

Fuzzy k -NN applied to moulds detection

M. Kuske^a, R. Rubio^b, A.C. Romain^a, J. Nicolas^a, S. Marco^b

^a *Surveillance de l'Environnement, Fondation Universitaire Luxembourgeoise, Avenue de Longwy, 185, B-6700 Arlon, Belgium*

^b *Sistemes d'Instrumentació i Comunicacions, Departament d'Electrònica, Universitat de Barcelona, Martí Franquès 1, 08028 Barcelona, Spain*

Abstract

The possibility to detect *Aspergillus versicolor* growing on different building materials by a metal oxide sensor array is studied. Results show that an accurate classification rate of $89 \pm 3\%$ can be obtained combining an extended linear discriminant analysis plus a fuzzy k -NN classifier. The classification ability of the classifier is assessed within the dataset by cross validation and also in a second dataset collected 5 months later. There is a slight decrease in the classification performance for all the algorithms, being the most sensitive the most accurate one.

Keywords: Fuzzy k -NN; Linear discriminant analysis; Moulds; Chemical sensor arrays; k -NN

1. INTRODUCTION

Excessive humidity and subsequent fungal development are one of the most frequent problems in buildings (Fig. 1). Moulds can provoke allergies, infections, toxic effects, and different symptoms characteristic for the "sick building syndrome". A relationship between the dampness and/or fungal contamination and those symptoms has been documented in many studies^[1-8].

Traditionally, fungal contamination in a building has been described as quantity of viable fungi determined from air, settled dust, surface and building material samples. The most important advantage of the method is the possibility to identify moulds. On the other hand, only viable spores can be detected, and the results are obtained only after several days. Other, more rapid and more universal methods have been investigated recently. They include determination of levels of fungal components and products such as ergosterol, (1-3)-beta-glucan, mycotoxins and microbial volatile organic compounds (MVOCs). MVOCs detection seems particularly interesting because the compounds can penetrate barriers not penetrable by spores, thus they can facilitate the detection of hidden moulds.

Several laboratories in the world use gas chromatography-mass spectrometry to detect volatile compounds produced by moulds. Electronic nose technology seems very interesting because, in difference with gas chromatography, the method is simpler, cheaper, and the results can be obtained in situ.

1.1. Microbial volatile organic compounds

Moulds produce a wide range of volatile organic compounds. Alcohols, ketones, terpenes, esters and sulphur compounds are the prevailing ones^[9-11]. The volatile production depends on the species, the substratum on which the fungi grow, and the environmental conditions^[10-17].

Various VOCs are products of microbial activity, however no single VOC seems to be a reliable indicator of biocontamination in building materials^[18]. Some of the volatile compounds are detected more often than others, and these compounds could eventually indicate fungal contamination in buildings. In fact, VOC analysis by GC/SM in buildings with a history of sick building syndrome symptoms and microbial growth revealed unusual levels of several MVOCs^[19, 20].

Regarding the great diversity of microbial volatile production and its dependency on many factors, fungal detection cannot be based on the presence of a single substance but on the coexistence of several relevant

compounds. These compounds, most often reported in different studies, are alcohols, ketones, terpenes, furanes, and sulphur compounds (Table 1).

1.2. Electronic nose for fungal detection

Up to now, the detection of microbial volatile organic compounds was carried out by sampling into carbon-based or TENAX adsorbents, and analysing by gas chromatography and mass spectrometry. This method, very specific and sensitive, requires however experience and special laboratory equipment, takes time and is relatively expensive. In the search for rapid and simple methods to detect microbial contamination electronic noses could be an alternative.

Fig. 1: Moulds growing under the wallpaper.



In fact, electronic noses seem to be particularly useful in cereal grains qualification, where the information on absence or presence of fungi and mycotoxins is required immediately. Several recent studies showed that sensor arrays and different techniques of statistical analysis should allow detecting and classifying selected fungal species growing on cereals^[24, 31-34]. Electronic noses used in these studies were able to differentiate between mycotoxigenic and non-mycotoxigenic strains^[33, 24], and to classify grains according to the degree of contamination^[31, 24].

Table 1: Microbial volatile organic compounds frequently reported in the literature

Chemical group	Compound	Reference
Alcohols	2-Methyl-1-propanol	[9-11, 16, 19, 21, 22, 26, 27, 30]
	3-Methyl-1-butanol	[9-11, 16, 18, 19, 22, 26-30]
	2-Methyl-1-butanol	[9, 10, 27]
	3-Octanol	[10, 19, 23, 27]
	1-Octen-3-ol	[9, 10, 18, 19, 21, 22, 25-29]
	1-Hexanol	[18, 16, 11]
	1-Pentanol	[18, 22]
	2-Methyl-isoborneol	[9,19]
	Geosmin	[16, 19, 21, 26, 27]
	Terpenes	Limonene
Pinene		[9, 16, 25]
Sesquiterpenes		[27, 30]
Ketones	3-Octanone	[16, 18, 19, 22, 24, 26-28]
	2-Heptanone	[11, 16, 19, 21, 30]
	2-Pentanone	[11, 16, 30]
Furanes	3-Methylfuran	[9, 19, 27]
Sulphur compounds	Dimethyl disulphide	[21, 27]

Schiffman et al. [35] studied the ability of electronic noses to detect moulds inside buildings. An array of 15 metal oxide sensors was capable of discriminating among the fungi with up to 96% accuracy. It was also able to recognise selected five volatile organic compounds that are emitted by fungi (ethanol, 2-methyl-1-propanol, 3-pentanone, 3-octanol and 3-octanone). The commercial e-nose MOSES II was successfully used to evaluate perceived indoor air quality. In the experiment with three mould species, it permitted to determinate that a mould is present, but it did not permit to differentiate among different species. An e-nose used by Persaud et al. [36] could discriminate between wood infected by *Serpula lacrimans* and uninfected samples.

