Comparative Study of Multivariate Methods to Identify Paper Finishes Using Infrared Spectroscopy

Jordi-Roger Riba[1]*, Member, IEEE, Trini Canals[2], Rosa Cantero[2]

[1]Escola d'Enginyeria d'Igualada, Universitat Politècnica de Catalunya, Electrical Engineering Department.

[2]Escola d'Enginyeria d'Igualada, Universitat Politècnica de Catalunya.

Plaça del Rei 15, 08700 Igualada, Catalunya, Spain

*Corresponding author: jordi.riba@eei.upc.edu

Abstract –**Recycled paper is extensively used worldwide. In the last decades its market has expanded considerably. The increasing use of recycled paper in papermaking has led to the production of paper containing several types of impurities. Consequently, wastepaper mills are forced to implement quality control schemes for evaluating the incoming wastepaper stock, thus guarantying the specifications of the final product. The main objective of this work is to present a fast and reliable system for identifying different paper types. Therefore, undesirable paper types can be refused, improving the performance of the paper machine and the final quality of the paper manufactured. For this purpose two fast techniques, i.e., Fourier transform mid-infrared (FTIR) and reflectance near-infrared (NIR) were applied to acquire the infrared spectra of the paper samples. Next, four processing multivariate methods, i.e., principal component analysis (PCA), canonical variate analysis (CVA), extended canonical variate analysis (ECVA) and support vector machines (SVM) were employed in the feature extraction –or dimension reduction– stage. Afterwards, the *k* nearest neighbors algorithm (*k*NN) was used in the classification phase. Experimental results show the usefulness of the proposed methodology and the potential of both FTIR and NIR spectroscopic methods. Using the FTIR spectrum in association with SVM and *k*NN the system achieved maximum classification accuracy of 100%, whereas using the NIR spectrum in association with ECVA or SVM and *k*NN the system achieved maximum classification accuracy of 96.4%.**

Index Terms – **Infrared spectroscopy, multivariate methods, paper finish, process control, quality improvement.**

I. INTRODUCTION

Nowadays the paper industry is highly productive and competitive due to the reduction of production costs and the use of highly automated continuous processes, thus minimizing chemicals consumption and labor. Moreover, the demand of adjusting processes to environmental protection requirements have brought about significant improvement in raw materials, production technology, process control and end-product quality. Paper-making mills recycle waste and raw materials for several decades [1]. It is well known that the paper sector has promoted the use of recycled paper and currently nearly two thirds of all paper consumed in Europe and in the U.S. is recovered for recycling [2-3].

The use of large amounts of recycled paper supposes a challenge for the paper industry because it requires the development of innovative and more strictly-controlled production processes to guarantee high quality specifications in the finished products. Thus, in order to ensure these specifications, wastepaper mills are forced to develop fast and reliable quality controls to evaluate the incoming wastepaper stock [4]. Currently, incoming wastepaper presents a broad variety of contaminants, which must be removed in order to meet the high quality standards of the final product. An unsuccessful contaminant removal reduces dramatically the paper machine efficiency and the quality of the final product, thus compromising the economic viability of the recycling process [1,5]. Unfortunately, removal of 100% of contaminants during processing is not possible. Consequently, it is essential to establish a criterion for selecting acceptable and unacceptable wastepaper. For example, some wastepaper mills that produce uncoated paper do not accept incoming coated paper because the adhesives in the coatings can cause white pitch deposition problems on the paper machine [5]. Therefore, for these wastepaper mills it is highly desirable to have a fast tool to identify coated paper and reject it, because this tool may play an important role in improving paper machine performance and paper quality. However, it is a very complex problem because of the presence of numerous compounds, which have a multidimensional contribution in determining the final properties of paper.

The paper industry has applied different spectroscopic techniques including infrared, Raman or X-ray photoelectron spectroscopy to characterize paper finishes. However, due to the complex composition of finished paper it is difficult to relate the information provided by these techniques with the industrial processes that determine the quality of the final product [6].

Infrared spectroscopy is an essential tool for studying paper structure and pulp chemistry [7-9]. It has been long used by the paper industry for noninvasive process control and for fast determination of specific parameters, including grammage and moisture content [10-16]. Infrared spectroscopy has been applied successfully in other areas, such as for measuring different ripeness parameters in wine grapes [17] and for monitoring the quality of dairy products [18] among others.

The primary aim of this work was to develop an effective method for the non-invasive, direct identification and classification of paper finishes from the infrared spectra. As explained, the incoming paper usually may have different origins, compositions and manufacturers, thus its composition may be extremely varied. Therefore a major challenge of this work is to deal with large and heterogeneous data sets since each sample has its particular composition and therefore the problem is not based in identifying a particular compound but a combination of different products.

In this work it is proved that with one instrument, an appropriately selected set of paper samples and by applying adequate dimension reduction and multivariate classification algorithms it is possible to perform a fast classification of different types of paper finish samples.

To carry out this task, a large number of samples supplied in various finishes -coated, offset and cast-coated- by diverse paper manufacturers were analyzed by means of Fourier transform mid-infrared (FTIR) and reflectance near-infrared (NIR) spectroscopy. The performance of both FTIR and NIR spectroscopic methods was compared. There was no need for sample pretreatment or reagent addition. The suggested instrumental spectroscopic methods (NIR and FTIR) avoid the need for the tedious analyses currently carried out to determine the content of different analytes in the paper samples. Obviously, the use of an analytic classical method presents some drawbacks: it is a time-consuming method, uses chemical products, requires a qualified laboratory technician and laboratory grade facilities and it is very difficult to automate. Therefore, it is difficult to be implemented in situ and is not able to provide on-line data required for fast identification and classification of paper samples.

