

## DESARROLLO DE UN MODELO SIMCA DE RECONOCIMIENTO DE PATRONES PARA LA CLASIFICACIÓN DE COMBUSTIBLE DIESEL

R. Fernández Fernández, A. Dago Morales, E. Linchenat Dennes  
Centro de Investigaciones del Petróleo

*En el control de calidad del combustible diesel se emplean procedimientos de ensayo oficialmente establecidos como métodos de referencia por diferentes organizaciones internacionales (ASTM, ISO, IP). Es conocido que estos métodos de referencia aunque son exactos, emplean un tiempo de medición considerable y requieren, por lo general, de apreciable cantidad de muestra. La aplicación de las técnicas quimiométricas de reconocimiento de patrones en la industria del petróleo ha hecho posible el desarrollo de procedimientos alternativos que permiten llevar a cabo la clasificación y control de calidad de diversos tipos de combustibles de una forma rápida y con el empleo de pequeña cantidad de muestra a partir de sus datos espectrales.*

*El objetivo de nuestro trabajo fue desarrollar un modelo SIMCA de reconocimiento de patrones para detectar muestras de combustible diesel que estén fuera de especificaciones técnicas, sobre la base del empleo de la espectroscopía FT-MIR. Las muestras de diesel se procesaron en el rango espectral entre 4000-400  $\text{cm}^{-1}$  con una resolución de 4  $\text{cm}^{-1}$  y 32 barridos por punto. Se utilizaron un total de 45 muestras, 30 para el cálculo del modelo y 15 para su validación.*

*El modelo SIMCA desarrollado posee una sensibilidad y selectividad adecuadas, y permite detectar combustibles diesel que están fuera de especificaciones técnicas de una forma rápida, económica y confiable, por lo cual puede constituir una herramienta útil para el control de calidad de este producto en las refinerías del país.*

**Palabras clave:** quimiometría, reconocimiento de patrones, modelo de clasificación, combustible diesel, método SIMCA.

*In the quality control of the diesel fuel used test procedures formally established as reference methods by different international organizations (ASTM, ISO, IP). It is known that these methods are accurate reference though, use a measurement time and require considerable, usually of a significant amount of sample. The application of the techniques of pattern recognition chemometrics in the oil industry has made possible the development of alternative procedures that enable to carry out the classification and quality control of various types of fuels in a fast and with the use of small amount Sample from their spectral data.*

*The aim of our study was to develop a model SIMCE pattern recognition to detect diesel fuel samples that are outside of technical specifications, based on the use of FT-MIR. Samples of diesel were processed in the spectral range between 4000-400  $\text{cm}^{-1}$  with a resolution of 4  $\text{cm}^{-1}$  and 32 sweeps per point. We used a total of 45 samples, 30 for the calculation of the model and 15 for validation.*

*The model developed SIMCE has an adequate sensitivity and selectivity, and the ability to detect diesel fuels that are out of technical specifications in a fast, economical and reliable, and therefore may be a useful tool for quality control of this product in refineries of the country.*

**Key words:** chemometrics, pattern recognition, classification, diesel fuel, SIMCE method.

### Introducción

En el control de calidad de los destilados medios del petróleo se emplean procedimientos de ensayo que han sido oficialmente establecidos como procedimientos de referencia por diferentes organizaciones internacionales; como por ejemplo: la ASTM (American Society for Testing Materials), la ISO (Internacional Standard

Organization) y el IP (Institute of Petroleum). Es conocido que estos métodos de referencia, aunque son exactos, emplean un tiempo de medición considerable y requieren, por lo general, de apreciable cantidad de muestra. Se debe destacar también que en muchas ocasiones cuando las muestras presentan pequeñas anomalías en su composición química, éstas no se pueden detectar a través de las propiedades físico químicas.

---

Los métodos quimiométricos han tenido gran aplicación en la industria de refinación del petróleo; por ejemplo, los métodos de calibración multivariada para el desarrollo de procedimientos alternativos para la predicción simultánea de propiedades físico químicas en derivados como gasolinas, kerosinas, turbocombustibles y diesel /1-9/. La exactitud de estos procedimientos alternativos es comparable con la exactitud de los métodos de referencia; presentan además las ventajas de utilizar pequeñas cantidades de muestras y de que el tiempo invertido en el análisis es pequeño. Por otra parte los métodos quimiométricos de clasificación y reconocimiento de patrones han sido ampliamente utilizados en la detección de adulteraciones de combustibles /10-13/. Una de las tendencias actuales en la industria petroquímica es el desarrollo de metodologías analíticas alternativas basadas en el empleo de técnicas espectroscópicas de infrarrojo cercano (FT-NIR), infrarrojo medio (FT-MIR) y ultravioleta visible (UV-Vis) para el control de calidad de diferentes productos, así como en el control de los procesos productivos.

