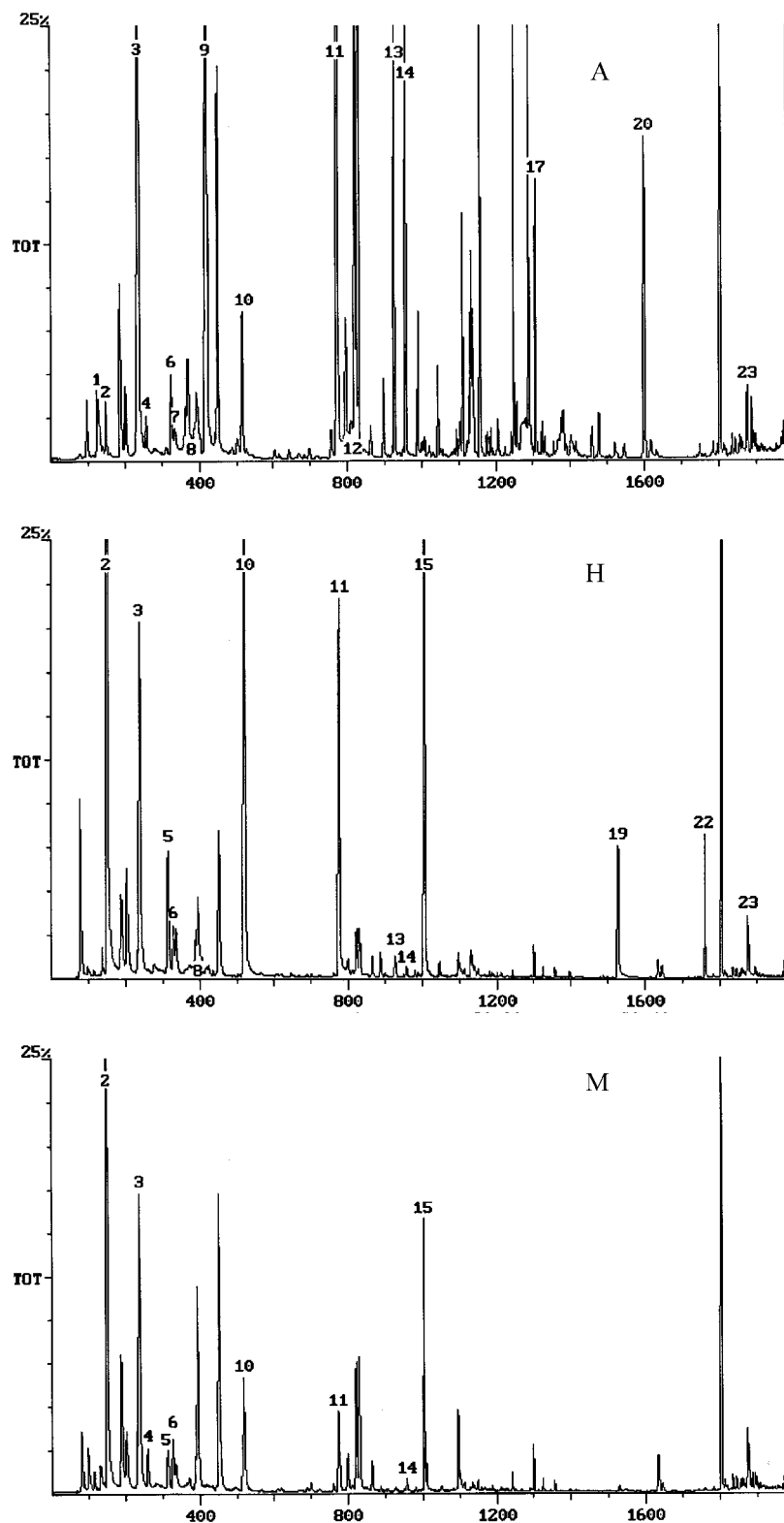
Figure 2. Chromatograms from Capron royal (A, top), Cilady (H, middle), and Marmolada (M, bottom) varieties. For peak identification see **Table 2**.

hierarchical treelike structure called a dendrogram. In this work, the method of average linkage between groups (*38*) was used.