Both FTIR and NIR spectroscopic methods provide a large amount of data. Clearly this is a challenging task for feature extraction and classification algorithms because of the inherent heterogeneous origin of the paper samples and the large amount of information provided by NIR and FTIR instrumental techniques. Therefore an expeditious multivariate statistical processing of such amount of data is essential in order to condense the analytically relevant information in a reduced set of latent variables [9,11,12,13,19], thus discarding most of the noise and perturbations usually present in the raw signal. For this purpose the performance of different dimension reduction algorithms–including principal component analysis (PCA), canonical variate analysis (CVA), extended canonical variate analysis (ECVA) and support vector machines (SVM)– were compared. Afterwards the *k* nearest neighbor algorithm (*k*NN) was applied in the classification stage. *k*NN provides as many outputs –in the range between 0 and 1– as classes, paper finish types, in the problem. This algorithm assigns an incoming paper sample to the type of finished paper with highest output. Fig. 1 summarizes the methodology applied in this work.

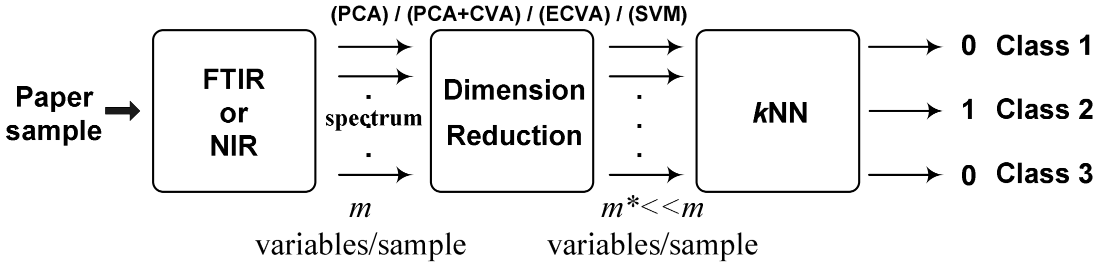

Fig. 1. Summary of the methodology applied to classify an input paper sample into different types of paper finishes.

Thus, different types of coated papers were identified by applying infrared spectroscopy jointly with multivariate statistical methods. The advantages of the system proposed here include high-speed results output, no sample pretreatment and no consumption of chemicals and reagents.

A challenge for wastepaper mill technicians is to apply infrared spectroscopic techniques jointly with multivariate classification algorithms in this pioneering stage to carry out a fast and accurate classification of paper finish samples.

II. INFRARED SPECTROSCOPY

The infrared spectral region is divided into different regions owing to instrumental and radiation interaction reasons. These regions are the near infrared (NIR, 800-2500 nm or 12500-4000 cm^{-1}), the mid-infrared (MIR, 2500-25000 nm or 4000-400 cm^{-1}) and the far infrared region (FIR, 50000-10^{6} nm or 200-10 cm^{-1}) [20]. When a beam of infrared light interacts with a sample, this absorbs energy (photons) from the incident field. The sample molecules absorb energy from the incident beam at the frequencies matching their vibration modes. Hence, by studying the transmitted or reflected spectrum it is possible to obtain information about the molecular structure of the sample. This absorption is usually measured as a function of the frequency, thus obtaining the spectrum. Solid samples are frequently analyzed by means of reflectance measurements.

Infrared spectroscopy allows identifying organic, inorganic, polymeric and biological molecules. Paper is composed of cellulose, binders and inorganic fillers among others. However, the majority of bands in a paper spectrum are due to cellulose which tend to mask the information about the products related to paper finish [21]. Hence, the selection and extraction of relevant information about the products applied during the paper finishing stage is a demanding task that requires suitable processing algorithms.

NIR and MIR are non-destructive and noninvasive instrumental methods. Appealing features of such methods include very fast response, small sample size requirements, they are environmentally friendly techniques, let in situ measurements, minimal or null sample preparation and reasonably cost per sample in regular use among others.

An important advantage of NIR over MIR is that NIR radiation has longer penetration depth than MIR light [19, 20]. However, most NIR apparatus present limited sensitivity, not being able to measure components with content less than 1% [20].

NIR spectra usually consist of broad overlapping absorption bands mostly due to overtones and combinations of vibrational modes involving different chemical bonds. Similarly, MIR spectroscopy measures fundamental molecular vibrations, which occur in the MIR spectral range. Unlike to what happens with NIR absorption bands which tend to be overlapping, broad and weak, MIR absorption bands are usually well-resolved and narrow.

However, in case of paper samples, NIR tends to mix the spectral data from the finish layer (superficial layer) with those from the inside of the paper sample. Contrarily, MIR provides better specificity and has a very short penetration depth, usually a few micrometers. NIR instruments with fiber optic probes allow acquiring the spectra of an untreated paper sample by placing the probe on the paper sample.

FTIR spectrometers operate in the MIR region and convert the data emerging from a Michelson interferometer with a movable mirror, into the spectrum by applying the Fourier transform. FTIR instruments frequently use an attenuated total reflectance (ATR) module which allows registering the reflectance spectrum of the paper sample by applying a slight pressure. It is based on measuring the changes that occur in the attenuated radiation due to the total internal reflection phenomenon produced when a MIR beam comes into contact with a paper sample by using a robust and chemically inert crystal.