En nuestro laboratorio se han desarrollado modelos de regresión multivariados para la predicción simultánea de propiedades físico químicas de destilados medios, mediante el empleo de la espectroscopía infrarroja de rango medio (FT-MIR) y el método de mínimos cuadrados parciales (PLS). El establecimiento de estas metodologías analíticas alternativas no culmina con el desarrollo y posterior validación del modelo; es necesario establecer un procedimiento de control de su validez con respecto al tiempo, principalmente en la industria de refinación del petróleo, donde ocurren con frecuencia variaciones de la materia prima y de los propios procesos productivos, que originan cambios en la composición química de los productos terminados; y por ende, problemas con la predicción de los modelos /14-16/. Para minimizar estas dificultades relacionadas con la detección de situaciones anómalas originadas por muestras atípicas, o no contempladas en el proceso de calibración, es necesario desarrollar modelos de clasificación que permitan identificar las mues-

tras que se encuentran fuera de especificación y definir además cuando es necesario realizar una actualización de las calibraciones, y de esta manera lograr modelos más robustos /16,17/.

El método SIMCA de reconocimiento de patrones (Soft Independent Modeling of Class Analogy) desde su introducción por Svante Wold en el año 1976 /18/ ha sido uno de los métodos de clasificación más utilizados en el control de calidad en las industrias alimenticia /19/ y farmacéutica <sup>[20]</sup>. En la industria del petróleo se ha utilizado para la detección de adulteraciones en gasolinas <sup>[10,13]</sup>. Es un método supervisado de reconocimiento de patrones que se basa en el principio de analogía entre las muestras que pertenecen a una misma clase, y emplea para el cálculo de las distancias los “scores” determinados mediante análisis por componentes principales (Principal Components Analysis, PCA). El método SIMCA calcula un modelo PCA para cada clase o categoría presente en el sistema objeto de estudio, posteriormente integra cada una de las clases y calcula sus límites o fronteras con una probabilidad dada, comúnmente del 95 % /21/.

En caso de que no esté disponible la información externa, las categorías o clases se pueden definir a partir de una evaluación de las agrupaciones, mediante el análisis exploratorio de datos (reconocimiento de patrones no supervisado) que brindan métodos como: el Análisis Jerárquico de Grupos (Hierarchical Cluster Analysis, HCA) y el Análisis por Componentes Principales. Una vez definidas las categorías, se construye y refina un modelo SIMCA para el conjunto de muestras de entrenamiento conocidas “training set”; y después éste se utiliza para predecir las clases de nuevas muestras desconocidas “test set”. Una muestra desconocida o muestra problema puede asignarse a una categoría, a más de una, o a ninguna; y se calcula la probabilidad de cada asignación.

El objetivo del trabajo es desarrollar un modelo SIMCA de reconocimientos de patrones para el control de calidad de combustible diesel, a partir de los datos que brinda la espectroscopía infrarroja FT-MIR.

## Materiales y métodos

Se utilizaron 45 muestras de diesel, de las cuales 10 están fuera de especificación de calidad. El conjunto de calibración se conformó con 30 muestras: 23 dentro de los parámetros de calidad y 7 fuera de especificación. El conjunto para validar el modelo desarrollado se conformó con un total de 15 muestras: 12 dentro de los parámetros de calidad y 3 fuera de especificación. Se utilizó la notación Dfe para las muestras de diesel que están fuera de especificación técnica y D para las muestras que están dentro del marco de las especificaciones de calidad.

Los espectros infrarrojos de las 45 muestras de diesel objeto de estudio se obtuvieron en un espectrómetro FT-MIR Avatar 360Esp, de la firma Nicolet Instrument Corp. Estos se midieron en absorbancia mediante la técnica de transmisión en el rango entre 4000-400  $\text{cm}^{-1}$  con una resolución de 4  $\text{cm}^{-1}$  y 32 barridos por punto. Se utilizó celda fija con ventanas de cloruro de sodio y un paso de luz de 0,03 mm. En el registro y procesamiento de los espectros se usó el sistema de programas OMNIC versión 5,2 a /22/.

