Original article

Electronic nose and neural network use for the classification of honey

Simona BENEDETTI[a]*, Saverio MANNINO[a], Anna Gloria SABATINI[b], Gian Luigi MARCAZZAN[b]

[a] Department of Food Science and Technologies, Chemistry Unit, via Celoria, 2, 20133 Milano, Italy
[b] Istituto Nazionale di Apicoltura, via di Saliceto, 80, 40128 Bologna, Italy

Abstract – Seventy samples of honey of different geographical and botanical origin were analysed with an electronic nose. The instrument, equipped with 10 Metal Oxide Semiconductor Field Effect Transistors (MOSFET) and 12 Metal Oxide Semiconductor (MOS) sensors, was used to generate a pattern of the volatile compounds present in the honey samples. The sensor responses were evaluated by Principal Component Analysis (PCA) and Artificial Neural Network (ANN). Good results were obtained in the classification of honey samples by using a neural network model based on a multilayer perceptron that learned using a backpropagation algorithm. The methodology is simple, rapid and results suggest that the electronic nose could be a useful tool for the characterisation and control of honey.

honey / electronic nose / classification / neural network analysis

1. INTRODUCTION

The honey production and composition is currently regulated in the European Community by the Council Directive 2001/110/EC and the member states have to follow this directive since 1 August 2003. According to this new directive, additional information about floral and geographical origin of honey should be provided to the consumers.

Several methods have been used for the determination of the floral and geographical origin of honey and they are mainly based on the analysis of its pollen content (melissopalynology), sensory analysis, amino acids, volatile compounds (characterized by solid phase micro extraction (SPME)), carbohydrates, phenolic compounds (flavonoids included), organic acids, and marker presence (Anklam, 1998). But some of these methods are generally too time-consuming, complex, and labour-intensive for routine quality control application or require very specialised personnel to interpret the results (Bogdanov and Martin, 2002). In industrial quality control applications, the need for speed and large number of samples to be tested significantly affects the type of test procedure that can be used.

One analytical tool, proposed in recent years to address the need for routine quality testing in the food industry, is the so-called "electronic nose". This device consists of an array of weakly specific or broad spectrum chemical sensors that intend to mimic the human olfactory system and convert sensor signals to data that can be analysed with appropriate statistical software. Such characteristics greatly facilitate the application of the electronic nose to rapid monitoring of the volatile components of food, providing real-time information about various characteristics of the food under study (Schaller et al., 1999).

---

* Corresponding author: simona.benedetti@unimi.it

A number of potential applications of the electronic nose in food industry have been reported, such as for quality estimation of ground meat (Winquist et al., 1993), detection of boar taint in meat (Bourrounet et al., 1995), detection of gender differences in meat products (Berdague and Talou, 1993), estimation of fish freshness (Scweizer-Berberich et al., 1994. Winquist et al., 1995), evaluation of shelf life of fresh vegetables (Riva et al., 2001), to follow beer production (Pearce et al., 2003) and to predict the degree of mouldy/musty odour in cereal (Borjesson et al., 1996).

Therefore today there is a great interest in using an array of different chemical sensors for identification of volatile compounds (Moore et al., 1993) and different odours (Holmberg et al., 1995). The sensor responses are evaluated by pattern recognition routines based on both statistical methods (Gardner, 1991) and artificial neural network (Gardner et al., 1990).

The aim of our work was to assess the suitability of an electronic nose coupled with appropriate data analysis for discriminating among honey samples of different botanical and geographical origins.

2. MATERIALS AND METHODS

2.1. Apparatus

Analyses were performed with an electronic nose (model 3320 Applied Sensor Lab Emission Analyser; Applied Sensor Co., Linkoping, Sweden) which consisted of three parts: an automatic sampling apparatus, a detector unit containing the array of sensors, and software for pattern recognition. The automatic sampling system supported a carousel of 12 sites for loading the samples under controlled temperature.

The sensor array was composed of 22 different sensors. Ten of the sensors were Metal Oxide Semiconductor Field Effect Transistors (MOSFET) and twelve were Taguchi type sensors (Metal Oxide Semiconductors – MOS). The MOSFET sensors were divided into two arrays of five sensors each, one array operating at 140 °C and the other at 170 °C, while the MOS, mounted in a separate chamber, were kept at 400–500 °C during all the process phases. We chose these sensors because in our experience MOS and MOSFET sensors are most suitable for food head space analysis as they work at high temperatures and thus are not sensitive to humidity, unlike polymer sensors.

