

# Recovered Paperboard Samples Identification by Means of Mid-Infrared Sensors

Jordi-Roger Riba, *Member, IEEE*, Trini Canals, and Rosa Cantero

**Abstract**— Paperboard is widely and increasingly applied as a packaging material and in many applications it is in direct contact with foodstuff. The increasing use of recovered paperboard has led to the production of paperboard containing several types of contaminants. In the case of using recovered paperboard some of these contaminants may migrate into the food in concentrations considered harmful to human health. To prevent this problem, a very fast and nondestructive method to identify recovered paperboard samples from those produced mainly from virgin fibers is developed in this paper. Therefore, recovered samples may be identified, so a special consideration may be given to these samples. To this end, Fourier transform mid-infrared spectroscopy was applied to acquire the mid-infrared spectra of the paperboard samples. Next, statistical multivariate feature extraction and classification methods were applied to identify incoming samples produced from recovered fibers. Experimental results presented here prove that the proposed scheme allows obtaining high classification accuracy with a very fast response.

**Index Terms**— Infrared spectroscopy, multivariate methods, paperboard, process control, classification.

## I. INTRODUCTION

PAPER and board (P&B) are broadly used as food packaging materials and are being increasingly applied due to continuous advances in both material formulation and package design and the tendency of using recyclable materials [1]. Since materials and food packaging manufacturers are responsible for the safety of their products [2], when selecting materials intended for direct contact with foodstuffs consumer's safety becomes a key issue.

P&B are mainly composed of different types of vegetable pulp and are often in direct contact with dry foodstuffs. However, P&B packaging materials may contain a multitude of other products in their composition, including fillers, starch, biocides, and retention aids among others and they may also be coated with waxes, polymers or polyethylene [3]. It is recognized that different chemicals may pass to the food

supply at different stages of the food chain, including pesticides, environmental contaminants or food additives among others [3]. The materials used in food packaging may release small amounts of their chemical components, therefore contaminating foodstuffs [3,4] and finally being ingested by the consumer [3]. This has been officially recognized by the Council of Europe [5] and the European Food Safety Authority (EFSA) [6]. When these chemical components migrate to foodstuffs in a concentration sufficient enough, it could involve a hazard for consumer's health, which is in discordance with the European regulations [7]. According to [8], the migration is a diffusion process, which is often influenced by an interaction between the packaging material and the food. This interaction may be significant, since in some cases it is reported to be one-hundred times superior to the contribution of pesticides or environmental pollutants [4]. While contamination of foodstuffs by environmental pollutants has been deeply studied, the contribution of food packaging as a supplementary contaminants' source has received much less attention, although food packaging is considered to play an important role in human xenobiotic exposure [9]. There are possible solutions to alleviate this problem, such as the use of functional barriers. However, they do not completely reduce the contamination risk [10].

The use of recovered P&B has expanded considerably in the last decades, which has irremediably led to the production of products containing several types of contaminants [11]. Recovered P&B may have different origins and compositions and may contain different components including printing inks, waxes, adhesives, mineral oils, biocides and surfactants among others [1]. Some of these chemical agents may migrate from the packaging materials to foodstuffs, thus causing a potential risk to human health. Because of this fact the content of materials based on recovered fiber must be limited below certain limits when intended for contact with foodstuff [1] since recovered fiber is a major migrants' source [3]. Recovered P&B is not used for liquid food contact [6] but they are mostly used in direct contact with solid dry foodstuffs like rice, salt, cereals, sugar or pasta among others [1,12]. Therefore, the migration phenomenon is virtually restricted to the volatile chemical components contained in the P&B material. They are evaporated and transferred through the gas phase [13] and finally they recondense in the food surface [14].

Diverse types of chemical compounds which may promote adverse health effects have been identified in P&B packaging,

Manuscript received December 17, 2012.

J.-R. Riba Ruiz is with the Electrical Engineering Department, Escola d'Enginyeria d'Igualada, Universitat Politècnica de Catalunya, 08700 Igualada, Spain (e-mail: jordi.riba@eei.upc.edu).

T. Canals Parelló and R. Cantero Gómez are with the Escola d'Enginyeria d'Igualada, Universitat Politècnica de Catalunya, 08700 Igualada, Spain.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

including phthalates [15,16] and mineral oil hydrocarbons (MOH) [17] among others. It has been reported that recovered P&B used in food packaging may contain mineral oil in such a concentration as to contaminate the food in contact at levels much higher than those considered safe for human health [18]. To minimize MOH exposure, EFSA proposes avoiding the use of materials containing MOH in the production of food packages [6]. To this end, recovered fiber sources should be segregated to avoid their use in P&B food packages.

This paper is focused to identify paperboard samples produced from recovered fibers from those produced mainly from virgin fibers in a straightforward manner, thus avoiding the need of a previous sample pretreatment, while providing improved analysis speed and simplicity. For food packagers it would be highly valuable to have a simple, fast and easy-to-use system to identify paperboard samples composed mainly of recovered paperboard in their incoming stock. In this work the identification is carried out by analyzing the mid-infrared spectra of the analyzed samples. This is because regulations are becoming more restrictive in regard to the use of paperboard produced from recovered fibers when applied in direct contact with the foodstuffs. In addition, in the certifications provided by paperboard suppliers, the issues concerning the use of recovered paperboard are usually not considered.

Mid-infrared (MIR) spectroscopy has been used for analyzing paper structure and pulp chemistry [19] since it allows noninvasive and close to real-time analysis [20]. MIR radiation is emitted from a broadband light source and consists of photons which are in resonance with fundamental rotational and vibrational modes related to most organic and inorganic molecules [20].

