
Identificación de muestras de papel mediante espectrometría IR y métodos multivariados

Jordi-Roger Riba

Escola d'Enginyeria d'Igualada, Universitat Politècnica de Catalunya, Plaça del Rei 15, 08700 Igualada, Barcelona, Spain

Paper samples identification by means of IR spectrometry and multivariate methods

Identificació de mostres de paper mitjançant espectroscòpia IR i mètodes multivariats

Recibido: 29 de febrero de 2012; revisado: 17 de junio de 2012; aceptado: 3 de julio de 2012

RESUMEN

Actualmente en todo el mundo se consume una gran cantidad de papel reciclado. Ello ha provocado la fabricación de papel que contiene muchos tipos de impurezas, por lo que las empresas papeleras se están viendo con la necesidad de desarrollar métodos para controlar la calidad del papel entrante para así garantizar su calidad. Para ello en este trabajo se investiga un método muy rápido y no destructivo para identificar diferentes tipos de papel que permita desechar el papel entrante que no cumpla con unos requisitos mínimos de calidad. De este modo se puede mejorar el comportamiento de la máquina de papel y al mismo tiempo asegurar la calidad del producto final. Ello se hace en base al estudio de los espectros FTIR (espectroscopia del infrarrojo medio por transformada de Fourier) y NIR (espectroscopia del infrarrojo cercano) tratados mediante ICA (análisis de componentes independientes) como técnica multivariable de extracción de características para reducir la cantidad de variables utilizadas y k-NN (k vecinos más cercanos) como técnica de clasificación. Los resultados experimentales muestran que es posible identificar con éxito más de un 90% de las muestras estudiadas, de una forma rápida, muy automatizada y de forma no destructiva.

Palabras clave: papel, espectroscopia infrarroja, métodos multivariados, control de calidad.

SUMMARY

Nowadays recycled paper is widely consumed worldwide. This fact has led to the production of paper containing numerous kinds of impurities. Therefore, wastepaper mills are trying to apply quality control systems to assess the incoming wastepaper stock, thus ensuring the specifications of the paper manufactured. For this reason this work is devoted to analyze a fast and non-destructive method to classify different paper types. Furthermore, this method allows refusing undesirable paper types while improving the performance of the paper machine and the quality of the paper manufactured. It is carried out by analyzing the FTIR (Fourier transform mid-infrared) and NIR (reflectance near-

infrared) spectra, which are post-processed by applying the multivariate ICA (independent components analysis) method. ICA allows greatly reducing the dimensionality of the problem before applying a classifier method, the k-NN (k nearest neighbors) algorithm in this case. Experimental results prove that it is possible an accurate, fast, automated and non-destructive identification of the analyzed paper samples.

Key words: Paper, infrared spectroscopy, multivariate methods, quality control.

RESUM

Actualment a tot el món es consumeix una gran quantitat de paper reciclat. Això ha dut a fabricar paper amb un alt contingut d'impureses i ha fet que les empreses papereres es vegin forçades a desenvolupar mètodes per controlar la qualitat del paper entrant per així garantir la seva qualitat. Per aquesta finalitat en aquest treball s'investiga un mètode molt ràpid i no destructiu per identificar diferents tipus de paper i que permeti rebutjar el paper entrant que no compleixi amb uns requisits mínims de qualitat. D'aquesta manera es pot millorar el comportament de la màquina de paper, tot i assegurant la qualitat del producte final. Això es realitza en base a l'estudi dels espectres FTIR (espectroscòpia de l'infraroig mitjà per transformada de Fourier) i NIR (espectroscòpia de l'infraroig proper) tractats mitjançant ICA (anàlisi de components independents) com a tècnica multivariable d'extracció de característiques per reduir la quantitat de variables utilitzades i k-NN (k veïns més propers) com a tècnica de classificació. Els resultats experimentals mostren que és possible identificar amb èxit més d'un 90% de les mostres estudiades de manera ràpida, no automatitzada i no destructiva.

Mots clau: paper, espectroscòpia d'infraroig, mètodes multivariats, control de qualitat.

*Autor para la correspondencia: jordi.riba@eei.upc.edu

INTRODUCCIÓN

En la actualidad la industria papelera, es altamente productiva y competitiva y ha evolucionado muy positivamente en la aplicación de procesos respetuosos con el medioambiente¹⁾ juntamente con el uso de materiales reciclados desde hace varias décadas²⁾. Dos tercios de todo el papel consumido en Europa y los Estados Unidos se recupera para ser reciclado^{3,4)}. Ello supone un reto tecnológico importante debido a que se debe garantizar la calidad y adecuarse a las especificaciones del producto final. Para garantizar las especificaciones de las materias primas, que suelen incorporar diferentes variedades de contaminantes, se requieren controles de calidad rápidos y confiables⁵⁾. Estos contaminantes muchas veces deben eliminarse para garantizar la calidad del producto final. Cuando su eliminación no es viable es posible descartar ciertas partidas de papel reciclado, por lo que es indispensable el uso de un sistema de control adecuado para determinar si una materia prima es aceptable o no lo es. Por lo tanto para la industria papelera es muy importante disponer de un método rápido para detectar algunos tipos de papel reciclados indeseados y así poderlos desechar. Este es un problema complejo debido a la presencia de numerosos componentes en las formulaciones de los acabados, los cuales presentan una contribución multidimensional en las propiedades finales del papel.