Los cálculos de los modelos se llevaron a cabo en el rango espectral entre 1750-690  $\text{cm}^{-1}$ . Para minimizar el efecto del corrimiento de la línea base, los espectros fueron corregidos mediante el procedimiento de corrección multiplicativa de la señal (Multiplicative Scatter Correction, MSC). Se utilizó como preprocesamiento el centrado en la media. El número óptimo de componentes principales se seleccionó teniendo en cuenta el por ciento de varianza que cada componente principal logra explicar. El cálculo del modelo SIMCA se realizó en base a las clases definidas previamente en los análisis PCA y HCA. Para la evaluación de los modelos se utilizaron los diagramas Coomans; se fijó en un 5 % el nivel de significación de los límites de las fronteras que definen cada clase. En el procesamiento matemático y estadístico se empleó el programa Pirouette versión 3.11 /23/.

## Resultados y discusión

Los espectros infrarrojos de las muestras de diesel en el intervalo entre 1750-690  $\text{cm}^{-1}$  se muestran en la figura 1.

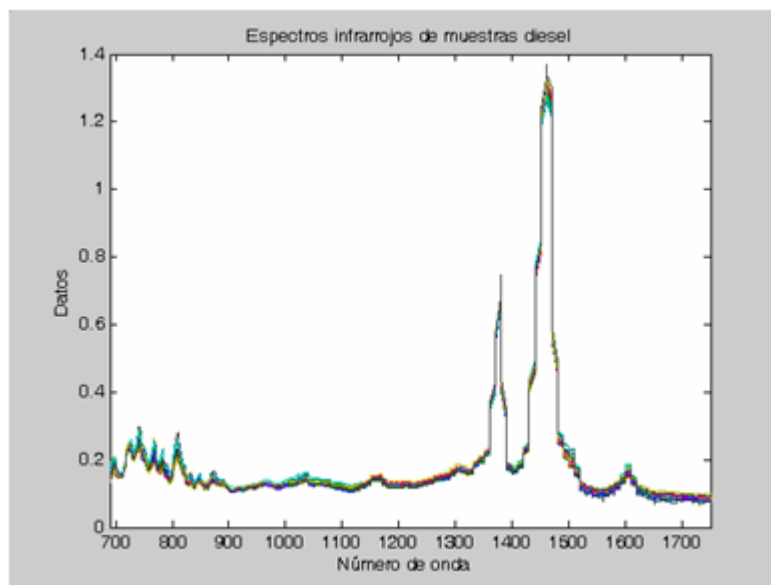


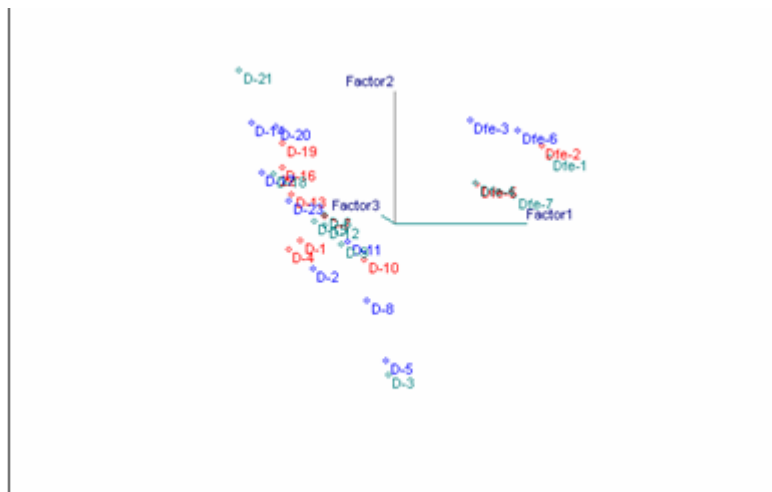
Fig. 1 Espectros infrarrojos de muestras de diesel.