The nondestructive method proposed in this work is based on the combined action of Fourier transform mid-infrared (FTIR) spectroscopy, one of the most popular infrared techniques since it provides high resolution spectra [21], and multivariate feature extraction and classification algorithms. The latter ones directly classify an unknown incoming paperboard sample as produced from recovered or virgin fibers by analyzing the spectral data provided by the FTIR spectrometer. Additional advantages of the method developed in this paper include no sample pretreatment, no consumption of chemicals and reagents, no need of both a qualified laboratory technician and laboratory grade facilities, and particularly fast classification response. The incoming paperboard samples often present extremely varied compositions since they have different origins and manufacturers. Because of the diverse origins and formulations, this is a challenging problem since the analyzed data sets are very heterogeneous.

## II. PAPERBOARD SAMPLES ANALYZED

A total amount of 57 paperboard samples are analyzed, which are split into two groups according to its origin, namely FBB (folding boxboard produced from virgin fibers, 29 samples) and WLC (coated recovered cartonboard, 28 samples). Whereas the FBB samples are produced essentially

from virgin fibers, the WLC samples are produced mostly from recovered fibers. All samples analyzed in this work present only one coated side. The samples were supplied by the firms Comart, Torraspapel, Union Papelera, Reno de Medeci, Papyrus and Attica Displayline. As the paperboard samples have been collected from different manufacturers and different catalogues, they constitute a highly heterogeneous group with different compositions. This fact makes the classification problem very difficult. This is especially true for the WLC samples, since they are mainly composed of recovered fibers.

To evaluate the classification models performance, the whole body of samples is divided into two sets, namely training and prediction sets. The training set is used to train or calibrate the mathematical classification model, whereas the prediction set is used to check the model accuracy when identifying samples different than those used to calibrate the model. Table I shows the paperboard samples dealt with, their origin, and the group in which they are assigned. Fig. 1 shows two of the 57 samples studied in this work.

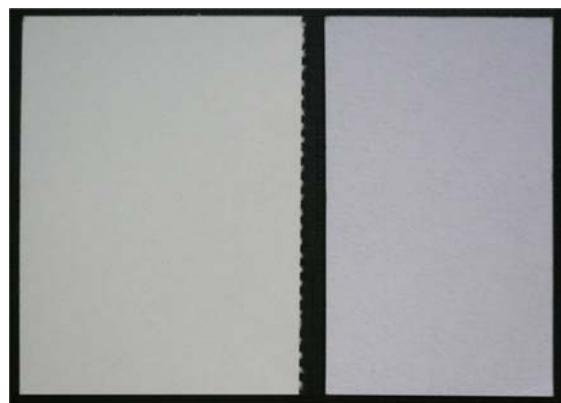


Fig. 1. Left side: FBB sample 13t. Right side: WLC sample 27p.

## III. DATA ACQUISITION BY MEANS OF FTIR SPECTROSCOPY

Infrared spectroscopy allows identifying different molecules types, including polymeric, organic, inorganic, and biological molecules. Infrared spectroscopy has been applied in the paper industry to analyze pulp chemistry and paper structure [18], for nonintrusive process control and for fast determination of grammage and moisture content [22] among others. In the infrared spectrum of a paperboard sample, the majority of spectral bands are originated by the cellulose which tends to mask the spectral information provided by other products already present in its composition [23]. Therefore, to design an efficient classification method it is mandatory to extract and select the relevant information of the paperboard spectrum.

This paper deals with Fourier transform mid-infrared (FTIR) spectroscopy because it is a nondestructive instrumental method which presents very fast response, requires small sample size, does not consumes chemicals or reagents, may be applied in situ, does not require sample preparation, provides improved spectral bands specificity and offers a reasonable cost per sample in regular use among others. It also allows avoiding the time-consuming analyses which are often applied to determine the content of different

analytes in the paperboard samples while providing a large amount of digitalized spectral data which may be used for fast identification of paperboard samples. However, due to the large amount of information provided by FTIR spectroscopy and the heterogeneous origin of the paperboard samples dealt with, this is a difficult task that requires further data processing by means of appropriate mathematical methods.

TABLE I  
PAPERBOARD SAMPLES ANALYZED IN THIS WORK.

Sample	Type	Reference	Grammage (g/m <sup>2</sup> )	Training set	Prediction set
1t	FBB	Comcote CM	235	x	
2t	FBB	Comcote CM	340	x	
3t	FBB	Comkraft RC	345	x	
4t	FBB	Crescendo	200	x	
5t	FBB	Crescendo	300	x	
6t	FBB	Simcote	295	x	
7t	FBB	Simcote	350	x	
8t	FBB	Comkraft	200	x	
9t	FBB	Comkraft	295	x	
10t	FBB	Comkraft	305	x	
11t	FBB	Simcote	295	x	
12t	FBB	Stompack	280	x	
13t	FBB	Incada exel	300	x	
14t	FBB	Trambritte	265	x	
15t	FBB	Trambritte	340	x	
16t	WLC	TBC	420	x	
17t	WLC	Libra SBS	350	x	
18t	WLC	115	210	x	
19t	WLC	115	250	x	
20t	WLC	512	320	x	
21t	WLC	512	400	x	
22t	WLC	512	500	x	
23t	WLC	513	350	x	
24t	WLC	513	450	x	
25t	WLC	DBC	280	x	
26t	WLC	Display 2	295	x	
27t	WLC	Gemini	350	x	
28t	WLC	Clubboard DK	250	x	
29t	WLC	Clubboard DK	400	x	
1p	FBB	Comcote CM	305		x
2p	FBB	Comkraft RC	295		x
3p	FBB	Comkraft RC	390		x
4p	FBB	Crescendo	240		x
5p	FBB	Simcote	255		x
6p	FBB	Simcote	320		x
7p	FBB	Comcote	270		x
8p	FBB	Comcote	340		x
9p	FBB	Simcote	245		x
10p	FBB	Simcote	270		x
11p	FBB	Stompack	255		x
12p	FBB	Incada exel	250		x
13p	FBB	Trambritte	250		x
14p	FBB	Trambritte	315		x
15p	WLC	TBC	330		x
16p	WLC	Libra SBS	280		x
17p	WLC	TBC	370		x
18p	WLC	115	230		x
19p	WLC	512	300		x
20p	WLC	512	350		x
21p	WLC	512	450		x
22p	WLC	513	320		x
23p	WLC	513	400		x
24p	WLC	Gemini	250		x
25p	WLC	Gemini	280		x
26p	WLC	Gemini	450		x
27p	WLC	Clubboard DK	350		x
28p	WLC	Ibiza	450		x