2.2. Samples

Seventy different unifloral honey samples of specific botanical and geographical origins were analysed. Fourteen were of Italian origin from *Robinia pseudoacacia* L., 30 from *Rhododendron* spp. and 20 from *Citrus* spp. Six samples were *Robinia pseudoacacia* honey of Hungarian origin.

All samples were provided by the Istituto Nazionale di Apicoltura (Bologna, Italy) and their unifloral authenticity was verified in the accredited laboratory of this institution by pollen, organoleptic, and chemical analyses. Data were in accordance with the unifloral honey characterization established by the Italian Minister of Agriculture (Persano Oddo et al., 2000).

Samples, once received, were stored at 4 °C until analysed and none of them was exposed to any treatment that might alter their composition.

2.3. Operative procedure

Three grams of each sample were placed in 40 mL Pyrex® vials with silicone caps and then introduced inside the automatic sampling carousel of the electronic nose. Preliminary trials indicated that using larger sample volumes did not significantly increase signal intensities and reproducibility. After an equilibration time of 20 min at room temperature, the measurement sequence started with the following temperature phases: standby at 20 °C for 10 min followed by incubation at 40 °C for 5 min.

The measurement procedure consisted of pumping reference air over the sensors (room air filtered through active carbon), at a constant flow rate (1 cm^3/s) for 10 s, in order to have a stable baseline. Then honey gas headspace sampled by an automatic syringe was pumped over the sensor surfaces for 30 s. The sensors were then exposed to reference air to recover the baseline. The total cycle time for each measurement was 5 min. No sensor drift during the measurement period was experienced.

Each sample was evaluated 3 times and the average of the results was used for subsequent analysis.

2.4. Data analysis

The data obtained from the sensor array for the 70 honey samples were analysed by Principal Component Analysis (PCA) performed with STATISTICA (v. 5.1 Statsoft Inc., Tulsa, Oklahoma, USA) and Artificial Neural Network (ANN) using NeuroSolutions software (v. 4.2 Neuro Dimension Inc., Gainesville, Florida, USA).

Principal Components Analysis was used for explorative data analysis as it identifies orthogonal directions of maximum variance in the original data, in decreasing order, and projects the data into a lower-dimensionality space formed of a subset of the highest-variance components. The orthogonal directions are linear combinations (principal components PCs) of the original variables and each component explains in turns a part of the total variance of the data. In particular, the first significant component explains the largest percentage of the total variance, the second one, the second largest percentage, and so forth.

Neural networks learn from examples through iteration, without requiring a priori knowledge of the relationship among variables under investigation. The neural network model used in this study was a multilayer perceptron (MLP) that learns using an algorithm called backpropagation with an adaptive learning rate and momentum of 0.01 and 0.70, respectively. The network was composed by 23 input neurons, a one-hidden-layer net with 4 hidden units in the hidden layer and 4 neurons in the output layer (one for each honey type). Sensor responses of honey samples were repeatedly fed into the net and training was accomplished by automatically adjusting the connection weights in order to reduce the error level.

The ANN training was terminated using the method of stopping with cross validation. Training was then stopped at the point of the smallest error in the validation set. This allowed to avoid overtraining. Sixty percent of the randomised honey data set were used as the training set (42 samples), 15% as the cross validation set (11 samples), and 25% as the testing set (17 samples).

Since a neural network can arrive at different solutions for the same data, if different values of the initial network weights are provided, the network was trained several times. The goal was to try and find a neural network model for which multiple training approaches the same final Mean Squared Error (MSE).

In order to know the effect that each input had on the network output, a sensitivity analysis was performed. This allowed to identify and to eliminate irrelevant inputs, and to produce a smaller network with better generalization.

3. RESULTS AND DISCUSSION

The cumulative variance explained by the first two principal components was 79.3%, with the first component providing 48.0% of the total (Fig. 1A). The first two principal components clearly separated the *Rhododen-dron* honey from the three others. The third component explained 13.2% of the total variance, and the fourth 4.4% (Fig. 1B). The third and fourth component separated the other three types of honey: the *Citrus* honey was located in the right part of the plot; the Hungarian *Robinia* in the upper central part and the Italian *Robinia* in the lower left part of the plot.

Once explored the sensor data by principal component analysis and determined that the four types of honey could be separated, a classification model was performed by the ANN technique.