Para definir las categorías o clases del modelo de clasificación se utilizó inicialmente el método no supervisado de reconocimiento de

patrones de análisis por componentes principales. El modelo PCA se calculó para 7 componentes principales con una varianza explicada

del 97 %. En la figura 2 se muestran los scores para los tres primeros componentes principales del modelo; se puede apreciar la

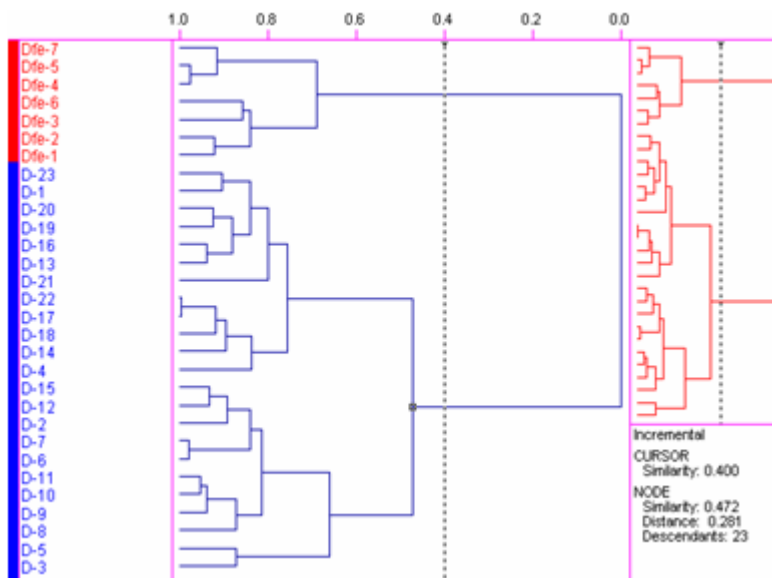
separación de las muestras de diesel en dos grupos diferentes D y Dfe sobre el primer componente principal (factor 1).



**Fig. 2 Modelo PCA del conjunto de calibración.**

Un resultado similar se obtiene también cuando se aplica el método no supervisado de análisis jerárquico de clusters. En la figura 3 se reporta el

dendograma del modelo HCA, en el cual se aprecia una buena separación (nivel de similitud cercano a 0,5) entre los dos grupos de diesel (D y Dfe).



**Fig. 3 Modelo HCA del conjunto de calibración.**

En la figura 4 se presenta la ampliación de una zona de los espectros infrarrojos en transmitancia para muestras representativas de cada clase. En el espectro de la muestra que pertenece a la clase

Dfe se observan apreciables diferencias en composición química en el intervalo espectral de 750-790  $\text{cm}^{-1}$  que tipifica la presencia de anillos aromáticos sustituidos.

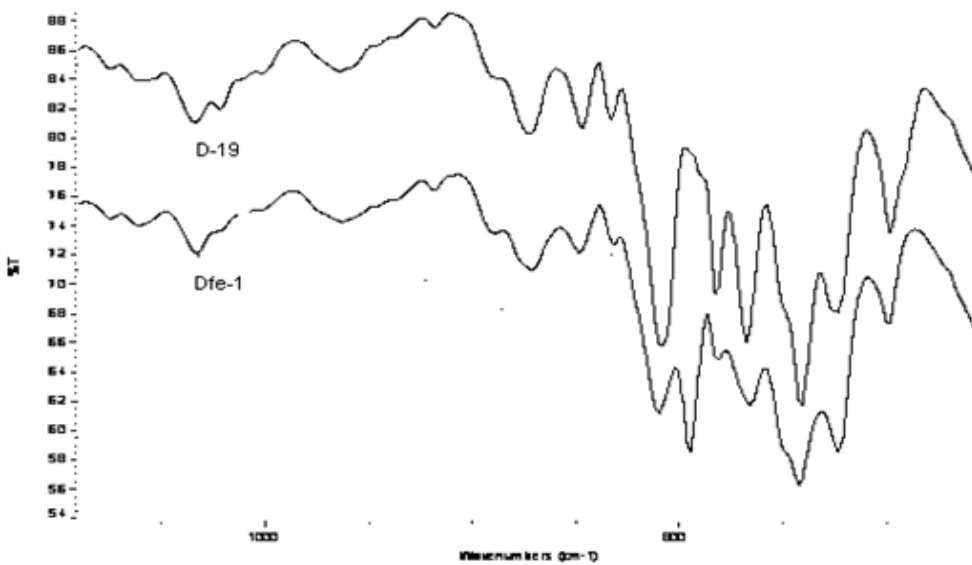


Fig. 4 Diferencias entre los espectros infrarrojos de muestras de cada clase.