FTIR spectrometers are basically composed of a Michelson interferometer and a movable mirror. The signal emerging

from the interferometer is processed via Fourier transform to obtain the spectrum. An attenuated total reflectance (ATR) module with a diamond crystal is used, which by applying a slight pressure, permits recording the reflectance spectrum of the analyzed paperboard sample. The ATR module measures the changes in the attenuated radiation due to the total internal reflection phenomenon produced when a mid-infrared beam comes into contact with the analyzed sample.

In this paper the FTIR spectra of the 57 paperboard samples are recorded at room temperature (25±1°C) by means of an ATR cuvette over the wavenumber range 4000–650cm<sup>-1</sup> without any treatment of the paperboard samples. Each record is obtained by averaging four consecutive scans made on an IR Spectrum One (S/N 57458) from PerkinElmer (Beaconsfield, UK) equipped with an ATR internal reflectance module (Universal Sampling Accessory, S/N P0DL01101418). The FTIR spectra of the 57 paperboard samples provide a data matrix consisting of 57 rows (paperboard samples) and 3351 columns (reflectance spectrum). These data in the reflectance mode are sent to the computer to process the information.

Fig. 2 shows the FTIR spectra of a cellulose sample, a FBB and a WLC paperboard samples. According to Fig. 2 the spectra of the paperboard and cellulose samples are similar but the former ones are much more attenuated, as it can be appreciated in the spectral stretching bands corresponding to the O-H, C-H and C-O-C groups. In addition, the C=O spectral stretching band is almost totally attenuated in the case of the FBB and WLC samples.

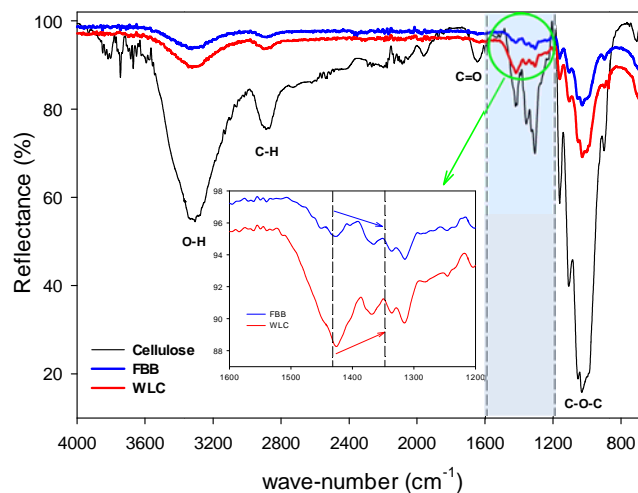


Fig. 2. FTIR reflectance spectra of different types of analyzed samples: pure cellulose sample, FBB and WLC samples.

When comparing the spectra of a paperboard sample obtained from virgin fibers (FBB) with that obtained from recovered fibers (WLC), the 1600-1200 cm<sup>-1</sup> spectral interval is that with a more accentuated difference. The shape of the 1600-1200 cm<sup>-1</sup> spectral interval for both the cellulose and FBB samples is very similar, but more attenuated for the latter. Contrarily, in the case of the WLC sample, the shape of this interval differs from that of the two other samples (cellulose and FBB).

#### IV. FEATURE EXTRACTION AND CLASSIFICATION METHODS

As exposed, the spectra of each P&B sample provided by the FTIR spectrometer consists of 3351 data points (reflectances at each wave-number). Therefore it is essential to apply efficient multivariate processing methods for concentrating the analytically significant information in a reduced set of latent variables [19,24]. The mathematical methods designed to calculate the latent variables are often called feature extraction methods, which also allow removing most of the noise typically present in the raw signals. Supervised feature extraction methods, unlike unsupervised algorithms, use class labels to evaluate the performance of the latent variables. A supervised algorithm is guided by an expert user that establishes the class labels of the training samples. To boost the discrimination between classes, supervised algorithms are needed [25]. The latent variables are often obtained by performing mathematical combinations of the original variables. To this end three feature extraction algorithms have been applied and their performance has been evaluated: principal component analysis (PCA), PCA linked with canonical variate analysis (CVA) and extended canonical variate analysis (ECVA). Once the problem dimensionality has been reduced it is necessary to apply a classification method. For this purpose the  $k$  nearest neighbor algorithm ( $k$ NN) has been applied in the classification stage.