En la industria papelera se han aplicado diferentes técnicas espectroscópicas para caracterizar los acabados del papel, entre ellas la espectroscopia Raman o la espectroscopia de rayos X⁶⁾ así como también la espectroscopia infrarroja⁷⁻¹⁰⁾.

Este trabajo tiene como objetivo principal el estudio, análisis y puesta a punto de un método eficaz, rápido y no destructivo para identificar y clasificar muestras de papel de orígenes y composiciones complejas y diversas a partir de su espectro infrarrojo. Ello supone un reto muy importante ya que se analizan grupos muy heterogéneos de muestras provenientes de diferentes fabricantes, con composiciones variadas y orígenes diversos. Debido a que cada muestra tiene su propia composición, el método no puede basarse en la identificación de un componente particular sino de la combinación de diferentes productos. El sistema utilizado se basa en subdividir el conjunto total de muestras disponibles en dos grupos, el de calibración y el de test o validación. Mientras que el conjunto de calibración sirve para calibrar o ajustar el modelo matemático de clasificación, el conjunto de muestras de test sirve para probar la eficacia de dicho modelo ante muestras diferentes de las de calibración.

Es conocido que la espectroscopia infrarroja permite identificar componentes orgánicos^{11,12)}, inorgánicos, poliméricos y moléculas biológicas. El papel está compuesto de celulosa, aglutinantes y cargas inorgánicas, entre otros. Pero la mayor parte de las bandas de un espectro de una muestra de papel son debidas a la celulosa, que tiende a enmascarar la información proporcionada por los productos utilizados en el acabado¹³⁾. Ello indica que se hace requisito necesario extraer toda la información relevante relacionada con dichos productos. Tanto la espectrometría FTIR como la NIR proporcionan grandes cantidades de datos por muestra analizada, siendo indispensable el uso de métodos matemáticos apropiados para reducir el número de variables del problema¹⁴⁾. En esta fase, denominada de reducción de dimensiones, se utilizan

algoritmos multivariantes que permiten condensar la información relevante contenida en un número elevado de variables iniciales (en el caso de la espectrometría infrarroja son absorbancias o reflectancias para cada longitud de onda) en un número reducido de variables latentes^{9,15-17)}. Estas variables latentes son las utilizadas por el algoritmo de clasificación. Es bien conocido que cuando se trabaja con un número muy elevado de variables los algoritmos de clasificación pierden capacidad de generalización a partir de un número crítico de variables para un número fijo de muestras de calibración¹⁸⁾. Además, la etapa de reducción de dimensiones permite reducir los requisitos computacionales asociados a la resolución del problema. Por todo lo explicado, el uso de un número reducido de variables latentes es muy práctico.

El método clásico para reducir las dimensiones del problema es PCA (análisis de componentes principales). No obstante, este trabajo se centra en analizar el comportamiento de ICA (análisis de componentes independientes, un algoritmo más apto para solucionar problemas de clasificación), comparándolo con el de PCA. Como algoritmo de clasificación se utiliza k-NN (k vecinos más cercanos) por ser uno de los clasificadores más sencillos, a la vez que aporta buenos resultados. Este algoritmo proporciona tantas salidas con valores entre 0 y 1 como clases tiene el problema (tipos de papel en este trabajo). Resumiendo, a continuación se detallan las etapas del método utilizado en este trabajo:

- Obtención de los espectros infrarrojos FTIR o NIR. Para cada muestra se obtiene un espectro con centenares o miles de datos espectrales.
- Reducción del número de variables (absorbancias). Para ello se aplica el algoritmo ICA aunque también se compara su comportamiento con el del método clásico PCA (análisis de componentes principales).
- Se clasifican las muestras mediante el algoritmo k-NN.

En este trabajo se utilizan los espectros FTIR y NIR de un total de 92 muestras de papel de diferentes fabricantes, divididas en tres grupos de acabados, llamados estucado, offset y alto brillo. A partir de los espectros y aplicando la metodología explicada se lleva a cabo una clasificación de dichas muestras según sus correspondientes tipos de acabado. El sistema propuesto tiene la ventaja de no requerir productos químicos, de utilizar las muestras de papel sin requerir tratamiento alguno, de tener una elevada velocidad de respuesta y de no dañar las muestras.

MÉTODOS MATEMÁTICOS APLICADOS

2.1 Análisis de componentes principales (PCA)

La PCA es una técnica muy utilizada para reducir las dimensiones o número de variables de un problema^{15,19,20)}. A partir de un elevado número de variables originales interrelacionadas, permite extraer información relevante, reduciendo el número de variables del problema mientras que tiende a desechar el ruido contenido en las variables originales. Por lo tanto, PCA transforma las variables originales o medidas (datos espectrales obtenidos para cada longitud de onda en el caso analizado) en un conjunto de variables latentes, llamadas componentes principales (PCs), perpendiculares entre ellos. Los PCs se obtienen a partir de combinaciones lineales de las variables originales. El primer PC (PC₁) explica la máxima variabilidad de los datos, el segundo (PC₂) que es perpendicular al pri-

mero, explica la máxima variabilidad residual y así sucesivamente. Ello indica que los PCs obtenidos se alinean en la dirección de máxima variabilidad de los datos en el espacio de las variables originales. En general se obtienen tantos PCs como variables originales tiene el problema, aunque en la práctica se suele retener un número reducido de PCs.