At the beginning all responses from the electronic nose sensors were used as inputs to the neural network that was built. The relative importance of each input varied considerably as illustrated by the reports of the sensitivity analysis (example shown Fig. 2). Only a few sensors were able to classify the honey samples and these were: 1 MOS and 2 MOSFET sensors (identified by the instrument manufacture as MO113, FE104A, FE104B respectively). With the array dimension reduced, the subsequent neural networks, consisting of three inputs and four outputs, were built and trained. The number of neurons in the hidden layer was changed in order to obtain the best results.

The results of ANN analysis can be visualised by a confusion matrix of the desired and actual network output and by a table showing the general performance indexes of the network. A confusion matrix displays the desired classification on the rows and the predicted classification on the columns. The ideal situation is to have all the exemplars end up on the diagonal cells of the matrix. The confusion matrix in Table I shows perfect classification. Every honey type was correctly classified by the network as *Rhododendron*, Italian *Robinia*, Hungarian *Robinia* and *Citrus*. This is confirmed by the Performance indexes reported in Table II.

However, for each set of samples, we used a two-way binomial test (Engels, 1988) to determine the lowest probability of correct identification that would have given a 95% chance of assigning all samples in this set to the correct category. Thus for the 6 samples of *Rhododendron* honey, the lowest probability
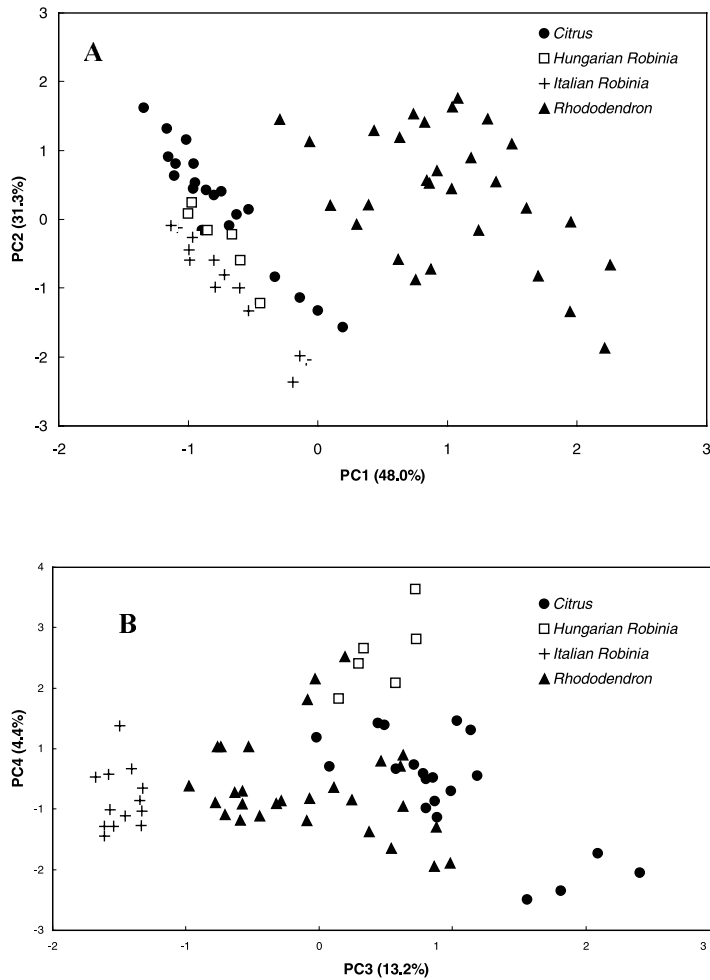
Figure 1. Principal Component Analysis plot of the scores produced by the 22 sensor responses along the first two principal components (A) and the third and fourth principal components (B). The number in parentheses indicate the proportion of the total variance explained by each principal component (PC).
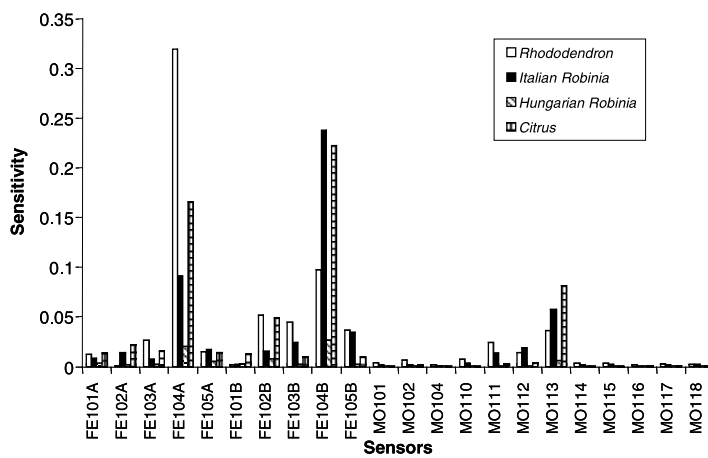