PCA is probably one of the most widely applied feature extraction techniques [24,26,27]. It is an unsupervised method focused to reduce the dimensions of the original data set while concentrating the relevant information contained in the original variable set into a reduced set of latent variables [28]. The original variables (reflectances at different wave-numbers) are transformed via PCA into orthogonal latent variables called principal components (PCs). The original variables are linearly combined to obtain the PCs, which represent the orthogonal directions with higher variance. The PCA algorithm provides as many PCs as original variables. However, only a few PCs are usually retained, those explaining a sufficient amount of the overall variance. To prevent overfitting it is essential to carefully determine the number of PCs to retain and to split the overall samples set into a training and a prediction set [26]. However, there is no general method to select the optimum number of PCs to be retained.

The CVA algorithm is a supervised multi-class feature extraction method specially intended for accentuating differences between data classes [29]. CVA is a generalization of the two-class Fisher linear discriminant analysis algorithm. Unlike PCA, CVA searches the directions in space that strengthens the differences between classes in the original data [30] by applying a criterion based on maximizing separation between classes while minimizing separation within classes. The latent variables calculated by the CVA algorithm, which are not necessarily orthogonal, are known as canonical variates (CVs). The number of CVs calculated by the CVA algorithm is equal to the number of classes in the problem minus one. A major disadvantage of CVA is that it cannot directly deal with data sets in which the number of samples is

less than the number of variables. This is the case of FTIR spectrometry, which provides 3351 variables. To surpass this drawback, the PCA algorithm is applied to reduce the number of variables prior to CVA application.

The supervised ECVA algorithm was developed by Norgaard *et al.* [29,31] to overcome the drawback of the CVA algorithm related to the impossibility of dealing with data sets containing a larger number of variables than samples. The advantage of ECVA compared to CVA, is that ECVA allows the direct calculation of the latent variables, called extended canonical variates (ECVs). Consequently, ECVA avoids the need of applying the PCA algorithm prior to ECVA. As in the CVA algorithm, ECVA calculates as many ECVs as number of classes in the problem minus one.

Once the problem dimensionality has been reduced by means of a suitable feature extraction method, the classification stage may be applied. For this purpose the nonparametric  $k$  nearest neighbor ( $k$ NN) algorithm is applied in this paper. It is one of the simplest and most effective classification algorithms [11].  $k$ NN classifies an input sample in a determined class by taking into account a weighted vote of the  $k$  nearest training set neighbors while applying the majority voting rule. To this end, the  $k$  nearest training set neighbors of an input sample are located. Then,  $k$ NN assigns a score  $k$  to the nearest neighbor's class,  $k - 1$  to the class of the second nearest neighbor and so on until reaching a score of 1. The input sample is assigned to the class with the highest score. Some authors have recommended using  $k$  values from 3 to 5 [32]. The  $k$ NN algorithm provides as many outputs as number of classes defined in the problem, whose values are comprised between 0 and 1. They denote the membership degree of the input samples to each class

#### V. RESULTS AND DISCUSSION

In this section the classification of the analyzed paperboard samples is carried out by means of the analysis of the FTIR spectra and the combined action of feature extraction and classification methods. The analyzed feature extraction methods are PCA, CVA in combination with PCA and ECVA. In all cases the  $k$ NN classifier is applied.

Since the 57 paperboard samples dealt with have different origins, it is assumed that these differences are reflected in their spectra. It is known that 29 of the total amount of 57 samples analyzed are produced from virgin fibers (FBB samples), whereas the remaining 28 samples are produced from recovered fibers (WLC). Once divided in these two groups, the 57 samples are randomly split into a training set and a prediction set. The training set contains 29 samples (15 FBB and 14 WLC samples) whereas the remaining 28 samples (14 FBB and 14 WLC samples) are assigned to the prediction set. Therefore, both the training and the prediction sets contain approximately 50% of the paperboard samples each.

Whereas the training set matrix contains 29x3351 elements, the prediction set matrix contains 28x3351 elements. The 3351 columns correspond to the spectral content of each sample and each matrix contains as many rows as samples analyzed.

However, the spectral interval considered may have a

notorious influence on the classification results. To study this effect two spectral intervals are analyzed, i.e. the whole FTIR spectrum (4000-650  $\text{cm}^{-1}$ ) and the spectral interval with more accentuated differences (1600-1200  $\text{cm}^{-1}$ ).

*A. Results considering the whole spectrum (4000-650  $\text{cm}^{-1}$ )*

In this section the feature extraction and classification algorithms are applied to the whole FTIR spectrum of the analyzed samples.

As explained, when applying the PCA algorithm, a reduced number of PCs are usually retained. However, there is not universal rule to select the appropriate number of PCs. To improve separation, PCA was applied on a reduced number of PCs, those explaining at least 99.5% of the overall variance. As shown in Fig. 3, this condition was attained by selecting the first six PCs. Therefore, PCA concentrates most of the useful information provided by the 3351 spectral variables in only six PCs, thus greatly reducing the problem dimensionality. This makes it possible the posterior application of the CVA algorithm.

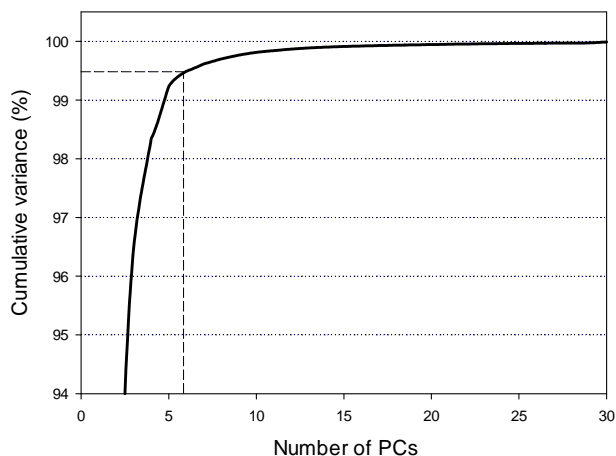


Fig. 3. Cumulative variance as a function of the number of retained PCs in the training data set when considering the overall 4000-650  $\text{cm}^{-1}$  spectral interval.