RESULTS AND DISCUSSION

Two sample preparations were realized for each variety, except for F which was processed 3 times, and two SPMEs have been done from all resultant aqueous solutions.

All obtained chromatograms correspond to raw aroma profiles of fruit samples of the selected 17 varieties, all assessed by the same described procedure. They indicate that the 23 selected strawberry aroma constituents can be easily characterized and quantified. For example, furaneol and its methyl ether (DHF and DMF, respectively), which are reported as important components in typical strawberry aroma but seldom detected (*4*), are clearly identified as peaks 13 and 14, respectively. Representative chromatograms of the varieties Capron Royal (A), Cilady (H), and Marmolada (M) are presented in **Figure 2**.
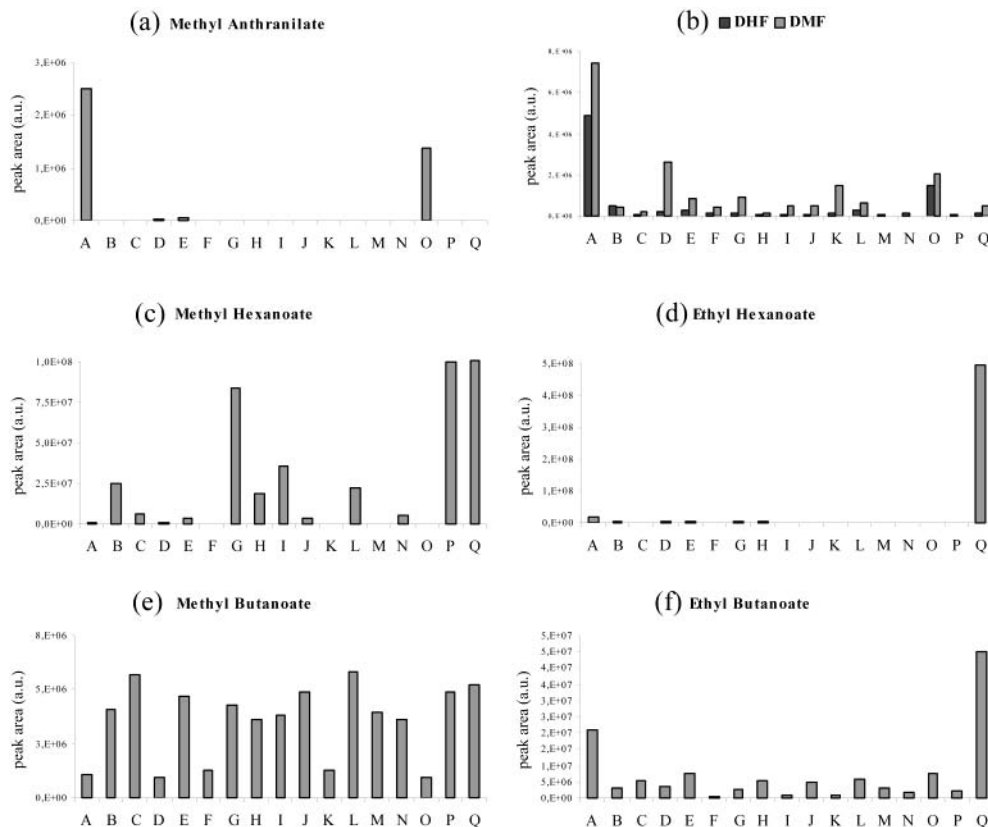
Figure 3. Comparison of the varieties on their content of methyl anthranilate (a), DHF and DMF (b), methyl hexanoate (c), ethyl hexanoate (d), methyl butanoate (e), and ethyl butanoate (f). Each bar represents the average value obtained from all the measures realized (at least 4 measures per variety).