III. Mathematical Background

The spectral response of both FTIR and NIR techniques for each input paper sample is composed of several hundred (or thousand) variables (absorbances or reflectances at each wavelength of the spectrum). Thus it is mandatory to concentrate the relevant information provided by this vast set of variables in a reduced set of latent variables, while retaining the most useful information to carry out the classification scheme. This constitutes the dimension reduction or feature extraction stage.

A) Principal Component Analysis (PCA)

PCA is an unsupervised multivariate technique widely applied to simplify structure in complex information [11,13,22]. Its main goal consists on extracting relevant information from data sets containing a large number of interrelated variables by reducing the dimensions of the original data set [23]. Therefore, PCA allows the dimensionality of data to be reduced while retaining as much information contained in them as possible. PCA transforms the original measured variables (spectral data obtained at different wavelengths in this case) into latent variables called principal components (PCs). PCs are linear combinations of the original measured variables. The first principal component (PC1) explains the greatest amount of total variance while the second (PC2) explains the greatest residual variance, and so on until the variance is totally explained. In practice, it is sufficient to retain only those first few PCs that explain a high enough proportion of the total variance. There is no universal method to determine the optimum number of PCs to be retained. In any case, overfitting can always be avoided by splitting samples into a training/calibration set and a test/prediction set [24]. Overfitting occurs when a too complex calibration model explains the noise jointly with the fundamental signal. An overfitted model fits the calibration data accurately but has poor predictive capability in new data, overstating slight variations in the new data. Hence, overfitting may be avoided by carefully determining the number of PCs to retain, thus reducing the complexity of the calibration model.

B) Canonical Variate Analysis (CVA)

CVA [16,25,26] is a supervised discriminant technique specially designed to accentuate differences between data classes. Thus, CVA estimates the directions in space that maximize the differences between classes in the original data according to a statistical criterion. CVA relies on discrimination criteria (classes separation) unlike PCA, which is based on regression criteria [25,27]. The CVA algorithm projects the original data into new axes called canonical variables (CVs), which are latent variables not necessarily orthogonal to one another. The separation criterion aims at obtaining the maximum separation between classes and the minimum separation within classes. However, CVA presents the disadvantage that it cannot directly deal with data where the number of variables is greater than that of samples. Thus, in these cases a dimensionality reduction stage (e.g. PCA) is required before CVA is applied.

C) Extended Canonical Variate Analysis (ECVA)

The ECVA algorithm represents an improvement of the classical CVA, since ECVA allows processing of data when the number of variables exceeds that of available samples. ECVA enables direct calculation of latent or canonical variates (ECVs) by simplifying the overall structure of the calculations without the need to previously reduce the dimensions of the problem. The number of latent variables arising from both CVA and ECVA are equal to the number of classes in the problem minus one. Details about this algorithm can be found in [26,28].

D) Support Vector Machines (SVM)

Support vector machines (SVM) are a set of supervised learning methods which are usually applied for classification or regression purposes. An *n*-class problem has *n*-1 decision boundaries may provide as output a set of unobserved latent variables or support variables, SVs. The number of SVs depends on the multi-class approach applied: one-against-one, one-against-all, minimum output code or error correcting output codes. Appealing features of the SVM algorithm include its ability of separating classes which cannot be separated by applying a linear classifier, accurate results using a small number of training sample and high accuracy at reasonably low computational burden [29].

Although the SVM is usually applied as a classifier, in this work it has been applied as a feature extraction method (like PCA, CVA or ECVA). In this way the latent variables outputted by the SVM algorithm (SVs) have been used as input of the *k*NN classifier. Hence, this system takes advantage of the power of both SVM and *k*NN algorithms. This allows comparing in the same conditions the ability of the four dimension reduction algorithms analyzed in this work.

E) k-Nearest Neighbor (kNN) Classifier

The nonparametric *k* nearest neighbor (*k*NN) classifier is one of the simplest and most widely used classification methods [30]. Despite its simplicity, *k*NN is one of the classifiers showing better performance [31,32]. It uses the majority voting rule, taking into account a weighted vote of the *k* nearest neighbors of the calibration set. The method is implemented by determining the *k* nearest neighbors for each object in the prediction set and assigning a score *k* to the class of the nearest neighbor, *k* - 1 to that of the second nearest and so on down to a score of 1. Finally, the analyzed object is assigned to the class with the highest score. Some authors have recommended using *k* values from 3 to 5 [16]. Appealing features of this algorithm include improved classification accuracy, and that the only user specified parameters are the value of *k* and the distance metric (usually the Euclidean distance) [31]. The output of the *k*NN algorithm consists of *n* variables (*n* is the number of classes) which value or degree of membership of each prediction sample to each class is comprised between 0 and 1.

IV. Data Acquisition and Paper Samples

This section describes the FTIR and NIR spectrometers used and the papers samples analyzed.

A) Data acquisition

Spectra for the whole set of paper samples dealt with were obtained by FTIR and NIR instrumental techniques. NIR and MIR spectra were acquired at room temperature (25 ± 1ºC). Fig. 2 details the two experimental set-ups.