Additionally, leave-one-out cross-validation was applied to predict the training samples class as an alternative method to select an appropriate number of PCs. It also resulted in the first six PCs.

Fig. 4 shows the results based on the PCA algorithm when considering 29 training samples and 28 prediction samples. After applying the  $k$ NN classifier, classification results according to the PCA (6 PCs) +  $k$ NN scheme resulted in 23 ( $k = 3$ ), 25 ( $k = 4$ ) and 26 ( $k = 5$ ) correctly classified samples out of 28 prediction samples (82.14%, 89.29% and 92.86% classification rates, respectively). As shown in Fig. 4, the WLC samples appear more dispersed than the FBB samples since the WLC samples are produced from recovered fibers, thus being a more inhomogeneous samples set.

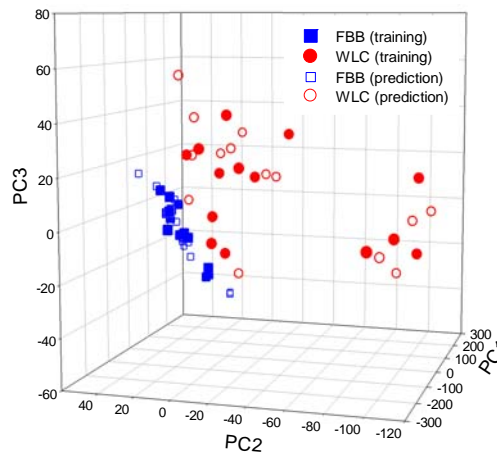


Fig. 4. Training and prediction samples plotted in the space defined by the first three PCs arising from the PCA algorithm in the 4000-650  $\text{cm}^{-1}$  spectral interval. This is a partial view since only three out of the six retained PCs are plotted.

Next, the performance of the CVA algorithm was evaluated. To this end the CVA was applied to the first six PCs arising from the PCA, whose results are presented in Fig. 5. Note that when dealing with a two-class problem, the CVA algorithm only provides one CV. The performance of the PCA (6 PCs) + CVA algorithms followed by the  $k$ NN classification method ( $k = 3,4,5$ ) was also examined. It resulted in 100% correctly classified prediction samples in all cases. Fig. 5 shows again that the WLC samples appear more dispersed than the FBB samples.

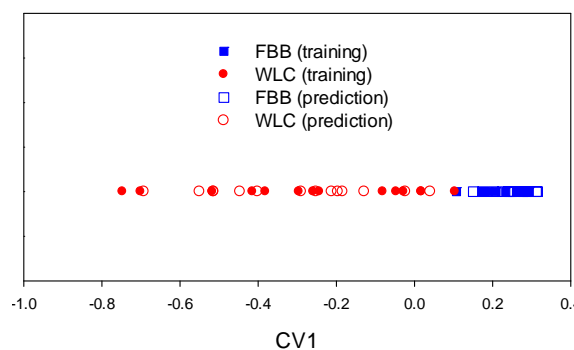


Fig. 5. Training and prediction samples plotted in the space defined by the only CV arising from the PCA (6 PCs) + CVA algorithms in the 4000-650  $\text{cm}^{-1}$  spectral interval.

The behavior of the ECVA algorithm was also tested, whose results are presented in Fig. 6. The prior reduction of the number of variables is not required when dealing with ECVA, so this algorithm is directly applied to the raw spectral data. Similarly to CVA, ECVA only returns one ECV when dealing with a two-class problem. After applying the  $k$ NN classifier, the classification results according to the ECVA +  $k$ NN scheme resulted in 26 ( $k = 3$ ) and 27 ( $k = 4, 5$ ) correctly classified samples out of 28 prediction samples (92.86% and 96.43% classification rate, respectively).

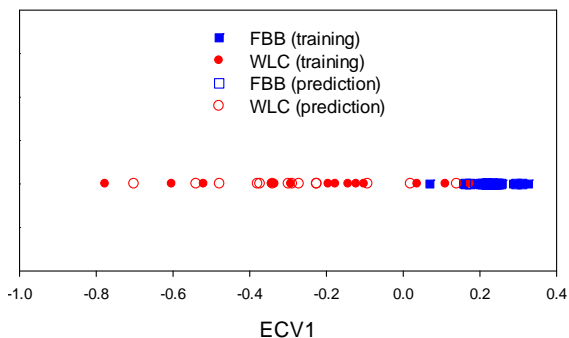


Fig. 6. Training and prediction samples plotted in the space defined by the only ECV arising from the ECVA algorithm in the 4000-650  $\text{cm}^{-1}$  spectral interval.

### B. Results considering the 1600-1200 $\text{cm}^{-1}$ interval

As explained, the overall 4000-650  $\text{cm}^{-1}$  spectral region is maybe not the most suited interval to accentuate the differences between paperboard samples produced from virgin and recovered fibers. So, in this section, the behavior of the proposed classification scheme is tested when considering the 1600-1200  $\text{cm}^{-1}$  spectral interval.

The PCA algorithm was applied to the 1600-1200  $\text{cm}^{-1}$  spectral interval. After analyzing the cumulative variance explained by the first few PCs arising from the PCA algorithm, it resulted that the first five PCs explained the 99.5% of the overall variance. The leave-one-out cross-validation algorithm was also applied to the training samples as an alternative method to select the most suitable number of PCs. It also resulted in the first five PCs.