Figure 2. Sensitivity report of the sensor responses for the 17 honey samples. FE stands for MOSFET type sensors and MO stands for MOS type sensors.

Table I. Confusion matrix for the 17 honey samples in the testing set analyzed with the artificial neural network to process the data from 3 sensors.

Sample honey type	Output of the artificial neural network			
	Rhododendron	Italian *Robinia*	Hungarian *Robinia*	*Citrus*
Rhododendron	6	0	0	0
Italian *Robinia*	0	2	0	0
Hungarian *Robinia*	0	0	1	0
Citrus	0	0	0	8

Table II. General performance of the artificial neural network for the 17 honey samples in the testing set analysed based upon processing the data from 3 sensors.

Performance	Honey type			
	Rhododendron	Italian *Robinia*	Hungarian *Robinia*	*Citrus*
MSE[a]	0.001	0.001	0.008	0.007
NMSE[b]	0.006	0.007	0.146	0.029
MAE[c]	0.033	0.021	0.064	0.069
r[d]	0.998	0.998	0.939	0.991
Percent correct[e]	100	100	100	100

[a] Mean square error, [b] normalized mean square error, [c] mean absolute error, [d] correlation coefficient, [e] percent correct performance measure.

of correct identification was 0.616. For the 8 samples of *Citrus* honey, this probability was 0.679. And for all 17 samples, the lowest probability was 0.835 indicating that there was a 95% chance that the probability of correct identification of the ANN used was 0.835 or better.

Volatile and semi-volatile organic compounds present in the headspace aroma contribute significantly to the honey flavour and to its variation in relation to botanical floral origin. The results suggest that the electronic nose could be a useful tool for screening honey volatile compounds and for characterizing the botanical origin. Honey classification was made possible using only five sensor responses and the data from chemical analysis were not necessary to achieve this aim.

Furthermore, the proposed methodology is simple, rapid, and does not require isolation of the volatile components. This makes the technique particularly useful for on line quality control since any alterations that can cause changes in the volatile fraction could in principle be detected. The extension of the methodology to other honey types could represent an original approach to characterize the origin of honey.

Résumé – Nez électronique et utilisation d'un réseau neural pour classer les miels. Le nez électronique offre aux industries alimentaires et de l'emballage une méthode d'analyse chimique rapide pour améliorer l'efficacité de la production et le contrôle de la qualité en classant un mélange complexe et volatil au cours du traitement et du stockage. Le nez électronique consiste en une batterie de détecteurs chimiques à spectre large ou spécifique, une unité qui reçoit les signaux et un logiciel de reconnaissance d'empreintes. La technique est particulièrement utile lors du contrôle de la qualité des matières premières ou des produits alimentaires élaborés ou lors du traitement. Dans cette étude nous donnons les résultats de l'analyse par le nez électronique de miels de quatre origines géographiques et botaniques différentes (miel italien de *Robinia*, miels hongrois de *Robinia*, de *Rhododendron* et de *Citrus*). L'analyse en composantes principales a été utilisée comme autre technique d'étude (Fig. 1).

Puis le réseau neural artificiel a été utilisé pour classer les miels d'après les réponses les plus sensibles des détecteurs (Fig. 2). La matrice de confusion des 17 échantillons tests (Tab. II) et les indices de performance d'erreur (Tab. II) indiquent que les miels sont été classés correctement.