La selección del número adecuado de PCs es un punto delicado y no hay ningún método universal para este propósito. En general, como ya se ha explicado, las muestras se dividen en dos conjuntos, el de calibración y el de test ²¹⁾. Este sistema permite evitar el fenómeno del sobreajuste. Un sistema está sobreajustado cuando su respuesta es capaz de ajustar con gran precisión los datos de calibración pero tiene muy poca capacidad para ajustar los valores de nuevos datos. El problema del sobreajuste en general es debido a la utilización de un modelo matemático con demasiados PCs, es decir, demasiado complejo. De ello se desprende que para evitar el sobreajuste del modelo se deben elegir cuidadosamente el número de componentes principales a incluir en el modelo matemático para realizar una correcta clasificación.

Para más detalles acerca de la base matemática de del algoritmo PCA y de los métodos de selección del número adecuado de PCs se puede consultar ²⁰⁾.

2.2 Análisis de componentes independientes (ICA)

Como se ha explicado, PCA se ha usado frecuentemente para reducir las dimensiones en problemas de clasificación pero al estar basado en criterios de regresión lineal, no es óptimo para abordar problemas de clasificación ¹⁴⁾. En este trabajo se sugiere el uso de ICA para reducir la dimensionalidad del problema como alternativa eficaz a PCA.

El método ICA es muy utilizado para solucionar el problema denominado separación ciega de fuentes (BSS) ^{22,23)}. El método BSS se refiere a la estimación de señales no observadas o fuentes (variables s o componentes independientes en el contexto de ICA) a partir de señales multivariantes originales. BSS hace la suposición que las señales originales están compuestas por una mezcla de señales fuente s desconocidas, siendo la composición de dicha mezcla desconocida ²³⁾.

ICA descompone los datos originales (espectros infrarrojos) en señales estadísticamente independientes o señales fuente, llamadas ICs (componentes independientes). Estas señales fuentes presentan una distribución no gaussiana. Ello significa que en caso de disponer de m^* señales fuente, su probabilidad conjunta es factorizable ²²⁾, es decir

$$p(s_1, s_2, \dots, s_m) = p(s_1)p(s_2)\dots p(s_m).$$

Sea $x = (x_1, x_2, \dots, x_m)^t$ el vector que contiene las m variables originales o observaciones (absorbancias o reflectancias en este caso) de una muestra obtenidas de m sensores. Sea $s = (s_1, s_2, \dots, s_m)^t$ el vector que contiene m^* componentes independientes (ICs) o fuentes, donde $m^* \leq m$. Los ICs son variables centradas respecto su valor medio y estadísticamente son mutuamente independientes ²⁴⁾. Las variables originales o medidas contenidas en el vector x se asumen combinaciones lineales de los ICs contenidos en el vector s . La relación lineal entre ellos se puede expresar matricialmente como,

$$x_{(m,1)} = A_{(m,m^*)} s_{(m^*,1)} \quad (1)$$

donde $A_{(m,m^*)}$ es la matriz de mezcla. Cada uno de sus elementos A_{ij} representan los pesos individuales de la variable fuente s_j en la variable original o medida x_i . ICA estima la matriz de desmezclado W como

$$W_{(m^*,m)} = A_{(m,m^*)}^{-1} \quad (2)$$

Las estimaciones ss de las fuentes se a partir de las variables originales en el vector x como:

$$ss_{(m^*,1)} = W_{(m^*,m)} x_{(m,1)} \quad (3)$$

Una vez centrado, el vector x se blanquea, es decir, se transforma de modo que sus componentes estén descorrelacionados y su variancia esté normalizada a la unidad. Las soluciones del algoritmo ICA (estimaciones de las fuentes ss) se obtienen maximizando su no gaussianidad a partir de la medida del coeficiente de kurtosis ²³⁾. Para más detalles acerca de la base matemática de ICA se puede consultar ²⁴⁾.

2.3 El clasificador basado en los k vecinos más cercanos (k-NN)

k -NN es un algoritmo de clasificación ampliamente utilizado debido a su simplicidad y excelentes resultados ^{20,25)}. Este método aplica la técnica del voto mayoritario, a partir de una ponderación del voto de cada uno de los k vecinos más cercanos del conjunto de calibración a la muestra que se pretende clasificar. En el caso de disponer de un número reducido de muestras de calibración se suelen aplicar valores de k comprendidos entre 3 y 5 ²⁶⁾. Los pasos en que se basa este algoritmo se detallan a continuación:

1. Identificar de modo supervisado la clase o grupo al que pertenece cada una de las muestras del grupo de calibración del modelo.
2. Para cada muestra de test localizar los k vecinos más cercanos de entre las muestras de calibración.
3. Cada uno de los k vecinos proporciona un voto ponderado para decidir la clase a la que pertenece dicha muestra de test. Por ejemplo, si $k = 3$, el vecino más cercano vota con 3 puntos a su clase, el segundo más cercano con 2 puntos a su clase, y el tercero con 1 punto. Así, de los 6 puntos obtenidos por la muestra de test se cuenta la clase mayoritaria o más votada, siendo la muestra asignada a dicha clase.
4. La salida de este algoritmo consiste en tantas variables por muestra clasificada como clases tiene definidas el problema, generalmente normalizadas entre 0 y 1. Cada salida expresa el grado de pertenencia de la muestra a dicha clase.