The results based on the PCA algorithm are presented in Fig. 7. After applying the  $k$ NN classifier, the classification results of the prediction samples according to the PCA (5 PCs) +  $k$ NN scheme resulted in 24 ( $k = 3$ ) and 27 ( $k = 4, 5$ ) correctly classified prediction samples out of 28 (85.71% and 96.43% classification rates, respectively). These results are slightly better than those presented in Fig. 4.

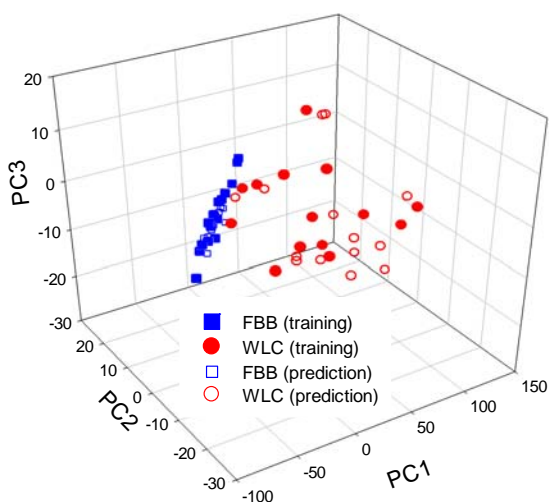


Fig. 7. Training and prediction samples plotted in the space defined by the first three PCs arising from the PCA algorithm in the 1600-1200  $\text{cm}^{-1}$  spectral interval. This is a partial view since only three out of the five retained PCs are plotted.

Next, the CVA was applied to the first five PCs, as shown

in Fig. 8. Next the  $k$ NN classifier was applied to the data supplied by the PCA (5 PCs) + CVA algorithms, resulting in 100% classified prediction samples.

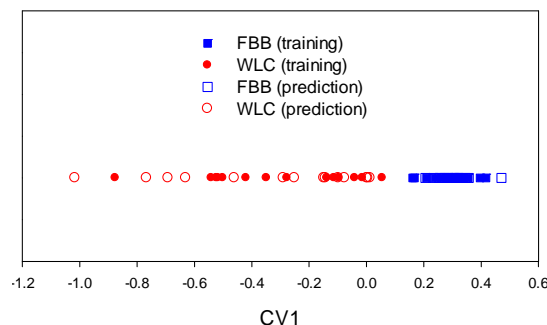


Fig. 8. Training and prediction samples plotted in the space defined by the only CV arising from the PCA (5 PCs) + CVA algorithms in the 1600-1200  $\text{cm}^{-1}$  spectral interval.

When comparing Figs. 5 and 8, it is clearly seen that the FBB and WLC samples appear more separated when considering the 1600-1200  $\text{cm}^{-1}$  than the overall 4000-650  $\text{cm}^{-1}$  spectral interval. This result corroborates the assumption that the 1600-1200  $\text{cm}^{-1}$  spectral interval is the most useful for this specific application.

Next, the performance of the ECVA algorithm was analyzed, whose results are presented in Fig. 9. This scheme lead to a classification rate of the prediction samples of 96.43% since 27 samples out of 28 were correctly classified in all cases ( $k = 3, 4$  and 5).

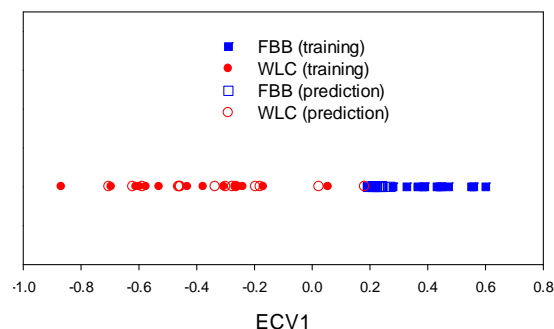


Fig. 9. Training and prediction samples plotted in the space defined by the only ECV arising from the ECVA algorithm in the 1600-1200  $\text{cm}^{-1}$  spectral interval.

### C. Results summary

A summary of the classification results attained by applying the three analyzed mathematical methods and considering the two studied spectral intervals is presented in Table II.

In this problem the  $k$ NN classifier generates two output values (one per class; i.e. FBB and WLC classes) per sample analyzed, which are normalized in the interval [0,1]. These values express the membership degree of the analyzed sample to each class. Values less than or equal to 0.5 indicate that the sample does not belong to the considered class, whereas output values greater than 0.5 indicate its membership to this class. Whereas an output value of 1 clearly indicates that the sample belongs to the considered class, an output value of 0.6 presents a high uncertainty degree. Therefore, to evaluate the classification results accuracy, the predictive residual error

sum of squares (PRESS) is used, which is calculated as in,

$$\text{PRESS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

$y_i \in [0,1]$  being the output of the  $i$ -th prediction sample provided by the  $k$ NN algorithm and  $\hat{y}_i \in \{0,1\}$  being the real value of  $y_i$ . The  $y_i$  values are known a priori since the prediction samples class is known. Note that the classification accuracy decreases with increasing values of the PRESS index.