El análisis exploratorio de datos preliminar demostró que, en principio, la espectroscopia infrarroja FT-MIR puede ser utilizada para el desarrollo de un modelo supervisado de clasificación con dos clases o categorías diferentes: la clase Dfe para las muestras de diesel que están fuera de especificación de calidad y la clase D para las muestras dentro de las especificaciones técnicas.

El conjunto de entrenamiento para el cálculo del modelo SIMCA está constituido por 23 muestras de

diesel de la clase D y 7 de la clase Dfe. El modelo se calculó con 3 componentes principales para ambas clases. En la tabla 1 se reporta un resumen de los resultados del modelo. En la figura 5 se reporta el diagrama de Coomans, en el cual se presentan las distancias ortogonales de todas las muestras del conjunto de calibración con respecto a las dos clases. Tanto en la tabla como en el diagrama se destaca que se logra una correcta clasificación de todas las muestras. La distancia entre las clases es de 3,92, lo que evidencia una adecuada separación entre las mismas.

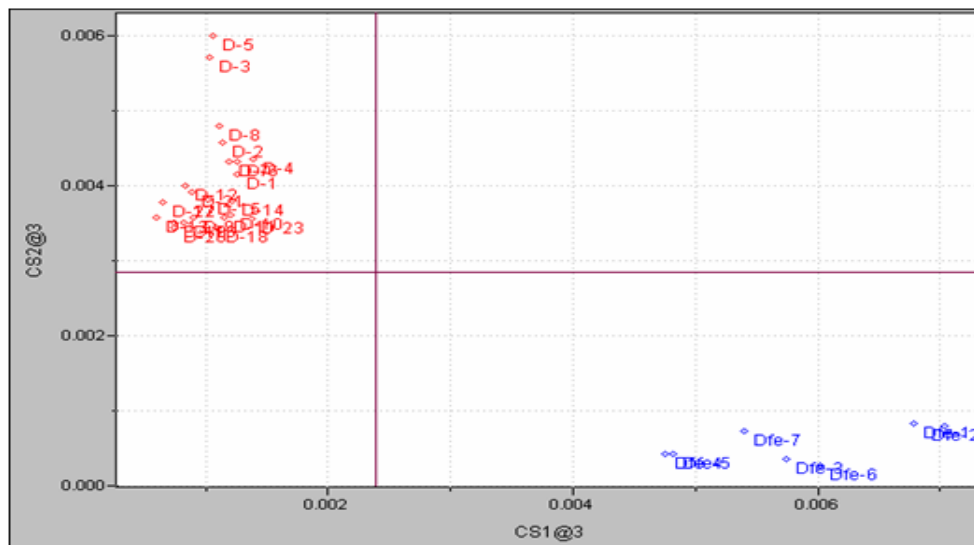


Fig. 5 Modelo SIMCA: distancias entre las muestras del conjunto de calibración con respecto a las clases D y Dfe (con líneas continuas se destacan los límites de las clases para un nivel de significación del 5%).

Tabla 1  
Resumen de los resultados del cálculo del modelo SIMCA

Clases	No. de muestras	No. de factores	% de varianza explicada	Clasificados correctamente	Clasificados incorrectamente	No clasificados
D	23	3	86,4	23	-	-
Dfe	7	3	94,0	7	-	-

La validación del modelo se llevó a cabo mediante un grupo conformado por 15 muestras: 12 del grupo D y 3 del grupo Dfe. En la

figura 6 se reportan los scores de las muestras del grupo de validación sobre los 3 componentes principales del modelo.