LAS MUESTRAS DE PAPEL

Es conocido que para satisfacer las elevadas exigencias que los sectores de la impresión y del embalaje demandan a las propiedades finales de la superficie de la hoja de papel (color, textura, brillo, etc.), es necesario recubrir los papeles con formulaciones adecuadas.

En este trabajo se dispone de un total de 92 muestras de papel de origen y composición conocidos, es decir, se conoce el fabricante y el tipo de acabado de cada muestra. Estas muestras corresponden a un total de tres tipos de acabados que se han clasificado según se detalla: 43 de ellas presentan un acabado tipo estucado, 24 son de tipo

alto brillo y las 25 restantes presentan un acabado tipo offset.

Los espectros FTIR de cada una de las 92 muestra de papel se juntaron para formar una matriz numérica de 92 filas y 1676 columnas, donde las columnas representan las componentes del vector de variables originales x . Al realizar la primera derivada se obtuvo una matriz de 92×1666 componentes, mientras que en el caso de la segunda derivada de 92×1656 componentes. A continuación, cada una de estas matrices se subdividió en dos (seleccionando aleatoriamente un número proporcional de muestras de cada tipo de acabado), para obtener una matriz de calibración de 63 filas o muestras y una matriz de test de 29 filas. En el caso de los espectros NIR se operó de modo similar, obteniendo una matriz de 92×550 columnas. Se descartaron 6 de las 92 muestras por ser o bien metalizadas o negras, lo que saturó la respuesta del detector. Por lo tanto resultó una matriz total de 86×550 elementos en modo absorbancia. Dicha matriz se convirtió en modo primera derivada (86×540) y a segunda derivada (86×530). A continuación se seleccionaron aleatoriamente 28 muestras para incluirlas en el conjunto de test y las 58 restantes se utilizaron como conjunto de calibración.

Para calcular la primera y la segunda derivada de los datos espectrales se utilizó el algoritmo de Savitzky y Golay aplicando un promediado móvil de cinco puntos para no dañar la relación señal-ruido debido a la derivación numérica. A continuación se aplicaron los algoritmos matemáticos explicados en la sección 2.

ADQUISICIÓN DE DATOS

Como se ha explicado, los espectros se han obtenido mediante espectrofotometría FTIR y NIR.

Para el registro de los espectros FTIR se ha utilizado el espectrofotómetro IR Spectrum One de la marca PerkinElmer (Beaconsfield, UK) equipado con un módulo de reflexión interna ATR con cristal de diamante. Dichos espectros se encuentran en el intervalo de números de onda comprendido entre 4000 y 650 cm^{-1} . Se promediaron los resultados de cuatro lecturas consecutivas según sugere el fabricante.

Para el registro de los espectros NIR se ha utilizado un espectrofotómetro Foss NIRSystem 5000 equipado con un detector de reflectancia y una sonda de fibra óptica Ortiprobe. Dicha sonda permite adquirir los espectros en modo absorbancia en el intervalo de longitudes de onda de 1100 a 2200 nm . En cada medida se promediaron 32 registros como sugiere el fabricante. Para tener en cuenta la heterogeneidad de cada muestra de papel y de este modo obtener espectros representativos de toda la muestra, la sonda de fibra óptica se desplazó sobre la superficie del papel, realizando lecturas en tres puntos distintos elegidos aleatoriamente. El espectro de cada una de las muestras de papel se obtuvo promediando los tres registros realizados.

Todas las adquisiciones se realizaron a temperatura ambiente, $25 \pm 1 \text{ }^\circ\text{C}$.

Las figuras 1 y 2 muestran los espectros FTIR y NIR de las muestras de papel del conjunto de calibración.

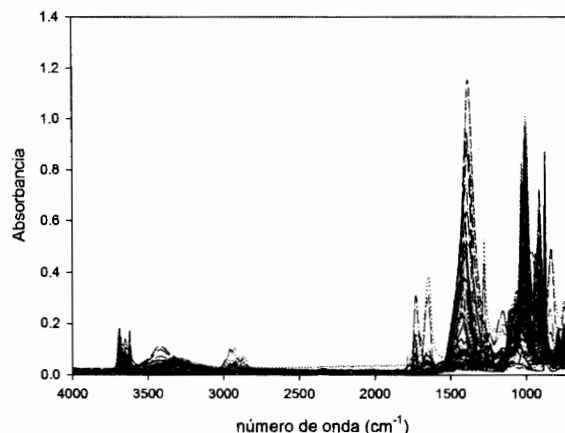


Figura 1. Espectros FTIR de las muestras de calibración utilizadas.