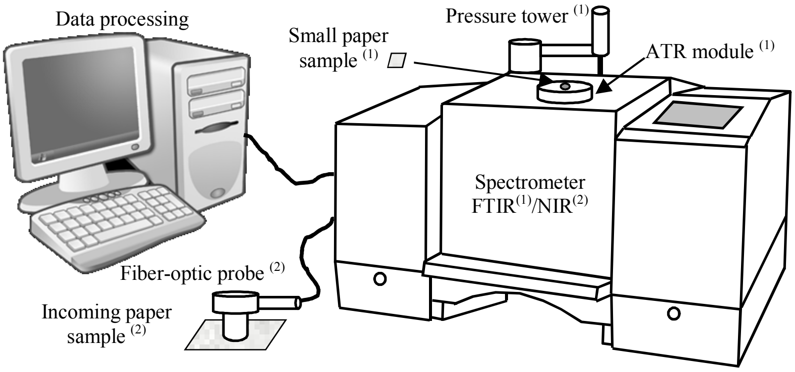

Fig. 2. Experimental set-ups. The FTIR spectrometer (1) is equipped with an ATR module whereas the NIR spectrometer (2) is equipped with a fiber-optic probe.

FTIR spectra were recorded over the range 4000-650 cm^{-1} with a resolution of 2 cm^{-1} using an IR Spectrum One (S/N 57458) from Perkin-Elmer (Beaconsfield, UK) equipped with an ATR internal reflectance module (Universal Sampling Accessory, S/N P0DL01101418) with a diamond crystal. FTIR spectra were recorded by placing a small-size sample of the incoming paper in direct contact with the ATR crystal. The sample is clamped to the crystal surface by applying pressure by means of a pressure device shown in Fig. 2. FTIR spectra were measured without any treatment of the paper samples. Each result was the average of four readings as suggested by the manufacturer.

NIR reflectance spectra were obtained in the absorbance mode, using a fiber-optic probe over the wavelength range 1100-2500 nm with a resolution of 2 nm using a Foss NIRSystems 5000 instrument (Silver Spring, MD) equipped with a reflectance detector and a fiber-optic probe. NIR spectra were measured by applying the fiber-optic probe perpendicular to the incoming paper sample and in direct contact with it. NIR spectra were recorded without any treatment of the paper samples. Each measurement was the average of 32 scans as suggested by the manufacturer.

Spectral data in the absorbance mode were sent to the computer which processes the information and they were converted into their first and second derivatives by means of the Savitzky–Golay routine. A five-point moving average was obtained at each point in order to avoid diminishing the signal-to-noise ratio during differentiation. Next, the algorithms described in Section III were applied to the pretreated spectral data.

B) Paper samples

A total amount of 92 paper samples whose origin and composition were known were used. They were split into three classes according to finish type: coated (43), cast-coated (24) and offset (25). The samples were supplied by different Spanish firms. The whole body of samples was split into a calibration set and a prediction set in order to evaluate the classification models.

The FTIR spectra for the 92 paper samples provided a data matrix consisting of 92 rows and 1676 columns which was used to obtain a 92x1666 first derivative matrix and a 92x1656 second derivative matrix. Next, the samples were divided into a calibration set and a prediction set. To this end, 29 samples were randomly included in the prediction set (14 of the 43 coated paper samples, 7 of the 25 offset samples and 8 of the 24 cast-coated samples), the remaining 63 being included in the calibration set.

Six of the 92 samples in the NIR matrix -either metallized or black-colored- were discarded due to saturation of detector response. Therefore, the starting NIR data matrix consisted of 86 rows and 550 columns obtained in the absorbance mode. This matrix was used to obtain a first derivative matrix (86x540) and a second derivative one (86x530). Next, 28 samples were randomly selected for inclusion in the prediction set (14 of the 43 coated paper samples, 7 of the 21 offset samples and 7 of the 22 cast-coated samples), the remaining 58 being included in the calibration set.

V. Results and Discussion

The finish differences in the paper samples lead to variations in physical and chemical properties, and consequently in spectral characteristics. Thus, it is supposed that by processing the spectral data by means of appropriate statistical methods, the paper samples may be identified and classified according to finish type.

We tested various types of models and carefully examined the effects of potentially influential variables such as data processing methods and wavelength ranges. This allowed the most suitable models for classifying and identifying paper finishes in the prediction set to be established. After this study the best results were achieved when dealing with FTIR spectral data with a zero mean preprocessing. In case of NIR spectrum, best results were found after performing a first derivative of the spectrum in the absorbance mode followed by a zero mean preprocessing.

A) Classification of paper samples from the FTIR spectra

As explained, the 92 samples were randomly split into a calibration set and a prediction set (containing 68.5% and 31.5% of the total samples, respectively). Therefore the starting calibration and prediction set matrixes consisted of 63x1676 and 29x1676 elements, respectively. Four feature extraction methods were applied, namely PCA, PCA+CVA, ECVA and SVM. Note that CVA requires the prior reduction of the number of variables, which was accomplished by a previous reduction of dimensions performed by means of PCA. Results are detailed below.

First, the performance of PCA processing followed by the *k*NN classification stage is evaluated. More than 99.5% of the total variance in the spectral data is available with the first 9 PCs. Thus the PCA algorithm concentrates most of the information provided by the 1676 raw variables in only 9 PCs, thus allowing reducing significantly the dimensionality of the problem. A 99.3% classification rate of the prediction samples set was achieved in this case. Fig. 3 plots the calibration and prediction samples in the space defined by only the three first PCs.