The obvious differences between these 3 chromatograms, in terms of constitution and relative composition, clearly point out the potential efficiency of the coupled SPME/GC method used for the discrimination of variety aromas. For example, according to the number of peaks and their relative intensities, the chromatogram obtained from Capron Royal (A) suggests higher aromatic properties compared to that of Marmolada (M). In the same conditions, the Cilady's chromatogram (H), appears as an intermediate between them. So, among the whole panel and according to this approach, Capron Royal (A), Profumata di Tortona (O), CF2024 (D), and Sengana (Q) can be considered as the highest aromatic varieties, and Marmolada (M), Miranda (N), and Selene (P) are the lowest. This first stage of classification is useful as long as it fits with elementary results of a rapid sensory survey of the corresponding varieties, but of no help for plant breeders for precisely comparing a new variety to those already known.

Nevertheless, because of their biological nature and the variability of the analytical method used, the result homogeneity of the analyzed strawberry samples remains limited. For example, in the case of Cigaline (F) for which 6 SPME measures have been run specifically, the relative standard deviation for each of the 23 constituents ranges between 5% and 75%. In these conditions, none of the chromatograms can be considered as characteristic of a given variety.

Moreover, no constituents have been found to be specific to any particular variety. Methyl anthranilate, which is often associated to the flavor of musk in wild strawberry varieties as mentioned by Ulrich et al., has been yet observed at a high level of concentration in two varieties, A and O (the 2 wild clones), as indicated by histograms in **Figure 3a**. In contrast, DMF appeared largely distributed over most of the varieties, while, at the same scale, DHF has only been found in A and O

(**Figure 3b**). Similarly, methyl hexanoate (**Figure 3c**) was found to be highly concentrated in at least six varieties (B, G, I, L, P, and Q), whereas ethyl hexanoate (**Figure 3d**) was only emerging in Q, and butanoic esters were scarcely found everywhere (**Figure 3e** and **f**). So, the complexity of these diagrams clearly makes their direct reading difficult in terms of classification and comparison of variety aromas.

On the contrary, if the whole matrix (23 columns) of the results obtained from the 70 analyzed samples (70 lines) is introduced into the SOM software described above the discrimination effect observed is strongly dependent on the number of neurons (hexagons) selected for the map. For example, with a 4-neuron map, the two cultivars Capron Royal (A) and Profumata di Tortona (O) clearly come apart because of their high content of methyl anthranilate (**Figure 4a**). All the other samples are distributed into the 3 remaining boxes according to their varieties, except those of Ciloé (I) and Cireine (J) which are found at the rate of 2 samples of the same variety (from the same blending) in 2 of the 3 boxes, alternatively. So, 62 of the 70 samples are already partitioned into 4 clusters, indicating that the varieties coming out in the same hexagon are chemically closer to each other than those appearing in the other hexagons.