1.3. Objectives of the study

The objective of this study is to test performances of an e-nose to detect the presence of moulds growing on different building materials. In the first stage, we are interested in only detection of mould presence. A side objective is to test the capabilities of fuzzy k -NN methods compared to conventional k -NN classifiers.

2. MATERIALS AND METHODS

2.1. Construction materials

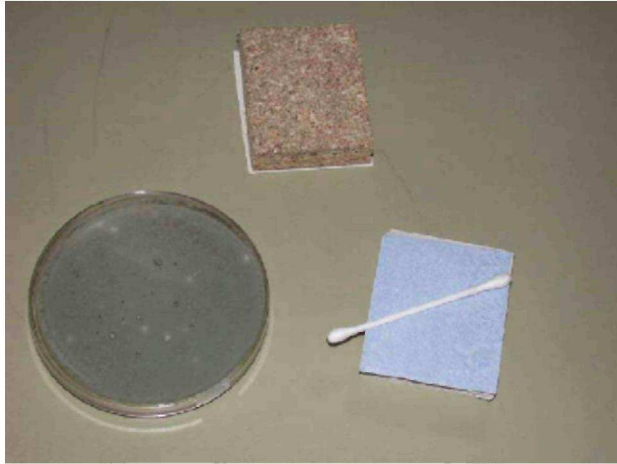
Five materials typical for Belgian houses were used as support to cultivate moulds: plasterboard, particle board, oriented strand board (OSB), wallpaper and glue. During the first stage of the study, 11 combinations of these materials were analysed (Fig. 2):

- plasterboard (PB);
- particle board or chip board (CB);
- oriented strain board (OSB);
- wall paper (WP);
- PB + WP;
- CB + WP;
- OSB + WP;
- PB + glue;
- CB + glue;
- OSB + glue;
- WP + glue.

Fig. 2: Pictures of the growing substrates (from upper left to lower right): wallpaper (WP), chipboard (CB), CB + WP, oriented strain board (OSB), OSB + WP, plasterboard (PB), PB + WP.



Fig. 3: Moulds growing in the culture media, pieces of substrates: chipboard and wallpaper and the stick used to transfer the moulds from the culture to the substrates.



Prior to contamination, material samples were stabilised during at least 24 h at elevated humidity. The samples were not sterilised.

2.2. Preparation of moulds

Four fungal species commonly found in indoor environments, *Aspergillus versicolor*, *Penicillium aurantiogriseum*, *Penicillium chrysogenum* and *Cladosporium sphaerospermum*, were separately incubated on culture medium (Malt Extract Agar, MEA) during 1 week. After this period, moulds were inoculated on samples of building materials (Fig. 3). However, in this paper we only report on the results for *A. versicolor*.

2.3. Cultivation and sampling equipment

Contaminated samples were incubated in air-tight glass jars (volume 500ml). The jars were closed with a cover in Teflon. Each jar contained two pieces of the same material of dimensions 5cm x 6cm contaminated with one fungal species. Forty millilitres of demineralized water was placed in the bottom of the jar. The samples were placed on top of small glass vials to avoid direct contact with water (Fig. 4). The jars were kept at room temperature (20-25°C) in the darkness.

Fig. 4: Moulds growing in plasterboard stored in glass jars with a water bed and a Teflon lid.



2.4. Electronic nose

An electronic nose instrument was designed and constructed at Fondation Universitaire Luxembourgeoise. It consists of a sampling unit, a sensor array, and a signal processing system (Figs. 5 and 6). Clean dry air at a 1.5 bar is fed to a mass flow controller, a humidifier (bubbling through demineralized water at room temperature), a

jar with samples, and three valves to direct the air flow. The sensor array is composed of 12 metal oxide semiconductor sensors that were selected in order to obtain different responses for volatile compounds produced by moulds. Six of the sensors are manufactured by Capteur, and six others by Figaro (see Fig. 5 for the exact references). They are fixed on separate electronic cards, and placed in two separate chambers placed in parallel. Each sensor chamber has a volume of 210 cm³.

2.5. Sampling procedure

With the system described above, the carrier gas was vapour saturated synthetic air at room temperature. A sample cycle included a passage of clean humidified air (reference air), followed by a passage of air from the samples. The air flow was constant, 200 ml/min; sampling time was 3 min (time of signal stabilisation). Recovery time for the next measurement was 50 min.

2.6. Measurement sequence

The experiment has taken place for about 8 months. During this time two experiments have been performed, resulting in two datasets. The contents of both datasets are summarized in Table 2. Note that because of time constraints the presence of glue was not considered in the second experiment. At the first dataset, clean and contaminated materials were intercalated as follows: clean materials were measured at the beginning of the experiment before being inoculated. A second set of clean materials were measured 1 month later. A third set of clean materials were measured 5 months later. *A. versicolor* was grown for 123 days. At the second dataset, *A. versicolor* was grown for 101 days. In this second dataset clean materials, were measured again at the beginning of the experiment, a second set 1 month later, a third set 1.5 months later, and a fourth set 2 months later. At every measurement session, the presentation of the materials were randomised. In summary, the first dataset consisted of 98 clean materials, 94 contaminated materials for a total of 192 samples. In the second dataset there were 55 clean samples, 49 contaminated samples for a total of 104 samples.