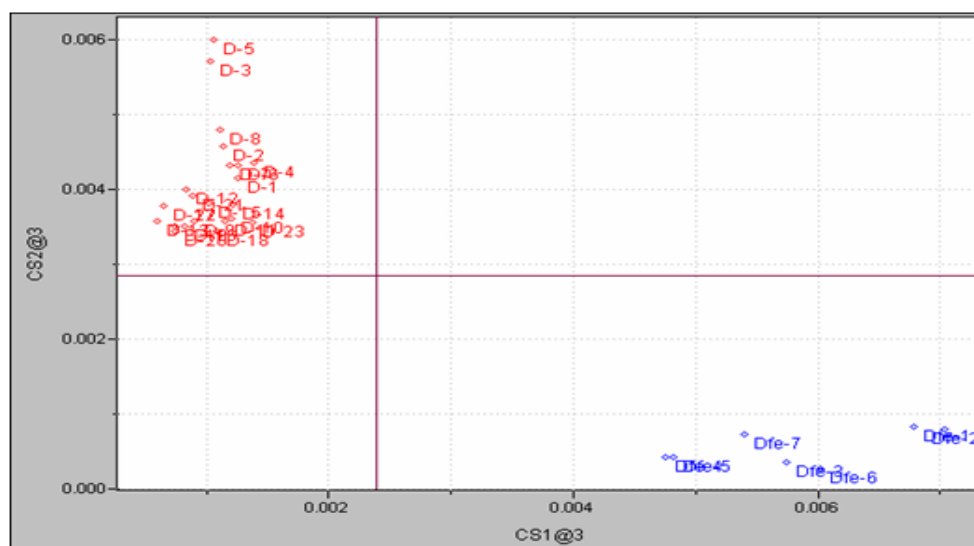


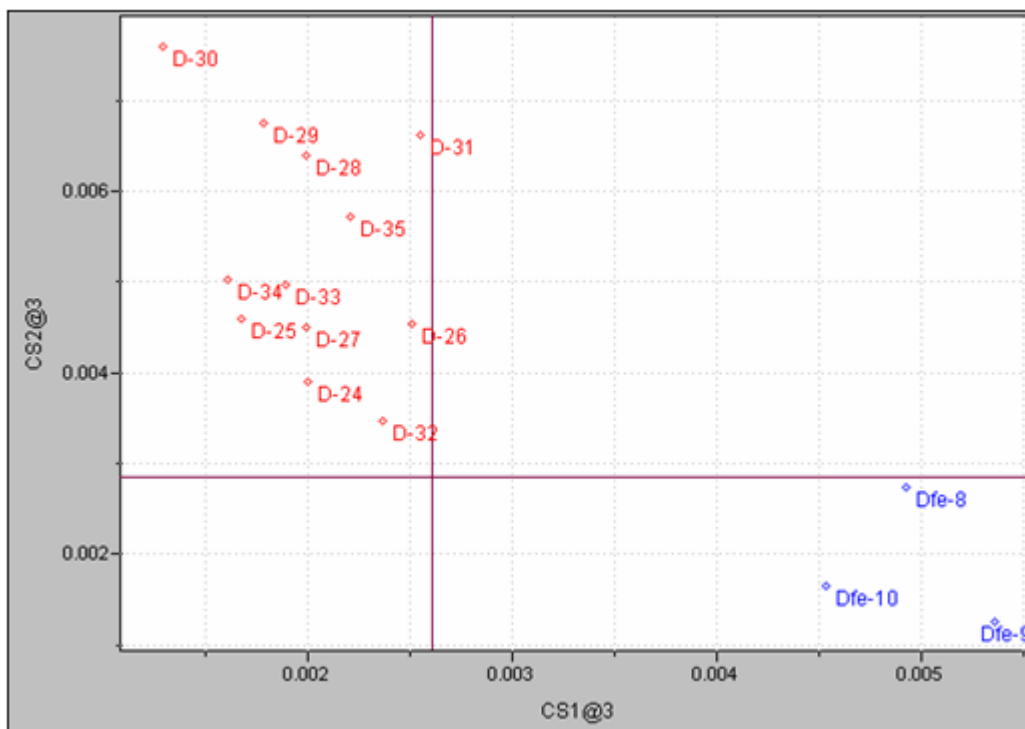
Fig. 6 Modelo SIMCA: Proyecciones de los scores de las muestras del conjunto de validación sobre los 3 componentes principales del modelo.

En la figura 7 se reporta el diagrama de Coomans, en el cual se reflejan las distancias de las muestras del conjunto de validación con respecto a las dos clases definidas.

En ambas figuras se destaca que los resultados de la validación del modelo SIMCA fueron

satisfactorios: el modelo desarrollado clasificó correctamente todas las muestras del conjunto de validación, no se presentaron falsos positivos ni falsos negativos, lo que da una idea de su robustez.