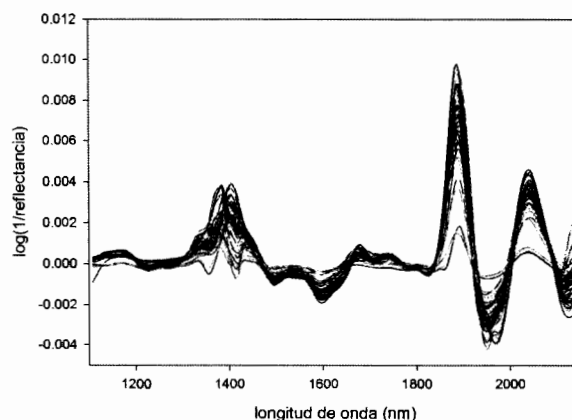


Figura 2. Primera derivada de los espectros NIR en modo absorbancia de las muestras de calibración analizadas

RESULTADOS Y DISCUSIÓN

En este trabajo se asume que la diferencia de composición entre las tres clases de acabados analizadas se ve reflejada en los espectros FTIR y NIR. Si es así, los métodos estadísticos utilizados utilizarán esta información para separar las muestras procesadas.

Después de analizar cuidadosamente diferentes modelos se constató que en el caso de los espectros FTIR, se obtenían los mejores resultados con los datos espectrales sin derivar y centrados respecto su media. Por el contrario, en el caso de los espectros NIR, se debe utilizar la primera derivada de los espectros en modo absorbancia con los datos centrados respecto su valor medio.

Como se ha explicado, en el caso de los espectros FTIR, las 92 muestras totales se dividieron aleatoriamente en 63 muestras de calibración y 29 de test. En el caso de los espectros NIR sólo se utilizaron 86 muestras, divididas en 58 de calibración y 28 de test.

Mientras que las muestras de calibración se utilizan para ajustar el modelo matemático, las muestras de test sirven para validar la bondad del modelo, ya que permiten utilizar muestras diferentes de las que se usaron para calibrarlo. El propósito de este trabajo es analizar el comportamiento del algoritmo de reducción de dimensiones ICA. Para ello,

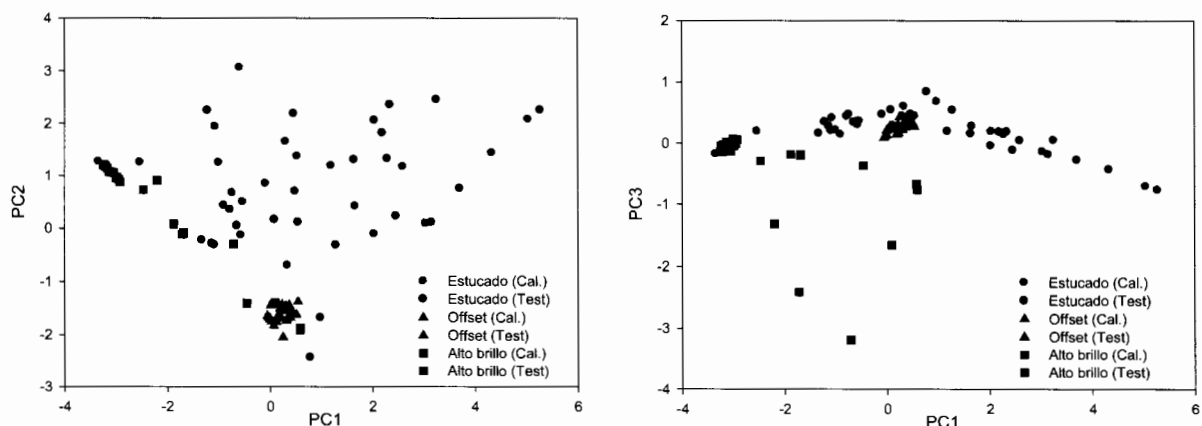


Fig. 3. Espectro FTIR de todas las muestras de los conjuntos de calibración y test vistas en el espacio de los tres primeros componentes principales. Este dibujo sólo es una vista parcial ya que se aprecian sólo tres de los nueve PCs que se utilizan para la clasificación.