In **Figure 4b**, the map has been made out of 12 neurons and the discrimination between varieties is much more efficient. Capron Royal (A) and Profumata di Tortona (O) are still together in a corner, close to CF2024 (D) which also contains small amounts of methyl anthranilate, whereas Sengana (Q), very rich in esters, is rejected to another corner of the map. The four samples of each of the 2 varieties Ciloé (I) and Cireine (J), respectively, are affected to the same hexagon due to the larger panel of differences given by the larger number of hexagons. The six samples of the Cigaline variety (F) are plotted in the same hexagon. One hexagon is empty, indicating that its
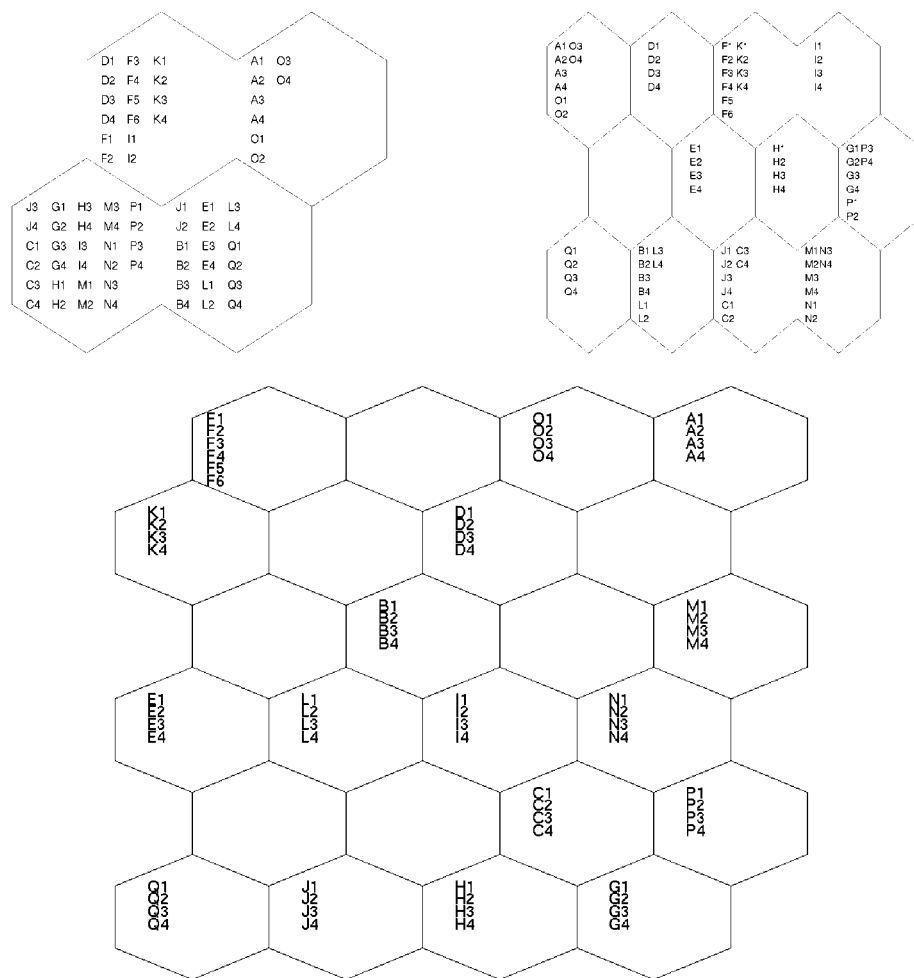
Figure 4. Distribution of the samples in the 4-unit, 12-unit, and 24-unit maps.

virtual constitution is chemically very different from any real samples, and no more than two varieties are sharing the same hexagon.

If the number of hexagons is doubled to 24 (**Figure 4c**), the discrimination effect between the 70 samples is complete and all the samples of the same variety are affected to the same hexagon. This means that among the 24 virtual samples calculated by using the SOM method, the same unit is selected as the BMU each time that a real sample of a defined variety is compared to the whole map. Moreover, Capron Royal (A) and Profumata di Tortona (O) are now fully discriminated but still neighbors of CF2024 (D) in the same corner of the map, and this may be due to the fact that all of them contain methyl anthranilate.

In the same way, it is interesting to note that the varieties Cigaline (F) and Earliglow (K) are isolated in another corner of the map, indicating that they have been chemically assessed close to each other but relatively different from all the other varieties, as they have no immediate neighbors. Relative to the aromatic classification of the 17 varieties, this observation may be correlated with the fact that Cigaline has been obtained from Earliglow by crossing with Garriguette, another typical and well-known variety. In the bottom of the map, Cilady (H) and Cireine (J), which are issued from the same crossing between the same parents, are located side by side, and so are several couples of varieties which are, at least, sharing one parent in the breeding tree.

In the frame of our objective, increasing the size of the map appeared pointless as long as the discrimination of the 70

samples according to the 17 studied varieties was obtained. Nevertheless, using a larger size map would afford a finer discrimination of samples, and this, at the extreme, would result in affecting only one real sample in only some of the hexagons. Anyway, it must be noted that, according to this approach and as the map size increases, the later two individual samples separate, the closer they are relative to the selected chemical aromatic constituents.