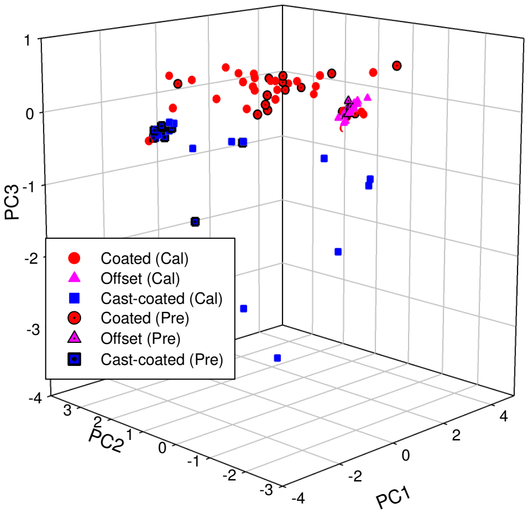

Fig. 3. FTIR spectrum. Calibration and prediction samples plotted in the space defined by the three first PCs arising from the PCA algorithm. This plot is a partial view since only three PCs are plotted out of a total of nine.

The performance of the CVA algorithm followed by the *k*NN classification stage was also examined. As explained, CVA requires the prior reduction of the number of variables, which has been carried out by means of PCA retaining the first 9 PCs. This scheme lead to a mean classification rate (prediction samples set) superior than 96.5%. When dealing with three classes, the CVA calculates two CVs. Fig. 4 shows the calibration and prediction samples in the space defined by the two CVs.

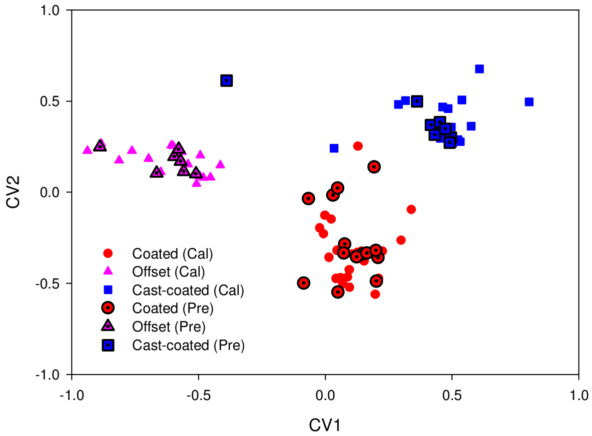

Fig. 4. FTIR spectrum. Calibration and prediction samples plotted in the space defined by the CVs arising from the PCA (9PCs)+CVA algorithm.

Now, the behavior of the ECVA + *k*NN algorithm is analyzed. As explained, ECVA does not require the prior reduction of the number of variables and in case of a three-class problem it calculates only two ECVs. This scheme lead to a mean classification rate of the prediction samples set superior than 96.5%. Fig. 5 shows the calibration and prediction samples in the space defined by the two ECVs.

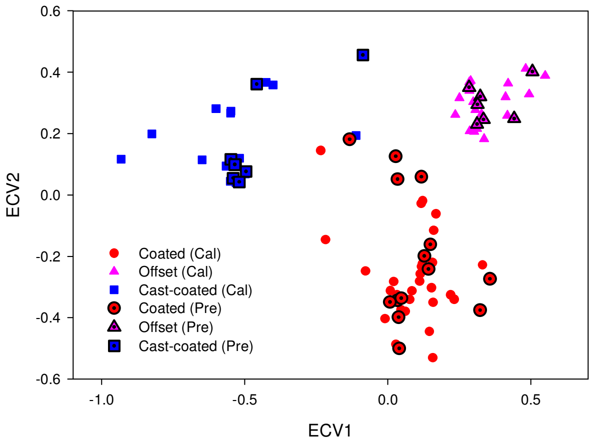

Fig. 5. FTIR spectrum. Calibration and prediction samples plotted in the space defined by the ECVs arising from the ECVA algorithm.

Finally, the behavior of the SVM + kNN method was studied. After careful selection of SVM-related options such as kernel and training types and different multi-class approaches, the SVM algorithm based on a RBF (Radial Basis Function) kernel, cross-validation training and multi-class one-against-all classification was applied. This scheme lead to 100% classification rate. Fig. 6 shows the calibration and prediction samples in the space defined by the three SVs.

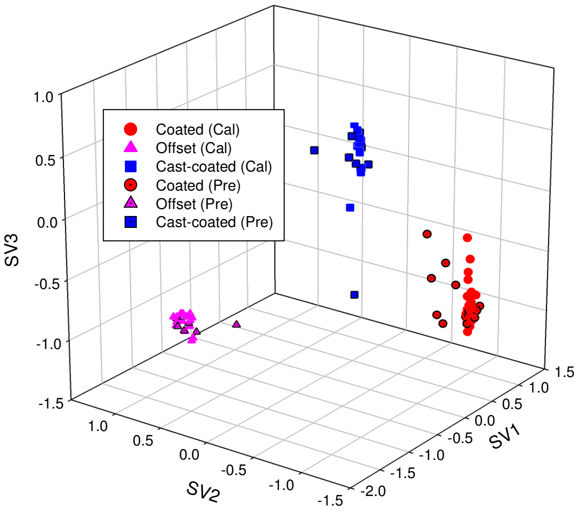

Fig. 6. FTIR spectrum. Calibration and prediction samples plotted in the space defined by the three SVs arising from the SVM algorithm.

B) Classification of paper samples from the NIR spectra

Six of the 92 samples in the NIR matrix were discarded. The 86 remaining samples were randomly split into a calibration set and a prediction set (containing 67.4% and 32.6% samples of the total samples, respectively). The primary matrix consisted of 550 columns per paper sample of NIR data obtained in the absorbance mode. Best results were achieved when using the first derivative of this matrix, obtaining 540 variables per sample. Thus, the matrixes of the calibration and prediction sets consisted of 58x540 and 28x540 elements, respectively.