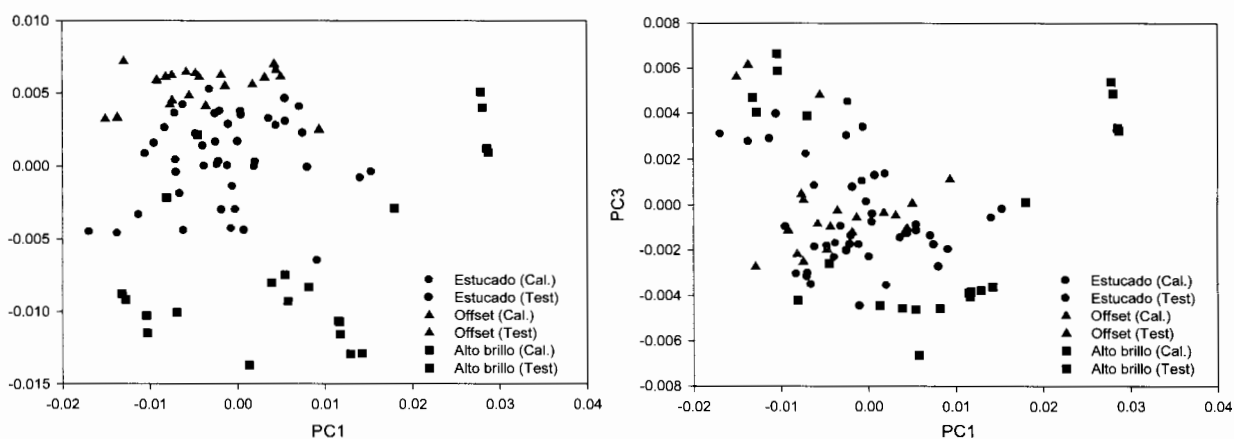


Fig. 4. Espectro NIR de todas las muestras de los conjuntos de calibración y test vistas en el espacio de los tres primeros componentes principales. Este dibujo sólo es una vista parcial ya que se aprecian sólo tres de los ocho PCs que se utilizan para la clasificación.

en esta sección los resultados obtenidos mediante dicho método se comparan con un método muy utilizado, PCA. Para comparar el comportamiento de ambos métodos en igualdad de condiciones se utiliza en ambos casos el clasificador k-NN.

5.1 Clasificación de las muestras de papel aplicando PCA + k-NN

En este apartado se evalúa el comportamiento del algoritmo PCA como sistema para reducir dimensiones seguido de la aplicación del clasificador k-NN. Para ello los datos espectrales conforman la entrada del algoritmo PCA, mientras que sus datos de salida o PCs sirven de entrada para el clasificador k-NN. Este último proporciona la decisión, es decir, clasifica cada muestra en una de las tres clases de acabados de papel analizadas.

En el caso de los espectros FTIR se retienen los nueve primeros PCs ya que ellos explican más del 99.5% de la variancia total de los datos ²⁰. Para los espectros NIR con sólo los ocho primeros PCs ya se explica el 99.5% de la variancia total ²⁰.

A continuación se aplica el clasificador k-NN para k = 3, 4 y 5 a los datos proporcionados por el algoritmo PCA, tanto para los espectros FTIR como NIR.

La tabla 1 muestra los resultados de la clasificación de las muestras de test, donde se parecía una tasa de aciertos ligeramente superior para los datos obtenidos a partir de los espectros FTIR.

Tabla 1. Resultados aplicando PCA + K-NN

Espectro	Método aplicado	Clasificador	Tasa de clasificación correcta
FTIR	PCA (9 PCs)	3 vecinos	27/29 (93.10%)
		4 vecinos	27/29 (93.10%)
		5 vecinos	27/29 (93.10%)
NIR	PCA (8 PCs)	3 vecinos	24/28 (85.71%)
		4 vecinos	24/28 (85.71%)
		5 vecinos	25/28 (89.29%)

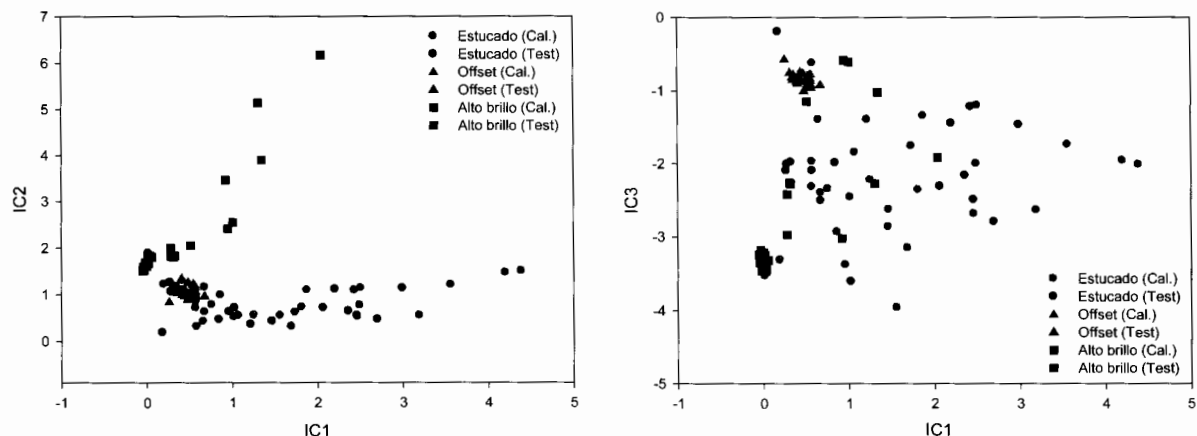


Fig. 3. Espectro FTIR de todas las muestras de los conjuntos de calibración y test vistas en el espacio de los tres primeros componentes independientes.