Fig. 5: Schema of the instrument used in the study.

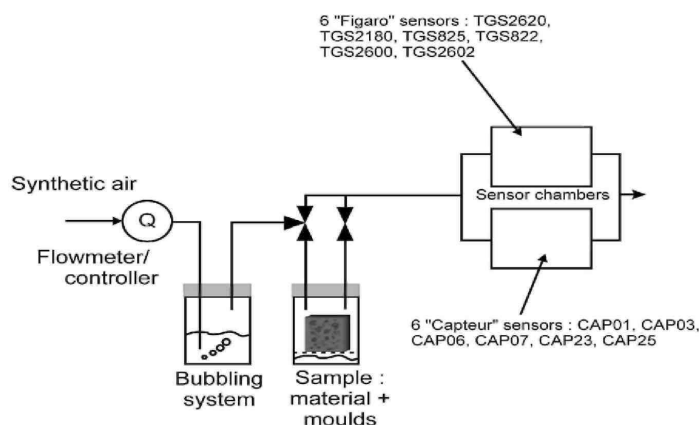


Fig. 6: Picture of the actual experimental setup showing the bubbler, the jar under analysis, the mass flow controller and the sensor chambers.



Table 2: Materials and number of samples contained in the first and second datasets

Material	First dataset	Second dataset
Oriented strain board (OSB)	21	14
OSB + wallpaper	19	16
OSB + glue	13	0
Plaster board (PB)	21	14
PB + wallpaper	19	16
PB + glue	13	0
Chipboard (CB)	21	14
CB + WP	19	15
CB + glue	12	0
WP + glue	15	0
WP +g	19	15

3. Fuzzy k -NN classifiers

The detection of moulds can be considered as a binary classification problem. However, first visual inspection of the patterns show that the signal level is very low. Moreover, due to the presence of different growing substrates, classes can be multimodal. k -nearest neighbours' decision rule, introduced by Fix and Hodges^[37] is often implemented for this kind of applications. Its computational simplicity, its well known asymptotic behaviour (bounded by twice the optimal Bayesian error), the lack of assumptions regarding the underlying probability distribution function for the classes and the good results obtained in small problems still make this algorithm widely used.

However, a drawback of the voting k -NN rule is that it implicitly assumes the k -nearest neighbours of the input vector, to be contained in a relatively small region, so good resolution can be obtained in the estimates of the conditional densities for every class. For small training datasets this is not usually the case, and the distance between the data point and the neighbours can become quite large outside the regions of maximum density. As a first thought, it becomes questionable if all the neighbours have to be given equal weight in the decision despite the distances to the point being classified. Although several weighted k -NN rules have been proposed^[38], numerical experiments have shown that the result is not always better than a simple k -NN^[39-41]. Moreover, the asymptotic behaviour of the conventional k -NN is better than the weighted counterpart. This means that for large datasets, the use of weighted k -NN versions brings no clear advantages. This seems not to be the case for small datasets, according to MacLeod et al.^[42], as those typical of electronic noses, where large datasets are the exception not the rule.

On the other hand, in cases where not all the classes can be assumed to be in the training set, this is typical of field analysis conditions where the matrix can contain analytes not present in the training set. In such situations, we consider wise to consider that a point far away from the training dataset most probably contains new analytes and do not permit to assign a class label on the basis of the information gathered in the training experiments. A hard option to implement such a concept is to introduce a *distance reject* as introduced by Dubuisson and Masson^[43]. This distance reject implies that any object too far away is classified as unknown. This distance reject can be learnt from the training set. However, at this point the wealth of techniques of outlier detection can be implemented as add-ons to the basic k -NN rule, either weighted or not.

Another problem encountered using the k -NN classifier is the lack of information about the "typicalness" of the samples used to label the input space. It may be argued that atypical vectors should be given less weight in the decision than those that are truly representative of the classes. Due to the crisp character of the input membership given to make the classification, no information about how the data is distributed in the input space is provided (apart from the samples themselves). For instance outliers are given full membership to a particular class. Another difficulty found is that the algorithm, once a input vector is classified, does not give information about the "strength" of membership to that class.

Fuzzy set theory provides a global framework to process information that is only partially addressed by other methods, namely, the consideration of the distances of the neighbours in the decision, the possibility to reject the classification of an incoming input pattern and the consideration of the "typicalness" of the training data for the different classes. In this sense, Keller et al. proposed in 1985^[44] a fuzzy version of the k -NN rule: the so called fuzzy k -NN algorithm.

The fuzzy set theory is introduced both, in the samples used to the classification and in the output given class. Concerning the samples, they are no longer described by a crisp label but with a fuzzy membership degree for each class:

$$u_{ij} \in [0, 1], \quad \sum_{i=1}^M u_{ij} \leq 1, \quad (1)$$

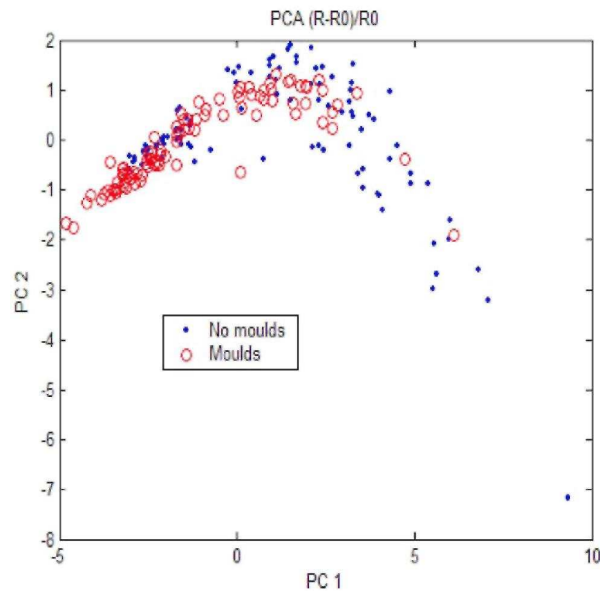
where u_{ij} is the membership value of training vector j to class i and M the number of classes. This membership assignment is a fuzzy partition of the training dataset. If the equality condition holds, the partition will be a strong fuzzy partition.