First, the performance of PCA processing followed by the *k*NN classification stage is tested. More than 99.5% of the total variance in the spectral data is available with the first 8 PCs. A mean 86.9% classification rate in the prediction samples set was achieved in this case. Fig. 7 shows plots the calibration and prediction samples in the space defined only by the three first PCs.

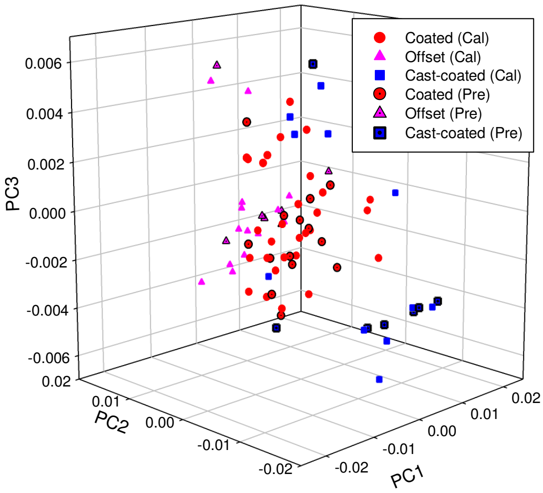

Fig. 7. NIR spectrum. Calibration and prediction samples plotted in the space defined by the three first PCs arising from the PCA algorithm. This plot is a partial view since only the three first PCs are plotted out of a total of eight.

The behavior of the PCA+CVA algorithm followed by the *k*NN method was also examined. The CVA requires the prior reduction of the number of variables, which has been carried out by means of PCA, retaining the first 8 PCs. This scheme lead to a mean classification rate of the prediction samples set superior than 94.0%. Fig. 8 shows the calibration and prediction samples in the space defined by the two CVs calculated by the CVA algorithm.

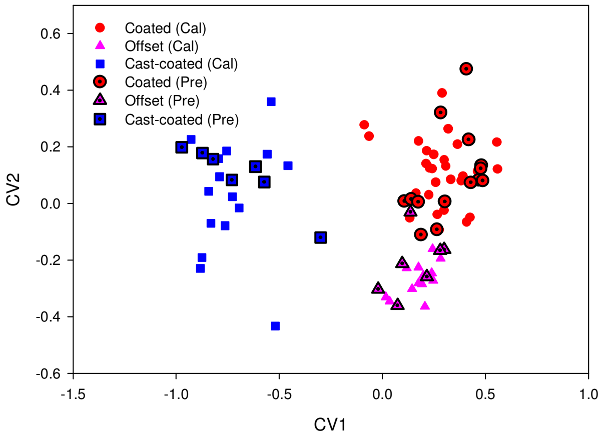

Fig. 8. NIR spectrum. Calibration and prediction samples plotted in the space defined by the CVs arising from the PCA (8PCs)+CVA algorithms.

Now, the performance of the ECVA + *k*NN algorithm is analyzed. This scheme lead to a mean classification rate of the prediction samples set greater than 95.2%. Fig. 9 shows the calibration and prediction samples in the space defined by the two ECVs.

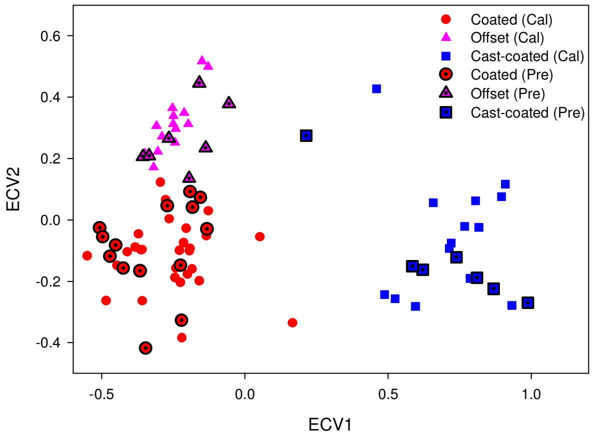

Fig. 9. NIR spectrum. Calibration and prediction samples plotted in the space defined by the ECVs computed by the ECVA algorithm.

Finally, the behavior of the SVM + *k*NN method is studied. This scheme lead to 96.4% classification rate. Fig. 10 shows the calibration and prediction samples in the space defined by the three SVs. As done with the FTIR samples, after careful selection of the SVM-related options, the SVM algorithm based on a RBF kernel, cross-validation training and multi-class one-against-all classification was applied.

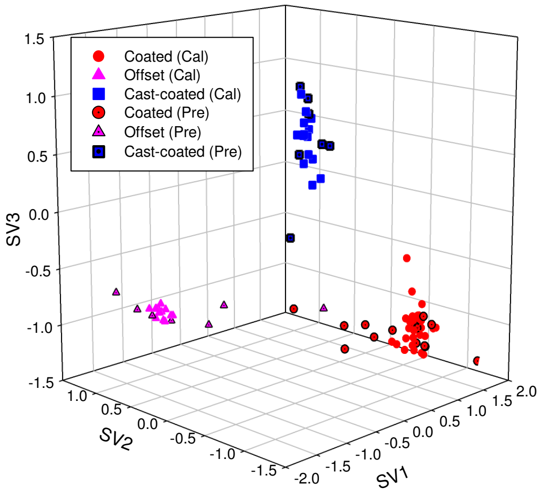

Fig. 10. NIR spectrum. Calibration and prediction samples plotted in the space defined by the three SVs arising from the SVM algorithm.