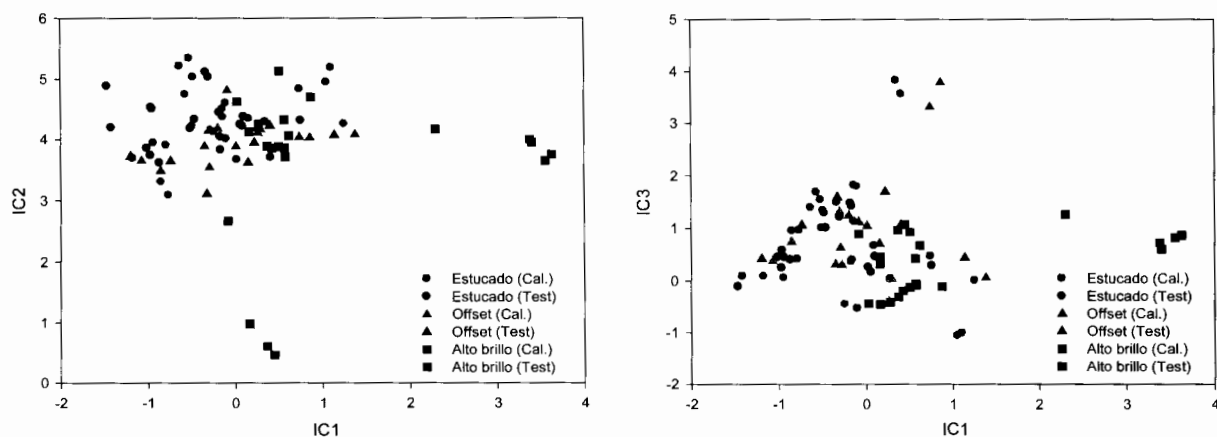


Fig. 4. Espectro NIR de todas las muestras de los conjuntos de calibración y test vistas en el espacio de los tres primeros componentes independientes. Este dibujo sólo es una vista parcial ya que se aprecian sólo tres de los cinco ICs que se utilizan para la clasificación.

5.2 Clasificación de las muestras de papel aplicando ICA + k-NN

En este apartado se evalúa el comportamiento del algoritmo ICA como sistema para reducir dimensiones seguido de k-NN aplicado como clasificador. El proceso aplicado es idéntico al explicado en la sección 5.1, de modo que sólo se cambia el algoritmo PCA por el ICA.

De modo similar al razonamiento seguido para el método PCA, en el caso de los espectros FTIR reteniendo los tres primeros componentes independientes es suficiente, mientras que para los espectros NIR se deben retener los cinco primeros ICs.

A continuación se aplica el clasificador k-NN para $k = 3, 4$ y 5 a los datos proporcionados por el algoritmo ICA, tanto para los espectros FTIR como NIR.

La tabla 2 muestra los resultados de la clasificación de las muestras de test, donde como en el caso de anterior, se aprecia una tasa de aciertos ligeramente superior para los datos obtenidos a partir de los espectros FTIR.

Tabla 2. Resultados aplicando ICA + K-NN

Espectro	Método aplicado	Clasificador	Tasa de clasificación correcta
FTIR	ICA (3 ICs)	3 vecinos	28/29 (96.55%)
		4 vecinos	28/29 (96.55%)
		5 vecinos	28/29 (96.55%)
NIR	ICA (5 ICs)	3 vecinos	26/28 (92.86%)
		4 vecinos	26/28 (92.86%)
		5 vecinos	27/28 (96.43%)

Al comparar las tablas 1 y 2 se observa que en todos los casos ICA proporciona una tasa de aciertos de clasificación superior de las muestras de test, siendo en promedio la mejora de un 5.3% respecto al uso de PCA.

CONCLUSIONES

Debido al elevado porcentaje de reciclaje de papel y a la importancia que tiene el papel reciclado en el sector papelerero, las empresas del sector se ven obligadas a realizar controles de calidad muy estrictos para asegurar una uniformidad y una calidad determinadas en el producto

final. Ello implica que es muy importante disponer de métodos de control rápidos, eficientes y respetuosos con el medioambiente. Para este fin, en este trabajo se ha desarrollado un método que permite a las empresas papeleras seleccionar el papel reciclado utilizado como materia prima, aceptándolo o rechazándolo en función de su composición. El método se fundamenta en la adquisición de los espectros FTIR o NIR y su tratamiento posterior mediante los algoritmos ICA y k-NN. Concretamente se han analizado un total de 92 muestras de papel con tres tipos de acabados diferentes. Los resultados muestran que los datos espectrales obtenidos mediante espectrofotometría FTIR permite una tasa de éxitos de clasificación superior de las muestras de test. En el caso de la formulación ICA + k-NN se ha obtenido una tasa de clasificaciones correctas del 96.55% para los espectros FTIR, mientras que para el caso de los espectros NIR es en promedio del 94.05%. Estos resultados muestran que el sistema propuesto permite una identificación rápida y fiable de las muestras de papel.

AGRADECIMIENTOS

Los autores quiere agradecer a las empresas Sarriópapel y Celulosa S.A. (Barcelona, España), Torraspapel S.A. (Barcelona, España) y Grup d'Impressió S.L. (Igualada, España) por proporcionar las muestras utilizadas en este trabajo.