TABLE II  
RESULTS SUMMARY

Statistical methods	kNN classifier	Prediction success rate	PRESS
<b>Spectral interval: 4000-650 cm<sup>-1</sup></b>			
PCA (6 PCs)	$k = 3$	23/28 (82.14%)	4.72
	$k = 4$	25/28 (89.29%)	4.50
	$k = 5$	26/28 (92.86%)	4.84
PCA (6 PCs) + CVA	$k = 3$	28/28 (100.0%)	0.00
	$k = 4$	28/28 (100.0%)	0.00
	$k = 5$	28/28 (100.0%)	0.00
ECVA	$k = 3$	26/28 (92.86%)	2.17
	$k = 4$	27/28 (96.43%)	1.56
	$k = 5$	27/28 (96.43%)	1.41
<b>Spectral interval: 1600-1200 cm<sup>-1</sup></b>			
PCA (5 PCs)	$k = 3$	24/28 (85.71%)	3.17
	$k = 4$	27/28 (96.43%)	2.62
	$k = 5$	27/28 (96.43%)	2.29
PCA (5 PCs) + CVA	$k = 3$	28/28 (100.0%)	0.00
	$k = 4$	28/28 (100.0%)	0.00
	$k = 5$	28/28 (100.0%)	0.00
ECVA	$k = 3$	27/28 (96.43%)	1.39
	$k = 4$	27/28 (96.43%)	1.39
	$k = 5$	27/28 (96.43%)	1.39

According to the results presented in Table II, the best results were attained when considering the 1600-1200 cm<sup>-1</sup> spectral interval. Additionally, in all analyzed cases, CVA was the feature extraction method with better performance, although ECVA also lead to very accurate classification results.

## VI. CONCLUSION

This paper has proposed a very fast, noninvasive, accurate and easy-to-apply method to discriminate between paperboard samples produced from recovered and virgin fibers. For this purpose FTIR spectroscopy has been analyzed in combination with feature extraction methods such as PCA, PCA + CVA and ECVA and the  $k$ NN classifier. The most appropriate spectral interval to be considered has also been analyzed, since as it has been shown, it may have a notorious influence on the results. The experimental results presented in this paper have shown that high classification rates may be achieved by applying the proposed methodology.

This application may be especially useful in packaging applications which require paperboard to be in direct contact with foodstuff, since these applications should avoid paperboard formulations produced mainly from recovered fibers, or at least special treatments must be applied to these materials.

## REFERENCES

- [1] V. I. Triantafyllou, K. Akrida-Demertzi, P. G. Demertzis, "A study on the migration of organic pollutants from recycled paperboard packaging materials to solid food matrices," *Food Chem.*, vol. 101, no. 4, pp. 1759–1768, 2007.
- [2] V. I. Triantafyllou, K. Akrida-Demertzi, P. G. Demertzis, "Migration studies from recycled paper packaging materials: development of an analytical method for rapid testing," *Anal. Chim. Acta.*, vol. 467, no. 1-2, pp. 253–260, 2002.
- [3] M. F. Poças, T. Hogg, "Exposure assessment of chemicals from packaging materials in foods: a review," *Trends Food Sci. Technol.*, vol. 18, pp. 219-230, 2007.
- [4] H. Gallart-Ayala, P. Lucci, O. Núñez, "Recent advances in LC-MS analysis of food-packaging contaminants," *Trends Anal. Chem.*, 2012; In Press.
- [5] Council of Europe, Consumer Health Protection Committee. Committee of experts on materials coming into contact with food policy statement concerning paper and board materials and articles intended to come into contact with foodstuffs, version 4 – 12.02.2009.
- [6] European Food Safety Authority (EFSA), Panel on Contaminants in the Food Chain (CONTAM), "Scientific Opinion on Mineral Oil Hydrocarbons in Food," *EFSA Journal*, vol. 10, no. 6:2704, pp. 1-185, 2012.
- [7] Council Directive 89/109/EEC of. On the approximation of the laws of the Member States relating to materials and articles intended to come into contact with foodstuffs. *Official Journal of the European Communities*, no. L, pp. 40/38-40/44, 21 Dec. 1988.
- [8] R. Sendón, A. Sanches Silva, I. Cooper, R. Franz, P. Paseiro, "Revision of analytical strategies to evaluate different migrants from food packaging materials," *Trends Food Sci. Technol.*, vol. 17, pp. 354–366, 2006.
- [9] J. Muncke, "Exposure to endocrine disrupting compounds via the food chain: Is packaging a relevant source?," *Sci.Total Environ.*, vol. 407, pp. 4549–4559, 2009.
- [10] J. Ewender, R. Franz, F. Welle, Permeation of Mineral Oil Components from Cardboard Packaging Materials through Polymer Films," *Packag. Technol. Sci.* 2012, Early view.
- [11] J.-R. Riba, T. Canals, R. Cantero, "Comparative Study of Multivariate Methods to Identify Paper Finishes Using Infrared Spectroscopy," *IEEE Trans. Inst. Meas.*, vol. 61, no. 4, pp. 1029-1036, April 2012.
- [12] C. Nerin, E. Asensio, "Behaviour of organic pollutants in paper and board samples intended to be in contact with food," *Anal. Chim. Acta.*, vol. 508, pp. 185–191, 2004.
- [13] M. Biedermann, K. Grob, "On-line coupled high performance liquid chromatography–gas chromatography for the analysis of contamination by mineral oil. Part 2: Migration from paperboard into dry foods: Interpretation of chromatograms," *J. Chromatogr. A*, vol. 1255, pp. 76–99, 2012.
- [14] E. Asensio, C. Nerin, "Evaluation of a Screening Method for Classifying Virgin and Recycled Paper and Board Samples," *Packag. Technol. Sci.*, vol. 22, pp. 311–322, 2009.
- [15] T. Fierens, K. Servaes, M. Van Holderbeke, L. Geerts, S. De Henauw, I. Sioen, G. Vanermen, "Analysis of phthalates in food products and packaging materials sold on the Belgian market," *Food and Chemical Toxicology*, vol. 50, pp. 2575–2583, 2012.
- [16] K. Fiselier, K. Grob, "Barriers against the Migration of Mineral Oil from Paperboard Food Packaging: Experimental Determination of Breakthrough Periods," *Packag. Technol. Sci.*, vol. 25, no. 5, pp. 285–301, 2012.
- [17] M. Biedermann, K. Grob, "On-line coupled high performance liquid chromatography–gas chromatography for the analysis of contamination by mineral oil. Part 1: Method of analysis," *J. Chromatogr. A*, vol. 1255, pp. 56–75, 2012.
- [18] M. Biedermann, Y. Uematsu, K. Grob, "Mineral Oil Contents in Paper and Board Recycled to Paperboard for Food Packaging," *Packag. Technol. Sci.*, vol. 24, no. 2, pp. 61–73, 2011.
- [19] J. Pan, K. L. Nguyen, "Development of the photoacoustic rapid-scan FT-IR-based method for measurement of ink concentration on printed paper," *Anal. Chem.*, vol. 79, no. 6, pp. 2259–2265, 2007.
- [20] S.-S. Kim et al., "Potential and Challenges for Mid-Infrared Sensors in Breath Diagnostics," *IEEE Sensors J.*, vol. 10, no. 1, pp. 145-158, Jan. 2010.