How this membership is determined is not specified in the original algorithm and several proposals have been reported in the literature. For instance, Kim et al. ^[45] propose a membership assignment that is based in the initial class label but also on the k -neighbours of every training vector. Basically, half the membership value is provided by the initial knowledge about the class labels, on the other half is provided by the local environment of the sample. In another proposal, Denoeux computes the initial membership based on the Dempster-Shafer theory ^[46]. More recently Marcelloni ^[47], has proposed to assign the training vector memberships using the supervised fuzzy c-means algorithm, originally proposed by Pedrycz and Waletzky ^[48]. On the other hand, the algorithm provides a membership degree for the input vector. This allows the user to make the last decision using this fuzzy information or to introduce thresholds of confidence level. Following Keller's fuzzy k -NN ^[44] the proposed output assignment rule of sample x to cluster i is given by

$$u_i(x) = \frac{\sum_{j=1}^K u_{ij} (1/||x - x_j||^{2/(m-1)})}{\sum_{j=1}^K (1/||x - x_j||^{2/(m-1)})}, \quad (2)$$

where x_j are the k -nearest neighbours and u_{ij} the membership of training vector j to class i . It can be observed that this formula do not only takes into account the membership of the training data but also incorporates a weight depending on the distance to the sample pattern. The exponent m takes usually the value of 2. In this case the weight is inversely proportional to the distance.

Fig. 7: Data distribution in the first two components (explained variance 85%): dots, clean samples; circles, mould polluted samples. Dataset 1.



Although better performance, compared with classic k -NN, of this algorithm is claimed in the literature, some a priori difficulties can be encountered. A membership function for the training samples needs to be defined. This can introduce some difficulties for small sample training datasets, especially if the input space dimensionality is high. When using clustering algorithms the optimal number of clusters has to be determined. This implies the use of a figure of merit for cluster quality ^[49]. In addition, if the membership values correspond to a hard fuzzy partition (e.g. using fuzzy c-means as clustering algorithm), sometimes outlier samples are given significant membership to classes that originally they do not belong to. Under these circumstances the algorithm does not become robust against outliers. Finally, the fuzzy output makes necessary the introduction of a classification

criterion, for instance:

$$h = \arg \max_i (u_{ij}), \quad (3)$$

where h is the given label and u_{ij} the membership vector for sample j .

4. DATA PROCESSING AND RESULTS

For every sample under analysis the component l of the pattern vector was formed as follows:

$$x_l = \frac{R_l - R_{l,0}}{R_{l,0}}, \quad (4)$$

where R_l is the conductance of sensor l in the array at the end of the sampling period (3 min), and $R_{l,0}$ the conductance of sensor l just before sampling when exposed to the vapour saturated carrier gas. It was observed that this simple preprocessing removed to a great extent the observed drift. In this way a 12-dimensional pattern vector was formed.

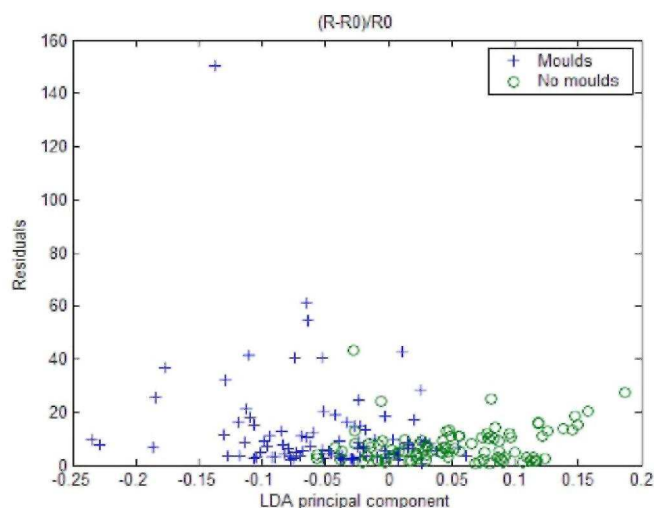
All the data processing algorithm development has been carried out using the first dataset. The second dataset has been reserved for evaluation.

Although k -NN classifiers can be applied directly, we introduced at this point a dimensionality reduction step: unsupervised principal component analysis (PCA) and supervised linear discriminant analysis (LDA).

A first look to the data is the scores plot in the first and second principal components (see Fig. 7). This projection explained 85% of the total variance. From this plot we can observe that the clusters greatly overlap, do not seem to be linearly separable and do not seem to be gaussian. Despite this some degree of separation can be appreciated between both classes.

In order to determine the optimal dimensionality for PCA projection, crossvalidation was used. Dataset 1 was subjected to 10 times random subsampling a subset with a 50-50% split. The maximum explained variance in cross-validation turned out to be in five dimensions, where a 98% of the total variance was explained.