To test the efficiency of this type of classifying map, two additional samples were blind-analyzed in duplicate by the same protocol. The four arrays of values obtained (X_1, X_2, Y_1, and Y_2) were input in the same computational program as that previously used for training the SOM as described above. The answer showed the four representative letters to plot in the same unit as that containing the four samples of Cilady (H) and the two unknown samples were effectively of this cultivar. This last result is of great interest because it demonstrates the ability of this method to classify each individual array of characteristic values, whereas most statistical methods used in the field of analytical chemistry necessitate the elaboration of mean values provided from several assays. In fact, the coupled techniques SPME and GC/MS act as providers of these arrays of parameters that can be compared to the signature of a writer and the SOM appears able to visualize this feature for itself and to compare it to others, independently of the exact shape and size of each of the letters used. This particular approach of the chemical constitution of natural matrixes should open large fields of applications for their determination and classification.

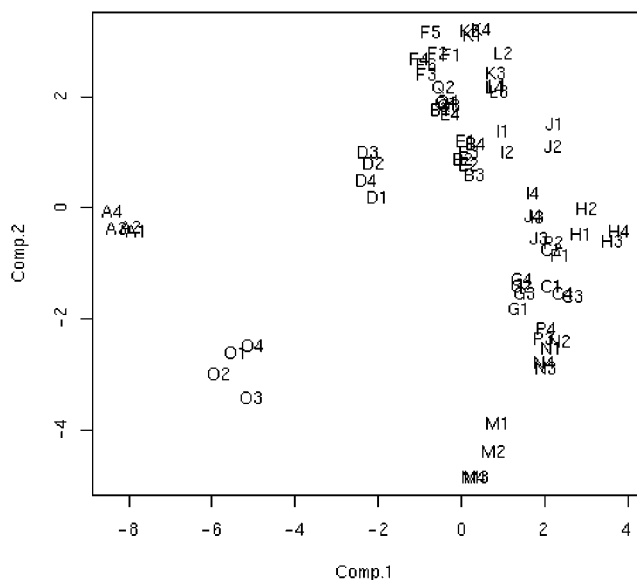The learning vector quantization (LVQ) algorithm (*32*) could

Figure 5. Results of the principal component analysis for the 17 varieties.

have also been used for our study. With LVQ, the self-organizing map is built using a supervised learning. Using the known varieties of the strawberries, LVQ could give a statistical classification, but no spatial order of the virtual strawberries. So, the SOM is preferable for future applications of the method.

To confirm these results, obtained data were submitted to the PCA and ALCA methods cited. For the three methods, some varieties (e.g., A and O) are clearly separated. Moreover, neighbor varieties on the SOM are close both on PCA and dendrogram. So, results afforded by the Kohonen map are validated by these 2 well-known and widely used statistical treatments. On another side, the PCA first and second axis accounted for 33.8 and 19.2% of the total variance, respectively (**Figure 5**). This rather low inertia of the two first axes constitutes itself a limitation of the reliability of the PCA results. Moreover, the "horseshoe" shape of the scatter plot confirms the very poor efficiency of this method in this case. The dendrogram obtained by using the second method (**Figure 6**),

is much more difficult to read than the Kohonen map, but affords some of the major results indicated by this map. For example, the discrimination of the 70 samples according to the 17 varieties is not totally obtained (i.e., J and C) but the relative levels of differentiation are easier to visualize. Nevertheless, this method leads the analyst to compare mean values obtained by averaging all the observed values considered two by two. In these conditions, it is clear that the whole work must be begun again for comparing each new analyzed sample to the preceding ones and this is not adapted to our purpose. Moreover, the apparent proximity of two varieties on the dendrogram is a randomly obtained feature, whereas in the SOMs the proximity of two hexagons is strongly related to the effective chemical proximity of the corresponding virtual units, and therefore, of the real samples that are plotted in there. This complementary point, relative to the significance of map topology, is presently being studied in the authors' laboratory.