- [21] J. Huang, R. Gosangi, and R. Gutierrez-Osuna, "Active Concentration-Independent Chemical Identification With a Tunable Infrared Sensor," *IEEE Sensors J.*, vol. 12, no. 11, pp. 3135-3142, Nov. 2012.
- [22] R. Hodges, H. Cullinan, G. Krishnagopalan, "Recent advances in the commercialization of NIR (near-infrared) based liquor analyzers in the pulping and recovery area," *TAPPI Journal*, vol. 5, no. 11, pp. 3-10, 2006.
- [23] B. Stuart, *Infrared Spectroscopy: Fundamentals and Applications*. Wiley: Chichester, U.K, 2004.
- [24] M. T. Bona, J. M. Andrés, "Reflection and transmission mid-infrared spectroscopy for rapid determination of coal properties by multivariate analysis," *Talanta*, vol. 74, no. 4, pp. 998-1007, 2008.
- [25] J. M. Leiva-Murillo, R. Santiago-Mozos, F. Pérez-Cruz, A. Artés-Rodríguez, "Comparison of Supervised Feature Extraction Methods for Multispectral Images," in *Proc. of the Learning'02*, Madrid, 2002.
- [26] N. Bhattacharyya, R. Bandyopadhyay, M. Bhuyan, B. Tudu, D. Ghosh, A. Jana, "Electronic Nose for Black Tea Classification and Correlation of Measurements With "Tea Taster" Marks," *IEEE Trans. Inst. Meas.*, vol. 57, no. 7, pp. 1313-1321, Jul. 2008.
- [27] O. M. Conde, L. Uriarte, P. B. García-Allende, A. M. Cubillas, F. Anabitarte, and J. M. López-Higuera, "Spectral and Optimized Marks for Qualitative Material Discrimination," *IEEE Sensors J.*, vol. 12, no. 1, pp. 230-236, Jan. 2012.
- [28] X. Wang, N. D. Georganas, E. M. Petriu, "Fabric Texture Analysis Using Computer Vision Techniques," *IEEE Trans. Inst. Meas.*, vol. 60, no. 1, pp. 44-56, Jan. 2011.
- [29] L. Nørgaard, R. Bro, F. Westad, and S.B. Engelsen, "A modification of canonical variates analysis to handle highly collinear multivariate data," *J. Chemometr.*, vol. 20, no. 8-10, pp. 425-435, Aug.-Oct. 2006.
- [30] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Englewood Cliffs, NJ: Prentice-Hall, 2007, ch. 8-10.
- [31] L. Nørgaard, G. Soletormos, N. Harrit, M. Albrechtsen, O. Olsen, D. Nielsen, K. Kampmann, and R. Bro, "Fluorescence spectroscopy and chemometrics for classification of breast cancer samples—A feasibility study using extended canonical variates analysis," *J. Chemometr.*, vol. 21, no. 10/11, pp. 451-458, Nov. 2007.
- [32] L. A. Berrueta, R. M. Alonso-Salces, and K. Héberger, "Supervised pattern recognition in food analysis," *J. Chromatogr. A*, vol. 1158, no. 1/2, pp. 196-214, Jul. 2007.



**Jordi-Roger Riba Ruiz** (M'09) received the M.S. degree in physics and the Ph.D. degree from Universitat de Barcelona, Barcelona, Spain, in 1990 and 2000, respectively. In 1992, he joined Escola d'Enginyeria d'Igualada, Universitat Politècnica de Catalunya, Barcelona, as a full-time Lecturer, and he joined the Department of Electric Engineering in 2001. His research interests include signal processing methods, electromagnetic devices modeling, electric machines, variable-speed drive systems, fault diagnosis in electric machines, and fault detection algorithms.



**Trini Canals Parelló** obtained the degree in Chemistry at the University of Barcelona (Spain) in 1980 and the Ph.D. degree in Chemistry at the University Autònoma of Barcelona in 1997. She received a Master's in Tanning Technical Management from the Universitat Politècnica de Catalunya (Spain) in 1999. Since 1984 she has been working as a lecturer at the Igualada Industrial Engineering Technical School - Igualada Tanning School at the Universitat Politècnica de Catalunya. She is mainly involved in the development of innovative analytical methods for leather and paper processing.



**M. Rosa Cantero Gómez** obtained the degree in Chemistry at the University of Barcelona in 1989 and the Ph.D. degree in Chemistry at the Autonomous University of Barcelona. She has been a lecturer at the Igualada Engineering School, Technical University of Catalonia (Barcelona) since 1990. In 1991 she carried out postgraduate studies on Tanning Techniques at this same School. Her research deals primarily with the application of Chemometrics to control analysis in different industrial sectors.