C) Results Summary

Table I summarizes the classification results obtained by means of the analyzed statistical methods.

TABLE I
RESULTS SUMMARY

Spectrum	Statistical methods	Classifier	Prediction success rate
FTIR	PCA	3NN	27/29 (93.10%)
		4NN	27/29 (93.10%)
		5NN	27/29 (93.10%)
	PCA + CVA	3NN	28/29 (96.55%)
		4NN	28/29 (96.55%)
		5NN	28/29 (96.55%)
	ECVA	3NN	27/29 (93.10%)
		4NN	28/29 (96.55%)
		5NN	28/29 (96.55%)
	SVM	3NN	29/29 (100.0%)
		4NN	29/29 (100.0%)
		5NN	29/29 (100.0%)
NI	PCA	3NN	24/28 (85.71%)
		4NN	24/28 (85.71%)
		5NN	25/28 (89.29%)
	PCA + CVA	3NN	26/28 (92.86%)
		4NN	26/28 (92.86%)
		5NN	27/28 (96.43%)
		3NN	27/28 (96.43%)

R	ECVA	4NN	27/28 (96.43%)
		5NN	26/28 (92.86%)
	SVM	3NN	27/28 (96.43%)
		4NN	27/28 (96.43%)
		5NN	27/28 (96.43%)

Results shown in Table I clearly indicate that the classification results obtained from the FTIR spectrum are slightly better that those obtained from the NIR spectrum.

The worst results were obtained when using the PCA + *k*NN scheme. It is so because PCA is an unsupervised method optimized for regression purposes, while the others (CVA, ECVA and SVM) are supervised methods based on classes separation criteria. Additionally, the best results were attained by means of the SVM algorithm.

VI. Conclusions

This work presents a system for selecting acceptable and unacceptable recycled paper in the incoming of wastepaper mills. Features of the proposed method include quasi-immediate response and high identification accuracy. As explained, wastepaper mills that manufacture uncoated paper and operate an on-site paper recycling plant do not accept coated paper in their incoming because the adhesives in the coatings can cause white pitch deposition problems on the paper machine. In this case the rejection of coated paper could help limiting the amount of fillers and adhesives in the recycled fiber stream. Therefore, the proposed system may be very useful to guarantee appropriate runnability of the paper machine as well as to meet high quality specifications –especially strength and stiffness– of the final product.

Both FTIR and NIR are non-invasive methods widely used for controlling different quality related parameters in several industrial sectors because they allow fast acquisition of spectral information and provide useful information about the composition of the analyzed samples. In this work their use has been extended by developing a methodology for identifying paper types by means of a fast statistical processing of the acquired FTIR or NIR spectra.

The experimental results presented in this work prove that for the 92 paper samples with different finishes dealt with, both FTIR and NIR methods provide excellent results. However, classification results obtained from the FTIR spectrum are slightly better than those obtained from the NIR spectrum. Near infrared light (NIR) penetrates more deeply into the paper surface than does mid-infrared light (FTIR). This may result in partial overlaps of the information extracted from the paper matrix with that obtained from the finish. This is a drawback of NIR spectroscopy in this particular application, which may be overcome by a careful selection of the feature extraction and classification algorithms applied in the calibration stage.

The four analyzed multivariate feature extraction methods (PCA, PCA+CVA, ECVA and SVM) present a very fast response (very similar in all cases) while providing a high success rate when applying the *k*NN classifier algorithm in classifying prediction samples from FTIR and NIR spectra. Consequently, they can be applied successfully to develop quality controls for fast and reliable evaluation of the incoming wastepaper stock. It is worth mentioning that the best performing algorithm has been the SVM, using both FTIR and NIR spectral data.

Acknowledgments

The authors are grateful to the firms Sarriópapel y Celulosa S.A. (Barcelona, Spain), Torraspapel S.A. (Barcelona, Spain) and Grup d'Impressió S.L. (Vilanova del Camí, Spain) for supplying the paper samples used in this work.

VII. References

[1]R.W.J. McKinney, *Technology of Paper Recycling,* Surrey, England: Chapman & Hall, Blackie Academic & Professional, chapter 3, 1997.

[2]European Recovered Paper Council, *European Declaration on Paper Recycling 2006–2010. Monitoring Report 2009*, Brussels, Belgium: European Recovered Paper Council, 2009.

[3]American Forest & Paper Association, *AF&PA Announces Increase in Paper Recovery to a Record 63.4 Percent, Industry Meets Goal Ahead of Schedule*, Washington, USA: http://www.afandpa.org/pressreleases.aspx?id=1316, Access 11 March 2011.

[4]M. Doshi, *Recycled Paper Technology. An Anthology of Published Papers* Atlanta, USA: TAPPI Press, pp 12–8, 67–76, 86–9, 1994.

[5]J.-R. Riba, T. Canals, R. Cantero and H. Iturriaga, “Potential of infrared spectroscopy in combination with extended canonical variate analysis for identifying different paper types,“ *Measurement Science and Technology*, vol. 22, no. 2, pp. 1-7, 2011.

[6]J. Vyörykkä, A. Fogden, J. Daicic, M. Ernstsson and A.-S. Jääskeläinen, “Characterization of Paper Coatings – Review and Future Possibilities,” in *Proc. of the 9th TAPPI Advanced Coating Fundamentals Symposium*, Turku, Finland, Feb. 2006. pp. 41-46.