BIBLIOGRAFÍA

- Puig R., Rius A., Martí E., Solé M., Riba J., Fullana P., Industrial Ecology in the Catalan Paper Industry, *Afinidad*, **65**(536), pp. 262-268, 2008.
- McKinney R.W.J., *Technology of Paper Recycling*, Surrey, England: Chapman & Hall, Blackie Academic & Professional, chapter 3, 1997.
- European Recovered Paper Council, *European Declaration on Paper Recycling 2006–2010. Monitoring Report 2009*, Brussels, Belgium: European Recovered Paper Council, 2009.
- 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.
- Doshi M., *Recycled Paper Technology. An Anthology of Published Papers* Atlanta, USA: TAPPI Press, pp 12–8, 67–76, 86–9, 1994.
- Vyörykkä J., Fogden A., Daicic J., Ernstsson M. and Jääskeläinen A.-S., Characterization of Paper Coatings – Review and Future Possibilities, *Proc. of the 9th TAPPI Advanced Coating Fundamentals Symposium*, Turku, Finland, Feb. 2006. pp. 41-46.
- Workman J.J., Review of Process and Non-invasive Near-Infrared and Infrared Spectroscopy: 1993–1999, *Applied Spectroscopic Reviews*, **34**(1&2), pp. 1 – 89, 1999.
- Workman J.J., Infrared and Raman Spectroscopy in Paper and Pulp Analysis, *Applied Spectroscopic Reviews*, **36**(2&3), pp. 139–168, 2001.
- Pan J. and Nguyen K.L., Development of the Photoacoustic Rapid-Scan FT-IR-Based Method for Measurement of Ink Concentration on Printed Paper, *Analytical Chemistry*, **79**(6), pp. 2259-2265, 2007.
- Hodges R., Cullinan H. and Krishnagopalan G., Recent advances in the commercialization of NIR (near-infrared) based liquor analyzers in the pulping and recovery area, *TAPPI Journal*, **5**(11), pp. 3–10, 2006.
- Nogués F., Carrillo F., Colom X., Characterization of natural cellulose substrates by means of FTIR spectroscopy during an oxidation process, *Afinidad*, **59**(498), pp. 104-110, 2002.
- Dena-Aguilar J. A., Jauregui-Rincon J., Bonilla-Petriciolet A. et al. Graft copolymerization of poly(acrylonitrile) and vinyl acetate on starch via free radical in solution: synthesis and characterization, *Afinidad*, **68**(553), pp. 195-202, 2011.
- Stuart B., *Infrared Spectroscopy: Fundamentals and Applications*. Chichester, England: John Wiley & Sons, 2004.
- Riba J.-R., Canals T., Cantero R., Comparative Study of Multivariate Methods to Identify Paper Finishes Using Infrared Spectroscopy, *IEEE Trans. on Instrumentation and Measurement*, in press.
- Bona M.T. and Andrés J.M., Reflection and transmission mid-infrared spectroscopy for rapid determination of coal properties by multivariate analysis, *Talanta*, **74**(4), pp. 998–1007, Jan. 2008.
- Câmara J.S., Alves M.A., Marques J.C., Multivariate analysis for the classification and differentiation of Madeira wines according to main grape varieties, *Talanta*, **68**(5), pp. 1512-1521, 2006.
- Dolmatova L., Ruckebusch C., Dupuy N., Huvenne J.P., Legrand P., Quantitative analysis of paper coatings using artificial neural networks Original Research Article, *Chemometrics and Intelligent Laboratory Systems*, **36**(2), pp. 125-140, 1997.
- Dalla Mura M., Villa A., Benediktsson J. A., Chanussot J., and Bruzzone L., Classification of Hyperspectral Images by Using Extended Morphological Attribute Profiles and Independent Component Analysis, *IEEE Geoscience And Remote Sensing Letters*, **8**(3), pp. 542-546, 2011.
- Bhattacharyya N., Bandyopadhyay R., Bhuyan M., Tudu B., Ghosh D., and Jana A., Electronic Nose for Black Tea Classification and Correlation of Measurements With Tea Taster Marks, *IEEE Trans. on Instrumentation And Measurement*, **57**(7), pp. 1313-1321, 2008.
- Riba J.-R., *Classificació invariant d'imatges aplicant tècniques multivariants. Visió artificial*. Editorial Académica Espanola, 2011.
- Krzanowski W.J., *Principles of Multivariate Analysis. A User's Perspective*, 2nd edn. New York, USA: Oxford University Press, pp 53–75, 2000.
- Thambirajah J., Thornhill N. F., and Pal B. C., A Multivariate Approach Towards Interarea Oscillation Damping Estimation Under Ambient Conditions Via Independent Component Analysis and Random Decrement, *IEEE Trans. on Power Systems*, **26**(1), pp. 315-322, 2011.
- Albarbar A., Gub F., Ball A.D., Diesel engine fuel injection monitoring using acoustic measurements and independent component analysis, *Measurement*, **43**(10), pp. 1376–1386, 2010.

-
24. Hyvärinen A. and Oja E., Independent component analysis: Algorithms and applications, *Neural Networks*, **13**(4–5), pp. 411–430, 2000.
 25. Li M., Crawford M.M., and Jinwen T., Local Manifold Learning-Based k -Nearest-Neighbor for Hyperspectral Image Classification, *IEEE Trans. on Geoscience and Remote Sensing*, **48**(11), pp. 4099 – 4109, 2010.
 26. Canals T., Riba J.R., Cantero R., Cansino J., Domingo D., Iturriaga H., Characterization Of Paper Finishes By Use Of Infrared Spectroscopy In Combination With Canonical Variate Analysis, *Talanta*, **77**(2), pp. 751-757, 2008.