Fig. 8: LDA projection in validation: some separation has been obtained but a great overlap still remains.



Regarding the use of LDA for binary classification a simple difficulty arises. For a c class problem the maximum LDA dimension is $c - 1$. Then for a binary problem, the maximum LDA dimension is restricted to one. For visualization purposes, we have used a two-dimensional plot where the LDA factor is in the x-axis, while the residuals are in the y-axis. In Fig. 8, LDA separation in validation is observed. Although there is still a great deal of overlap, some separation is consistent.

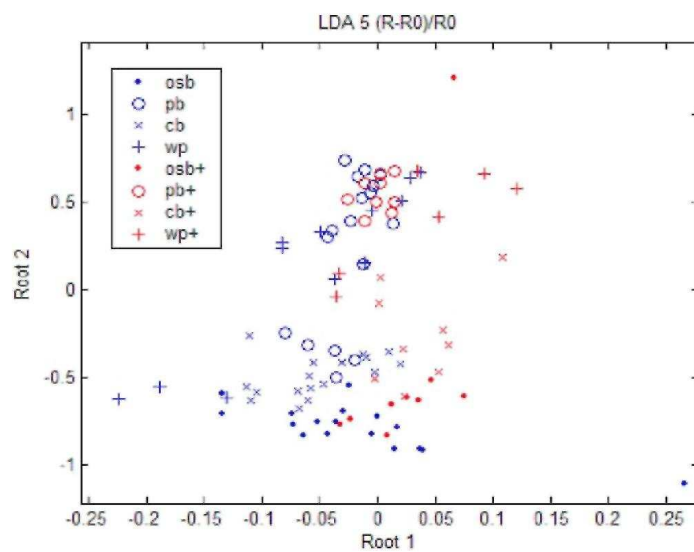
A possibility to increase the number of possible dimensions in an LDA analysis is to introduce additional divisions in the original classes. Here in this case, we further divided every class (moulds, no-moulds) in four classes using the growing substrate as additional label. In this way, we generated a synthetic problem with eight

classes. It is clear that in this case, we will have a maximum of seven dimensions. The distribution in the first two factors can be observed in Fig. 9. Again some separation can be observed but overlap remains. The final discriminative power will be assessed when implementing the classifiers.

Regarding the implementation of the classifiers, while k -NN is straightforward, the implementation of fuzzy k -NN requires some decisions regarding how to compute the fuzzy partition of the training set. As introduced before several options have been proposed in the literature. Several options have been explored by the authors including the proposal of Kim et al.. Finally best results were obtained with a simplified version where each class (or pseudo-class) was modelled by a spherical gaussian. In the case each class, consist of four pseudo-classes (for the sake of augmenting the LDA dimension), the final fuzzy set and its membership function is computed using the fuzzy AND operator for combining the four fuzzy sets corresponding to the growing materials. Among the different options for the fuzzy AND operator, here we have chosen the maximum:

$$(\mu_A \cup \mu_B)(x) \equiv \max(\mu_A(x), \mu_B(x)). \quad (5)$$

Fig. 9: Data distribution in validation in the first two factors for the synthetic problem with eight classes: red symbols are polluted samples, blue symbols are clean samples. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.



Before commenting the final classification results of the k -NN and fuzzy k -NN several points need further details. First point is the final classification rule. For this study the proposal of Eq. (3) was used. We have to remark that in this way, all the points are classified either as moulds or no-moulds and no ambiguous answer is permitted. The advantages of an ambiguous answer has not been addressed in this work. However, this question requires probably the evaluation of the risk or cost associated to false positive and false negatives.

Second point to remark is the selection of the optimal k . This again was selected by crossvalidation within dataset 1. The optimal result was $k = 3$. It was noted that the dependence on k was in general weaker for fuzzy k -NN classifiers. Probably this is due to the weights given depends on the distance. Too far away training points do not contribute too much to the classifier output.

Let us first evaluate both classifiers working either on the raw input space or in a projection to the power optimal five dimensions or just to two dimensions. Results can be examined in Table 3. First thing to note are remarkable classification rates about 80% for both algorithms. No significant difference is observed between conventional k -NN and the Fuzzy version. The best results are obtained when working directly on the input space or in the projection to five dimensions. Classification rate degrades significantly if the classifier considers only the first two principal components. At this point our interpretation is that for this data distribution, nor the method to compute the training set fuzzy partition or the presence of weights do not bring advantages in terms of final classification rates. While the second point has already been raised in some studies in the literature, we can also question the method chosen to model each class since visual inspection of the bidimensional PCA plots does not show gaussian classes.

Table 3: Estimation of the classification rate and its variance using 10 times random subsampling in dataset 1

Method	Raw	PCA 2 dim	PCA 5 dim
k -NN ($k = 3$)	81 ± 2	70 ± 3	79 ± 4
Fuzzy k -NN ($k = 3$)	79 ± 1	68 ± 3	78 ± 2

Effect of the dimensionality reduction by PCA.

Table 4: Estimation of the classification rate and its variance by 10 times random subsampling in dataset 1

Method	LDA (2 classes) (%)	LDA (8 p-classes) (%)
k -NN ($k = 3$)	74 ± 14	82 ± 4
Fuzzy k -NN ($k = 3$, 8 prototypes)	71 ± 8	89 ± 3

Effect of the dimensionality reduction by LDA.