This work clearly demonstrates that the HS−SPME/GC−MS technique is an efficient analytical tool to globally and rapidly characterize the chemical nature of strawberry variety aromas, according to a defined selection of constituents. The reliability of the described method is sufficient to allow recognition and classification of measured values by the statistical treatment tool called the Kohonen self-organizing map. One of the main advantages of the whole analysis system is its ability to let the user choose the number of neurons of the map, and then, the degree of differentiation that must control the discrimination between two studied samples. A second advantage is that the resulting map relative to a given panel of varieties will be comparable to another one, made a different year, in different agronomic conditions, or with a different climate, for the same panel. A third advantage is that, once the training of the map is done, the comparison of a new variety to the whole panel can be performed within the few hours necessary for collecting fruits, extracting them by SPME, and chromatographing the adsorbed vapors. Statistical treatment of the results needs only a few minutes. And once this first step of classification is completed, the results have to be correlated to those given by the sensory analysis for the same varieties. This work is also presently in progress in the authors' laboratory.
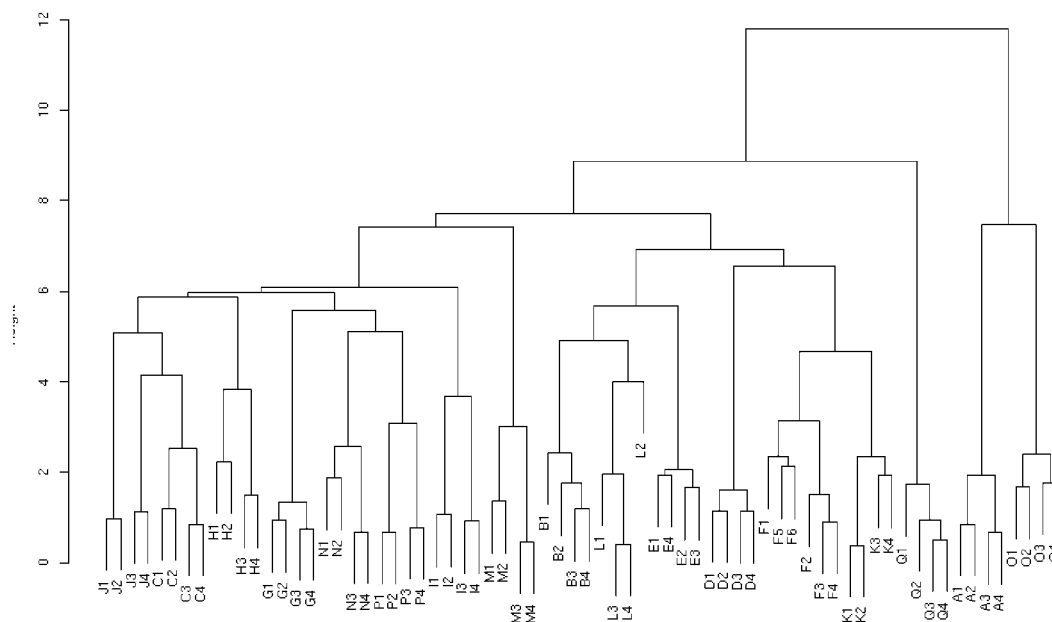

Figure 6. Results of the cluster analysis for the 17 varieties.

3136 *J. Agric. Food Chem.,* Vol. 50, No. 11, 2002

Urruty et al.

LITERATURE CITED

(1) Latrasse, A. Fruits III. In *Volatile Compounds in Foods and Beverages*; Maarse, H., Ed.; Dekker: New York, 1991; pp 333–387.

(2) Honkanen, E.; Hirvi, T. The flavour of berries. In *Food Flavours, Part C: The Flavours of Fruits*; Morton, I., MacLeod, A., Eds; Elsevier: Amsterdam, 1990; pp 123–193.

(3) Douillard, C.; Guichard, E. The aroma of strawberry (*Fragaria ananassa*): characterisation of some cultivars and influence of freezing. *J. Sci. Food Agric.* 1990, *50*, 517–531.