miel / classification / nez électronique / analyse par réseau neural

Zusammenfassung – Eine elektronische Nase und die Nutzung eines neuralen Netzwerks für die Klassifikation von Honig. Die elektronische Nase bietet der Nahrungs- und Verpackungsindustrie eine Methode der schnellen chemischen Analyse zur Verbesserung der Produktionsleistung und der Qualitätskontrolle durch die Klassifizierung komplexer flüchtiger Stoffgemische während der Herstellung und der Lagerung. Sie besteht aus einer Anordnung von Sensoren mit unterschiedlichen Selektivitätsmustern, einer Einheit zum Sammeln der Signale und einer das Muster erkennenden Software. Diese Technik ist besonders nützlich bei der Qualitätskontrolle von Rohstoffen und fertigen Nahrungsmitteln oder während der Verarbeitung. In dieser Arbeit werden die Ergebnisse einer Analyse von 70 Proben 4 verschiedener Honige mit der elektronischen Nase vorgestellt, die sich botanisch und geographisch unterscheiden (italienische *Robinia*, ungarische *Robinia*, *Rhododendron* and *Citrus*). Als weitere Untersuchungstechnik kam PCA zur Anwendung (Abb. 1). Das künstliche neurale Netzwerk (Artificial Neural Network - ANN) wurde angewendet, um die Honige auf der Basis der empfindlichsten Reaktionen der Sensoren zu klassifizieren (Abb. 2). Die "confusion matrix" die beim Test von 17 Proben erhalten wurde (Tab. I) und der "error performance" Index (Tab. II) weisen auf eine gute Klassifikation der untersuchten Honigtypen hin.

Honig / elektronische Nase / Klassifikation / neurale Netzwerk Analyse

REFERENCES

Anklam E. (1998) A review of the analytical methods to determine the geographical and botanical origin of honey, Food Chem. 63, 549–562.

Berdague J.L., Talou T. (1993) Examples of semiconductor gas sensors applied to meat products, Sci. Aliments 13, 141–148.

Bogdanov S., Martin P. (2002) Honey authenticity, Mitt. Geb. Lebensmittelunters Hyg. 93, 232–254.

Borjesson T., Eklov T., Jonsson A., Sundgren H., Schnurer J. (1996) Electronic nose for odor classification of grains, Cereal Chem. 73, 457–461.

Bourrounet B., Talou T., Gaset A. (1995) Application of a multigas-sensor device in the meat industry for boar-taint detection, Sensors Actuat. B 26, 250–254.

Engels W. (1988) Binomial test 1.01. Genetics Department, Univ. Wisconsin, Madison.

Gardner J.W. (1991) Detection of vapours and odours from a multisensor array using pattern recognition. Part 1: principal components and cluster analysis, Sensors Actuat. B 4, 108–116.

Gardner J.W., Hines E.L., Wilkinson M. (1990) The application of artificial neural networks in an electronic nose, Meas. Sci. Technol. 1, 446–451.

Holmberg M., Winquist F., Lundstrom I., Gardner J.W., Hines E.L. (1995) Identification of paper quality using a hybrid electronic nose, Sensors Actuat. B 27, 246–249.

Moore S.W., Gardner J.W., Hines E.L. Gopel W., Weimar U. (1993) A modified multiplayer perceptron model for gas mixture analysis, Sensors Actuat. B 15-16, 344–348.

Pearce T.C., Gardner J.W., Friel S., Barlett P.N., Blair N. (1993) Electronic nose for monitoring the flavour of beers, Analyst 118, 371–377.

Persano Oddo L., Sabatini A.G., Accorti M., Colombo R., Marcazzan G.L., Piana M.L., Piazza M.G., Pulcini P. (2000) I mieli uniflorali italiani, Nuove schede di caratterizzazione, Ministero delle Politiche Agricole e Forestali, Tipografia Benedettucci Carlo.

Riva M., Benedetti S., Mannino S. (2001) Shelf life of fresh cut vegetables as measured by an electronic nose: preliminary study, Ital. J. Food Sci. 13, 201–211.

Schaller E., Bosset J.O., Escher F. (1999) Practical experience with the "electronic nose" systems for monitoring the quality of dairy products, Chimia 53, 98–102.

Schweizer-Berberich P.M., Vaihinger S., Gopel W. (1994) Characterisation of food freshness with sensor array, Sensors Actuat. B 18, 282–290.

Winquist F., Hornsten E.G., Sundgren H., Lundstrom I. (1993) Performance of an electronic nose for quality estimation of ground meat, Meat Sci. Technol. 4, 1493–1500.

Winquist F., Sundgren H., Lundstrom I. (1995) A practical use of electronic nose: quality estimation of cod fillet bought over the counter, in: Digest of Technical Papers, 8th Int. Conf. on Solid-State Sensors and Actuators and Eurosensors IX (vol. 1). Royal Swedish Academy of Engineering Sciences, IVA, Stockholm, Sweden 695–698.