The second option explored in this work has been the evaluation of the classifiers on an LDA projection. Results can be observed in Table 4. First point that clearly appears is that the performance of both classifiers degrades clearly when we project the data in a single dimension. However, a clear difference appears when we project the data in an LDA subspace of five dimensions (optimal value computed using crossvalidation). While the performance of the conventional k -NN recovers the values found before in Table 3, or even slightly better 82 ± 4 , the performance of fuzzy k -NN submits to 89 ± 3 . Due to the computed variance values, this improvement seems to be significant. An important difference to explain this difference is the modelling of every class by the union of four fuzzy sets corresponding to the different growing substrates. It can be argued that the lack of gaussianity observed in the PCA plots originates from the presence of different growing materials. Although probably the pseudo-classes are not perfect spherical gaussians, a better improvement of the description of the input data seems to be accomplished. Other attempts to increase the complexity of this behaviour, for instance using two gaussians per pseudo-class failed to provide good results. This can be understood in the terms of the complexity control taking into account the reduced set of data available.

Finally, it is important to asses the stability of the classifier for future classification of unknown samples. For this purpose we have reserved a the second dataset: an experiment performed in the same conditions as the first one. This second experiment started 5 months after the first experiment and was carried out during more than 3 months.

In order to evaluate this stability we have plotted the second dataset in first and second principal axis extracted from the first dataset (Fig. 10). Basically the structure of the data is pretty similar, however an slight drift can be observed. The reason of this drift can be attributed to sensor ageing or just changes of the average room temperature.

Fig. 10: Projection of dataset 2 (new data) on the principal axis of dataset 1 (old data).

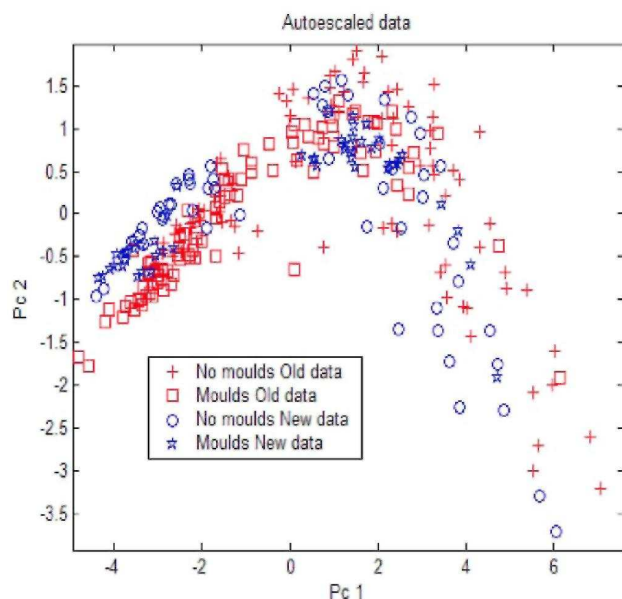


Table 5: Stability of the classifier

	PCA (5 dim)	LDA (S classes)
<i>k</i> -NN	79 ± 4	82 ± 4
	75 ± 4	74 ± 3
Fuzzy <i>k</i> -NN	78 ± 2	89 ± 3
	73 ± 3	78 ± 3

On white background results for the dataset 1. On grey background results for dataset 2.

The classifiers designed with the first dataset have been applied to this second dataset. The results are compared in Table 5. It can be observed how the classification ability decreases in all algorithms. This remark is noticeable although not drastic taking into account that no recalibration was performed. The most important decrease, however, appears in the most accurate classifier in the first dataset. It seems to be clear, that the fuzzy partition of the input space obtained from the first dataset is no longer representative in the second dataset, being possibly this the origin of the decrease in classification ability. This result could be expected, and from our point of view do not contradicts the results obtained in the first dataset. It just shows that accurate modelling of the training data distribution suffers under drift conditions. It seems then, that in order to keep high classification results, additional recalibration or drift counteraction is needed.

5. CONCLUSION

In the presented laboratory conditions the presence of *A. versicolor* can be detected with a high level of accuracy: 89 ± 3%, using a fuzzy *k*-NN classifier, despite the use of a high variety of growing substrates. The best conventional *k*-NN classifier gave slightly worse results with a 82 ± 4%. The highest accuracy was obtained by projecting the patterns using LDA with eight pseudo-classes, obtained as the result of the binary presence of moulds in four different kinds of substrates. The classifier used was a fuzzy *k*-NN classifier, that used a fuzzy input partition obtained modelling each class with a spherical gaussian membership function. The final binary fuzzy partition of the training dataset was obtained by the fuzzy AND of the fuzzy sets of the individual substrates for each class.

The obtained classifiers were quite stable when its performance was measured in a second dataset measured 5 months later. However, it was observed that even if the LDA + fuzzy *k*-NN classifier still gave the best results in this verification set (78 ± 3% C.R.), it was the classifier that showed a higher decrease in the classification rate. This is possibly due to the more accurate modelling of the input fuzzy partition that was not stable under the drift.

Acknowledgements

Spanish team has been funded by Cycit DPI2001-3213-C02-01 and Generalitat de Catalunya ACI2001-11. Belgian team has been funded by Region Wallone. One of the authors RR also acknowledges a PhD grant from the Spanish Ministry of Science and Technology. The assistance of CERTECH is also acknowledged.