[7]J.J.Workman, “Review of Process and Non-invasive Near-Infrared and Infrared Spectroscopy: 1993–1999,” *Applied Spectroscopic Reviews*, vol. 34, no. 1 & 2, pp. 1 – 89, 1999.

[8]J.J.Workman, “Infrared and Raman Spectroscopy in Paper and Pulp Analysis,” *Applied Spectroscopic Reviews*, vol. 36, no. 2 & 3, pp. 139–168, 2001.

[9]J. Pan and K.L. Nguyen, “Development of the Photoacoustic Rapid-Scan FT-IR-Based Method for Measurement of Ink Concentration on Printed Paper,” *Analytical Chemistry*, vol. 79, no. 6, pp. 2259-2265, Feb. 2007.

[10] R. Hodges, H. Cullinan and 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, Nov. 2006.

[11] M.T. Bona and 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, Jan. 2008.

[12] J.S. Câmara, M.A. Alves, J.C. Marques "Multivariate analysis for the classification and differentiation of Madeira wines according to main grape varieties," Talanta, vol. 68, no. 5, pp. 1512-1521, Feb. 2006S.

[13] S. López-Feria, Cárdenas, J.A. García-Mesa and M. Valcárcel, "Classification of extra virgin olive oils according to the protected designation of origin, olive variety and geographical origin," *Talanta*, vol. 75, no. 4, pp. 937–4, May 2008.

[14] D. Anderson, "Coatings," *Anal. Chem*., vol. 73, no. 12, pp. 2701–2704, June 2001.

[15] O. Berntsson, L.G. Danielsson and S. Folestad, "Estimation of effective sample size when analysing powders with diffuse reflectance near-infrared spectrometry," *Anal. Chim. Acta*, vol. 364, no. 1-3, pp. 243–51, May 1998.

[16] 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, July 2007.

[17] M. Larraín, Andrés R. Guesalaga, and Eduardo Agosín, "A Multipurpose Portable Instrument for Determining Ripeness in Wine Grapes Using NIR Spectroscopy," *IEEE Trans. on Instrumentation and Measurement*, vol. 57, no. 2, pp. 294-302, Feb. 2008.

[18] S.C. Mukhopadhyay, C.P. Gooneratne, G.S. Gupta, and S.N. Demidenko, "A Low-Cost Sensing System for Quality Monitoring of Dairy Products," *IEEE Trans. on Instrumentation and Measurement*, vol. 55, no. 4, pp. 1331-1338, Aug. 2006.

[19] L. Dolmatova, C. Ruckebusch, N. Dupuy, J.P. Huvenne, P. Legrand, "Quantitative analysis of paper coatings using artificial neural networks Original Research Article," *Chemometrics and Intelligent Laboratory Systems*, vol. 36, Issue 2, pp. 125-140, April 1997.

[20] D.-W. Sun, Infrared spectroscopy for food quality analysis and control. Burlington, USA: Academic Press, Elsevier Inc, 2009

[21] B. Stuart, Infrared Spectroscopy: Fundamentals and Applications. Chichester, England: John Wiley & Sons, 2004

[22] N. Bhattacharyya, R. Bandyopadhyay, M. Bhuyan, B. Tudu, D. Ghosh, and A. Jana, "Electronic Nose for Black Tea Classification and Correlation of Measurements With "Tea Taster" Marks," *IEEE Trans. on Instrumentation And Measurement*, vol. 57, no. 7, pp. 1313-1321, July 2008.

[23] X. Wang, N.D. Georganas, and E.M. Petriu, "Fabric Texture Analysis Using Computer Vision Techniques," *IEEE Trans. on Instrumentation And Measurement*, vol. 60, no. 1, pp. 44-56, Jan. 2011.

[24] W.J. Krzanowski, *Principles of Multivariate Analysis. A User's Perspective,* 2nd edn. New York, USA: Oxford University Press, pp 53–75, 2000.

[25] R.A. Johnson and D.W. Wichern, Applied *Multivariate Statistical Analysis,* 6th edn. Englewood Cliffs, USA: Prentice-Hall, chapters 8–10, 2007

[26] 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, August-October 2006.

[27] M.M. Paradkar, S. Sivakesava, and J. Irudayaraj, "Discrimination and classification of adulterants in maple syrup with the use of infrared spectroscopic techniques," *J. Sci. Food Agric.,* vol. 82, no. 5, pp. 497–504, April 2002.

[28] 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.

[29] H. Zareipour, A. Janjani, H. Leung, A. Motamedi, and A. Schellenberg, “Classification of Future Electricity Market Prices,” *IEEE Trans. on Power Systems*, vol. 26, no. 1, pp. 165 – 173, Feb. 2011.

[30] M. Li, M.M. Crawford, and T. Jinwen, “Local Manifold Learning-Based k -Nearest-Neighbor for Hyperspectral Image Classification,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4099 – 4109, Nov. 2010.

[31] P. Ciosek, and W. Wróblewskia, “The analysis of sensor array data with various pattern recognition techniques,” *Sensors and Actuators B: Chemical*, vol. 114, no. 1, pp. 85-93, March 2006.

[32] K. Choi, S. Singh, A. Kodali, K.R. Pattipati, J.W. Sheppard, S.M. Namburu, S. Chigusa, D.V. Prokhorov, and L. Qiao, “Novel Classifier Fusion Approaches for Fault Diagnosis in Automotive Systems,” *IEEE Trans. on Instrumentation and Measuremen*t, vol. 58 , no. 3, pp. 602 – 611, March 2009.