(4) Gomes da Silva, M. D. R.; Chaves das Neves, H. J. Differentiation of strawberry varieties through Purge-and-Trap HRGC-MS, HRGC–FTIR and Principal Component Analysis. *J. High Resolut. Chromatogr.* 1997, *5*, 275–283.

(5) Song, J.; Fan, L.; Beaudry, R. M. Application of Solid-Phase Microextraction and Gas Chromatography/Time-of-Flight Mass Spectrometry for rapid analysis of flavor volatiles in tomato and strawberry fruits. *J. Agric. Food Chem.* 1998, *46*, 3721–3726.

(6) Larsen, M.; Poll, L. Odour thresholds of some important aroma compounds in strawberries. *Z.-Lebensm.-Unters. Forsch.* 1992, *195*, 120–123.

(7) Ulrich, D.; Hoberg, E.; Rapp, A.; Kecke, S. Analysis of strawberry flavour – discrimination of aroma types by quantification of volatile compounds. *Z.- Lebensm.-Unters. Forsch. A* 1997, *205*, 218–223.

(8) Schieberle, P.; Hofman, T. Evaluation of the character impact odorants in fresh strawberry juice by quantitative measurements and sensory studies on model mixtures. *J. Agric. Food Chem.* 1997, *45*, 227–232.

(9) Pawliszyn, J. *Solid-Phase Microextraction: Theory and Practice*; Wiley-VCH: New York, 1997.

(10) *Applications of Solid-Phase Microextraction*; Pawliszyn, J., Ed; The Royal Society of Chemistry: Cambridge, UK, 1999.

(11) Coleman, W. M.; Perfetti, T. A.; Lawrence, B. M. Automatic injection solid-phase microextraction–chiral gas chromatography–mass selective detection analyses of essential oils. *J. Chromatogr. Sci.* 1998, *36*, 575–578.

(12) Stanfill, S. B.; Ashley, D. L. Solid-phase microextraction of alkenylbenzenes and other flavor-related compounds from tobacco for analysis by selected ion monitoring gas chromatography–mass spectrometry. *J. Chromatogr. A* 1999, *858*, 79–89.

(13) Falqui Cao, C.; Wang, Z.; Urruty, L.; Pommier, J. J.; Montury, M. Focused microwave assistance for extracting some pesticide residues from strawberries into water before their determination by SPME/HPLC/DAD. *J. Agric. Food Chem.* 2001, *49*, 5092–5097.

(14) Zhang, Z.; Pawliszyn, J. Headspace Solid-Phase Microextraction. *Anal. Chem.* 1993, *65*, 1843–1852.

(15) Harmon, A. D. Solid-Phase Microextraction for the analysis of flavors. In *Techniques for Analyzing Food Aroma*; Marsili, R., Ed.; Marcel Dekker: New York, 1997; pp 81–112.

(16) Chin, H. W.; Bernhard, R. A.; Rosenberg, M. Solid-phase microextraction for cheese volatile compound analysis. *J. Food Sci.* 1996, *61*, 1118–1122.

(17) Steffen, A.; Pawliszyn, J. Analysis of flavor volatiles using headspace Solid-Phase Microextraction. *J. Agric. Food Chem.* 1996, *44*, 2187–2193.

(18) Bicchi, C. P.; Panero, O. M.; Pellegrino, G. M.; Vanni, A. C. Characterization of roasted coffee and coffee beverages by Solid-Phase Microextraction–Gas Chromatography and Principal Component Analysis. *J. Agric. Food Chem.* 1997, *45*, 4680–4686.

(19) Fitzgerald, G.; James, K. J.; MacNamara, K.; Stack, M. A. Characterisation of whiskeys using solid-phase microextraction with gas chromatography–mass spectrometry. *J. Chromatogr. A* 2000, *896*, 351–359.

(20) Ibáñez, E.; López-Sebastián, S.; Ramos, E.; Tabera, J.; Reglero, G. Analysis of volatile fruit components by headspace solid-phase microextraction. *Food Chem.* 1998, *63*, 281–286.