References

- [1] B. Brunekreff, A.P. Verhoeff, R.T. van Strien, J.H. Wijnen, Home dampness and childhood respiratory symptoms: the role of sensitization to dust mites and moulds, in: Samson, et al. (Eds.), *Health Implications of Fungi in Indoor Environments*, 1994, pp. 189-199.
- [2] E. Johanning, R. Biagini, D. Hull, P. Morey, B. Jarvis, P. Landsbergis, Health and immunology study following exposure to toxigenic fungi (*Stachybotrys chartarum*) in a water-damaged office environment, *Int. Arch. Occup. Environ. Health* 68 (1996) 207-218.
- [3] K. Klanova, The concentrations of mixed populations of fungi in indoor air: rooms with and without mould problems rooms with and without health complaints, *Central Eur. J. Publ. Health* 8 (2000) 59-61.
- [4] O. Koskinen, T. Husman, A. Hyvärinen, T. Reponen, A. Nevalainen, Two moldy day-care centers: a follow-up study of respiratory symptoms and infections, *Indoor Air* 7 (1997) 67-76.
- [5] D. Norback, G. Wieslander, K. Nordstrom, R. Walinder, Asthma symptoms in relation to measured building dampness in upper concrete floor construction, and 2-ethyl-1-hexanol in indoor air, *Int. J. Tuberc. Lung Dis.* 4 (2000) 1016-1025.
- [6] R. Walinder, D. Norback, G. Johanson, Pulmonary reactions after exposure to 3-methylfuran vapour, a fungal metabolite, *Int. J. Tuberc. Lung Dis.* 2 (1998) 1037-1039.

- [7] G.H. Wang, C.S. Li, Indoor endotoxin and glucan in association with airway inflammation and systemic symptoms, *Arch. Environ. Health* 54 (1999) 172-179.
- [8] CE. Wilson, Sudden infant death syndrome and Canadian aboriginals: bacteria and infections, *FEMS Immunol. Med. Microbiol.* 25 (1999) 221-226. [9] G. Fischer, T. Muller, R. Schwalbe, R. Ostrowski, W. Dott, Exposure to airborne fungi, MVOC and mycotoxins in biowaste-handling facilities, *Int. J. Hygiene Environ. Health* 203 (2000) 97-104.
- [10] H. Kiviranta, A. Tuomainen, M. Reiman, S. Laitinen, J. Liesivuori, A. Nevalainen, Qualitative identification of volatile metabolites from two fungi and three bacteria species cultivated on two media, *Central Eur. J. Publ. Health* 6 (1998) 296-299.
- [11] A.L. Sunesson, C.A. Nilsson, B. Andersson, G. Blomquist, Volatile metabolites produced by two fungal species cultivated on building materials, *Ann. Occup. Hygiene* 40 (1996) 397-410.
- [12] J. Bjurman, J. Kristensson, Volatile production by *Aspergillus versicolor* as a possible cause of odor in houses affected by fungi, *Mycopathologia* 118 (1992) 173-178.
- [13] T. Borjesson, U. Stollman, J. Schnurer, Volatile metabolites and other indicators of *Penicillium aurantiogriseum* growth on different substrates, *Appl. Environ. Microbiol.* 56 (1990) 3705-3710.
- [14] K. Fiedler, E. Schutz, S. Geh, Detection of microbial volatile organic compounds (MVOCs) produced by moulds on various materials, *Int. J. Hygiene Environ. Health* 204 (2001) 2-3.
- [15] W. Horner, A. Worthan, B. Epstein, M. Black, MVOC Emission Patterns of Indoor Fungi: Strain and Culture Differences, 1999. <http://www.aiha.org/abs99/9pos3.html>.
- [16] A. Korpi, A.-L. Pasanen, H. Viitanen, Volatile metabolites of *Serpula lacrymans*, *Coniophora puteana*, *Poria placenta*, *Stachybotrys chartarum* and *Chaetomium globosum*, *Build. Environ.* 34 (1999) 205-211.
- [17] L.J. Rose, R.B. Simmons, S.A. Crow, D.G. Ahearn, Volatile organic compounds associated with microbial growth in automobile air conditioning systems, *Curr. Microbiol.* 41 (2000) 206-209.
- [18] A. Korpi, A.-L. Pasanen, P. Pasanen, Volatile compounds originating from mixed microbial cultures on building materials under various humidity conditions, *Appl. Environ. Microbiol.* 64 (1998) 2914-2919.
- [19] G. Strom, J. West, B. Wessen, U. Palmgren, Quantitative analysis of microbial volatiles in damp Swedish houses, in: Flannigan, et al. (Eds.), *Health Implications of Fungi in Indoor Environments*, Elsevier, Amsterdam, 1994, pp. 291-305.
- [20] K. Wilkins, M.O. Hansen, K. Larsen, MVOC analysis for detection of microbial growth in buildings II. A new sampling method?, 1999. http://www.asc.amr.dk/forskning/43rapporter/unders_bygn/Under_byg.pdf
- [21] R. Keller, K. Senkpiel, H. Ohgke, Use of MVOC measurements and odor perception as indicator of mould in indoor areas, in: *Bioaerosols, Fungi and Mycotoxins: Health Effects, Assessment, Prevention and Control*, E. Johannig, New York, 1999.
- [22] T.O. Larsen, J. Frisvad, Production of volatiles and presence of mycotoxins in conidia of common indoor *Penicillia* and *Aspergillii*, in: Flannigan, et al. (Eds.), *Health Implications of Fungi in Indoor Environments*, Elsevier, Amsterdam, 1994, pp. 251-279.
- [23] J. Martin, P. Gao, Volatile Metabolites Produced by *Stachybotrys chartarum* on Rice and Gypsum Board, 2000. <http://www.aiha.org/abs00/0iaq.html>.
- [24] J. Olsson, T. Borjesson, T. Lundstedt, J. Schnurer, Volatiles for mycological quality grading of barley grains: determinations using gas chromatography-mass spectrometry and electronic nose, *Int. J. Food Microbiol.* 59 (2000) 167-178.
- [25] A.-L. Pasanen, S. Lappalainen, P. Pasanen, Volatile organic metabolites associated with some toxic fungi and their mycotoxins, *Analyst* 121 (1996) 1949-1953.
- [26] J. Schnurer, J. Olsson, T. Borjesson, Fungal volatiles as indicators of food and feeds spoilage, *Fungal Genet. Biol.* 27 (1999) 209-217.
- [27] A.-L. Sunesson, W.H.J. Vaes, C.A. Nilsson, G. Blomquist, B. Andersson, R. Carlson, Identification of volatile metabolites from five fungal species cultivated on two media, *Appl. Environ. Microbiol.* 61 (1995) 2911-2918.
- [28] D. Tuma, R.N. Sinha, W.E. Muir, D. Abramson, Odor volatiles associated with microflora in damp ventilated and non-ventilated bin-stored bulk wheat, *Int. J. Food Microbiol.* 8 (1989) 103-119.
- [29] C.K. Wilkins, S. Scholl, Volatile metabolites of some barley storage molds, *Int. J. Food Microbiol.* 8 (1989) 11-17.
- [30] K. Wilkins, K. Larsen, Variation of volatile organic compound patterns of mold species from damp buildings, *Chemosphere* 5 (1995) 3225-3236.
- [31] A. Jonsson, F. Winquist, J. Schnurer, H. Sundgren, I. Lundstrom, Electronic nose for microbial quality classification of grains, *Int. J.*