**Fig. 7 Validación del modelo SIMCA: Distancias de las muestras del conjunto de validación con respecto a las clases D y Dfe (con líneas continuas se destacan los límites de las clases para un nivel de significación del 5 %).**

## Conclusiones

Se desarrolló un modelo SIMCA de reconocimiento de patrones para el control de calidad de combustible diesel, mediante el empleo de la espectroscopía infrarroja de rango medio FT-MIR. El modelo desarrollado posee una sensibilidad y selectividad adecuadas y permite detectar combustibles diesel que están fuera de especificaciones técnicas de una forma rápida, económica y confiable. La metodología desarrollada puede constituir una herramienta útil para el control de calidad de este producto en las refinerías del país.

## Bibliografía

1. Fodor, G. E.; Mason, R. A.; Hutzler, S. A.; Appl. Spectrosc., 53, 1292, (1999).
2. Wold S., Trygg J, Berglund A and Antti H.; Chemometrics and Intelligent Laboratory Systems, 58, 131-150, (2001).
3. Andrade, J. M.; Garrigues, S; de la Guardia, M.; Gómez-Carracedo, M.; Prada, D.; Anal. Chim. Acta, 482, 115, (2003).
4. Vianney Santos, O. Jr.; Flavia Oliveira, C. C.; Daniella Lima, G.; Andrea Petry, C.; et.al.; Anal. Chim. Acta, 547, 188, (2005).
5. Côcco, L. C.; Yamamoto, C. I.; von Meien O. F.; Chem. Intel. Lab. Syst., 76, 55, (2005).
6. Pasadakis, N.; Sourligas, S.; Foteinopoulos, Ch.; Fuel, 85, 1131, (2006).
7. Brudzewski, K.; Kesik, A.; Kolodziejczyk, K.; Zborowska, U.; Ulaczyk, J.; Fuel, 85, 553, (2006).
8. Dago-Morales, A.; Fernández, F. R.; Ruiz-Martínez, M. D.; Balmayor, M. M.; et.al.; Rev. CENIC, Ciencias Químicas, 37, 3, (2006).
9. Balabin, R. M.; Safieva, R. Z.; Lomakina, E. I.; Chem. Intel. Lab. Syst., 88, 183, (2007).
10. Santos de Oliveira, F.; Gomes Teixeira L. S.; Ugulino Araujo, M.C.; Korn, M.; Fuel, 83, 917, (2004).
11. Wiedemann, L. S. M.; d' Avila, L., A.; Azevedo, D. A.; Fuel, 84, 467, (2005).
12. Pasadakis, N.; Kardamakis, A.; Anal. Chim. Acta, 578, 250, (2006).
13. Flumignan D. L.; Tininis, A. G.; Ferreira F. O.; de Oliveira, J. E.; Anal. Chim. Acta, 595, 128, (2007).

- 
14. Andrade, J. M.; García, M. V.; López-Mahía, P.; Prada, D; *Talanta*, 44, 2167, (1997).
  15. Swierenga, H.; de Weijer, A. P.; Van Wijk, R. J.; Buydens, L. M.; *Chemom. Intell. Lab. Syst.*, 49, 1, (1999).
  16. García-Mencía, M. V.; Andrade, J. M.; López-Mahía, P.; Prada, D.; *Fuel*, 79/ 4, 1823-1832, (2000).
  17. Daszyłowski, M.; Kaczmarek, K.; Vander Heyden, Y.; Walczak, B.; *Chem. Intell. Lab. Syst.*, 85, 203, (2007).
  18. Wold, S.; *Pattern Recognition*, 8, 127, (1976).
  19. Forina, M.; Lanteri, S.; "Data analysis in Food Chemistry" in Kowalski, B. R.; Ed. *Chemometrics. Mathematics and Statistics in Chemistry*; Reidel Publishing Co., 305, (1984).
  20. Gabrielsson, J.; Lindberg, N. O.; Lundstedt, T.; *J. Chem.*, 16.141, (2002).
  21. Ferreira, M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; *Quim. Nova*, 22, 724, (1999).
  22. Nicolet's OMNIC Spectroscopy Software, version 5.2a, Nicolet Instrument Corporation, Copyright (1992-2000).
  23. Pirouette. Infometrix Inc., Woodinville WA, version 3.11, (2003).