(21) Ulrich, D.; Hoberg, E.; Rapp, A.; Sandke, G. Flavour analysis in plant breeding – solid-phase microextraction of strawberry aroma compounds; *Proc. 5th Wartburg Aroma Symp.,* Eisenach (Germany), March 17–20, 1997; pp 283–295.

(22) Pérez, A. G.; Sanz, C.; Olías, R.; Ríos, J. J.; Olías, J. M. Aroma quality evaluation of strawberry cultivars in southern Spain. *Acta Hortic.* 1997, *439*, 337–340.

(23) Mochizuki, T.; Noguchi, Y.; Sone, K.; Morishita, M. Aroma components of amphiploid strawberries derived from interspecific hybrids of *Fragaria × Ananassa* and diploid wild species. *Acta Hortic.* 1997, *439*, 75–80.

(24) Larsen, M.; Poll, L.; Olsen, C. E. Evaluation of the aroma composition of some strawberry (*Fragaria ananassa* Duch) cultivars by use of odour threshold values. *Z.-Lebensm.-Unters. Forsch.* 1992, *195*, 120–123.

(25) Lek, S.; Giraudel, J. L.; Guégan, J. F. Neural networks: Algorithms and architectures for ecologists and evolutionary ecologists. In *Artificial Neuronal Networks: Application to ecology and evolution*; Lek, S., Guégan, J. F., Eds; Springer: Berlin, Heidelberg, 2000; pp 3–27.

(26) Goodacre, R.; Neal, M. J.; Kell, D. B. Quantitative analysis of multivariate data using artificial neural networks: a tutorial review and applications to the deconvolution of pyrolysis mass spectra. *Anal. Chem.* 1994, *66*, 1070–1085.

(27) Goodacre, R.; Kaderbhai, N.; McGovern, A. C.; Goodacre, E. A. Chemometric analyses with self-organising feature maps: a worked example of the analysis of cosmetics using Raman spectroscopy. In *Kohonen Maps*; Oja, E., Kaski, S., Eds.; Elsevier: Amsterdam, 1999; pp 335–347.

(28) Goodacre, R.; Pygall, J.; Kell, D. B. Plant seed classification using pyrolysis mass spectrometry with unsupervised learning: the application of auto-associative and Kohonen artificial neural networks. *Chem. Intell. Lab. Sys.* 1996, *34*, 69–83.

(29) Kaartinen, J.; Hiltunen, Y.; Kovanen, P. T.; Ala-Korpela, M. Application of self-organizing maps for the detection and classification of human blood plasma lipoprotein lipid profiles on the basis of ^{1}H NMR spectroscopy data. *NMR Med.* 1998, *11*, 168–176.

(30) Jolliffe, I. T. *Principal Component Analysis*; Springer-Verlag: New York, 1986.

(31) Krzanowski, W. J. *Principles of Multivariate Analysis: a User's Perspective*; Oxford University Press: Oxford, 1988.

(32) Kohonen, T. *Self-Organizing Maps*; Springer: Berlin, 1995.

(33) Pearson, K. On lines and planes of closest fit to a system of points in space. *Philos. Mag.* 1901, *2*, 557–572.

(34) Lance, G. N.; Williams, W. T. A general theory for classificatory sorting strategies. 1. Hierarchical systems. *Comput. J.* 1967, *9*, 373–380.

(35) Giraudel, J. L.; Lek, S. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecol. Modell.* 2001, *146*, 329–339.

(36) Ihaka, R.; Gentleman, R. R. A language for data analysis and graphics. *J. Comput. Graph. Stat.* 1996, *5*, 299–314.

(37) Melssen, W. J.; Smits, J. R. M.; Buydens, L. M. C.; Kateman, G. Using artificial neural networks for solving chemical problems. Part II: Kohonen self-organising feature maps and Hopfield networks. *Chemom. Intell. Lab. Syst.* 1994, *23*, 267–291.

(38) Kaufman, L.; Rousseeuw, P. J. *Finding Groups in Data: an Introduction to Cluster Analysis*; Wiley: New York, 1990.