Food Microbiol. 35 (1997) 187-193.

[32] G. Keshri, N. Magan, P. Voysey, Use of an electronic nose for the early detection and differentiation between spoilage fungi, *Lett. Appl. Microbiol.* 27 (1998) 261-264.

[33] G. Keshri, N. Magan, Detection and differentiation between myco-toxigenic and non mycotoxigenic strains of two *Fusarium* spp. using volatile production profiles and hydrolytic enzymes, *J. Appl. Microbiol.* 89 (2000) 825-833.

[34] J. Olsson, T. Borjesson, T. Lundstedt, J. Schnurer, Detection and quantification of ochratoxin A and deoxynivalenol in barley grains by GC-MS and electronic nose, *Int. J. Food Microbiol.* 72 (2002) 203-214.

[35] S.S. Schiffman, D.W. Wyrick, G.A. Payne, G. O'Brian, H.T. Nagle, Effectiveness of an electronic nose for monitoring bacterial and fungal growth, in: *Proceedings of the ISOEN 2000, Brighton, July 20-24, 2000*, pp. 173-180.

[36] K.C. Persaud, P.D. Wareham, P.A. Payne, H. Chueh, J.V. Hatfield, A portable electronic nose for dry rot detection, in: *Actes des conférences "La métrologie des odeurs et les nez électroniques"*, Eurodeur, Airodeur, 2001, Paris.

[37] Fix, Hodges, Discriminatory analysis, non-parametric discrimination: consistency properties, Technical Report 4, USAF School of Aviation Medicine, Randolph Field, TX, 1951.

[38] S.A. Dudani, The distance weighted k-nearest neighbour rule, *IEEE Trans. Syst. Man. Cyber.* 6 (1976) 325-327.

[39] T. Baily, A.K. Jain, A note on distance-weighted k-nearest neighbour rules, *IEEE Trans. Syst. Man. Cyber.* 8 (1978) 311-313.

[40] B.V. Dasarathy, *Nearest Neighbor Norms: NN Pattern Classification Techniques*, IEEE Computer Society Press, Los Alamitos, CA, 1991.

[41] R.L. Morin, D.E. Reaside, A reappraisal of distance weighted k-NN classification for pattern recognition with missing data, *IEEE Trans. Syst. Man. Cyber.* 11 (1981) 241-243.

[42] J.E. MacLeod, A. Luk, D.M. Titterington, A re-examination of the distance weighted k-NN classification rule, *IEEE Trans. Syst. Man Cyber.* 15 (1987) 580-585.

[43] B. Dubuisson, M. Masson, A statistical decision rule with incomplete knowledge about the classes, *Pattern Recog.* 26 (1993) 155-165.

[44] J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nearest neighbor algorithm, *IEEE Trans. Syst. Man. Cyber. SMC-15* (4) (1985) 580-585.

[45] Y.K. Kim, J.H. Han, Fuzzy k-NN Algorithm using modified k-selection, in: *Proceedings of the Fourth IEEE International Conference on Fuzzy Systems*, vol. 3, March 20-24, 1995, pp. 1673-1679.

[46] T. Denoeux, A k-nearest neighbor classification rule based on Dempster-Shafer theory, *IEEE Trans. Syst. Man. Cyber.* 25 (1995) 804-813.

[47] F. Marcelloni, Recognition of olfactory signals based on supervised c-means and k-NN algorithms, *Pattern Recog. Lett.* 21 (2001) 1007-1019.

[48] W. Pedrycz, J. Waletzky, Fuzzy clustering with partial supervision, *IEEE Trans. Syst. Man. Cyber. B* 27 (1997) 787-795.

[49] C. Rita de Franco, L. Silva Vidal, A validity measure for hard and fuzzy clustering derived from Fisher's linear discriminant, in: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'2002)*, 2002, May 12-17, 2002, pp. 1493-1498.