

Control de propiedades de gasoil mediante Tecnología NIR

Héctor Menez, Laura Markiewicz, Gustavo Marchetti, Horacio Fernández

REPSOL YPF S.A.
Dirección de Tecnología, CTA - Argentina

SINOPSIS

Las técnicas de análisis en línea basadas en métodos espectroscópico y modelos de regresión multivariante, se están imponiendo en el área del refino de petróleo e industria petroquímica, como una de las tecnologías más activas y promisorias para el control de calidad de los productos y aportación de datos en tiempo real a los sistemas expertos de control de unidades de producción. En particular, el uso de sistemas basados en espectroscopia de infrarrojo cercano (NIR) ha experimentado el mayor desarrollo y en los últimos años el número de aplicaciones a la industria ha crecido notablemente.

Un nuevo concepto, “*the on-line laboratory*”, está emergiendo en el lenguaje de los especialistas en análisis y control de procesos. Sus implicancias en la economía y organización de las unidades de refino será materia de análisis en los próximos años.

Por su parte los laboratorios deberán acompañar este proceso modificando su rol responsabilidades y capacidades. La incorporación al laboratorio de estas técnicas de medida multivariante permitirá incrementar significativamente su productividad y contribución a la mejora de los procesos.

REPSOL-YPF ha instalado o está en proceso de implantación de varios de estos sistemas en sus diferentes complejos industriales. Sus aplicaciones abarcan desde control de sistemas de *blending* de gasolinas hasta el control de fabricación de diversos productos petroquímicos.

Un proyecto de implantación de esta tecnología se desarrolla en diferentes fases y grados de integración que van desde estudios de viabilidad de la técnica mediante su instalación en laboratorio, su eventual verificación en plantas piloto, y finalmente su implantación en plantas industriales en conjunción con sistemas de control avanzado.

En este trabajo se mostrarán los estudios de aplicabilidad de la tecnología NIR en el control de propiedades críticas en corrientes base y formulaciones de gasoil.

INTRODUCCIÓN

La viabilidad de usar la espectroscopia infrarroja cercana (NIR) combinada con la calibración multivariante para predecir simultáneamente varias características químicas y físicas de los productos de petróleo ha sido mostrada en diversas publicaciones.

Esta técnica es capaz de proporcionar análisis rápidos y de bajo costo, tanto “*off-line*” como “*on-line*”, de gasolinas terminadas, corrientes de procesos y componentes de *blending*, resultando la información tan precisa como los métodos ASTM de referencia.

La aplicación de NIR a destilados medios para la medición de propiedades tales como densidad, destilación, punto de enturbiamiento, punto de inflamación, POFF, hidrocarburos aromáticos, número de cetano, también ha sido materia de estudio en publicaciones especializadas.

La espectroscopia infrarroja ha sido reconocida por muchos años como una de las herramientas analíticas más poderosas para la caracterización de compuestos químicos. Pero esta capacidad de generar una “*impresión digital*” a escala molecular ha sido aprovechable, hasta hace muy pocos años, sólo dentro de los límites del laboratorio.

El avance tecnológico de los años recientes en fibras ópticas de alta eficiencia han permitido invertir la ecuación tradicional, ya no debe “*viajar*” la muestra desde el proceso al laboratorio para producir la información, ahora se ha hecho posible llevar la espectroscopia al corazón del proceso.

El otro aspecto esencial ha sido el avance en sistemas ópticos prácticamente de “*estado sólido*”, una consecuencia de los desarrollos tecnológicos requeridos por la carrera espacial y reciclados posteriormente al ámbito industrial.

Así es como en la actualidad se puede disponer, a costo razonable, de equipamiento de muy alta velocidad de adquisición espectral y prácticamente invulnerables a las condiciones agresivas habituales de un ambiente industrial.

Vale mencionar como ejemplo que un solo analizador provisto de multiplexado óptico es capaz de monitorear hasta 16 puntos de muestreo o corrientes de proceso diferentes en un tiempo que va de 4 a 6 minutos para cerrar el ciclo de lecturas.

De las tres regiones en que habitualmente se divide la radiación infrarroja, cercano (NIR, medio (MIR) o lejano (FIR) es el infrarrojo medio el que es capaz de proveer información más específica sobre la naturaleza química de un material, sin embargo el mayor desarrollo de los últimos años se ha desplazado hacia el infrarrojo cercano, esto a pesar de que su señal es menos específica y por ende de más difícil interpretación.

Hay más de una razón que explica esto pero una de ellas ha sido determinante, y tiene que ver con el espectro infrarrojo del cuarzo. Este material constituye el núcleo de una fibra óptica y su espectro de absorción es muy intenso en el infrarrojo medio y relativamente débil en el cercano. La consecuencia práctica es inmediata, sólo la radiación de la región NIR es posible de ser transportada por fibra óptica, decenas de metros desde el sistema óptico hasta el punto de muestreo en la línea de proceso.

Esta característica de la radiación NIR con respecto al cuarzo y la alta velocidad de adquisición de los sistemas ópticos modernos, ha puesto a la espectroscopia NIR *on-line* como un proveedor privilegiado de información para la optimización en tiempo real de numerosos procesos químicos

Pero frente a estas decisivas ventajas de la espectroscopia NIR, la falta de especificidad de la señal dificulta la interpretación e introduce algunas complicaciones, aquí la información no es tan evidente como en el infrarrojo medio, donde con una sola longitud se puede establecer una relación confiable de intensidad/concentración. De hecho, buena parte de la analítica basada en infrarrojo se asienta en esta premisa y las bandas de absorción típicas de los grupos químicos están tabuladas y disponibles en la literatura.

En el caso del NIR la información útil está igualmente contenida en el espectro, sólo que no es evidente, está latente, para decirlo en la terminología usual.

El desarrollo de métodos matemáticos y estadísticos apropiados para extraer esta información oculta, latente, es el campo de investigación de una disciplina relativamente reciente, de principios de los '70, denominada quimiometría (del término inglés *chemometrics*).

El avance de la quimiometría ha sido posible por la aparición de procesadores rápidos y accesibles, junto al desarrollo de algoritmos de cálculos igualmente eficaces en resolver las ecuaciones matriciales requeridas en el tratamiento de datos espectrales. Estos métodos quimiométricos son capaces hoy de extraer la información latente contenida en los datos espectrales y de ese modo establecer una relación cuantitativa con los parámetros de interés. La desventaja es que demandan una etapa de desarrollo de la metodología más laboriosa que en los análisis espectrofotométricos clásicos, hasta alcanzar el estado de rutina analítica.

Asimismo la transferencia exitosa de esta tecnología analítica desde el laboratorio a líneas de proceso, requiere monitorear el funcionamiento y la confiabilidad del analizador a fin de evitar mediciones erróneas o deterioro en la calidad de las predicciones. Esto se encuentra fuertemente ligado a una actividad de soporte permanente, que incluya, para cada variable de interés y en el rango de calibración definido, la optimización del modelo, la validación de las predicciones contra los métodos de referencia y la investigación de datos anómalos.

MATERIALES Y MÉTODOS

HERRAMIENTAS MATEMATICAS Y ESTADISTICA UTILIZADAS PARA EL DESARROLLO DE MODELOS

ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)

El análisis en componentes principales (PCA) se aplica en química con diferentes finalidades [Jackson, 1991; Wold, 1987]. Entre ellas cabe destacar la *clasificación de muestras y la reducción en la dimensionalidad de los datos*, siendo éste el paso previo de muchos métodos multivariantes de análisis de datos. En la técnica multivariable PCA, un número elevado de variables (longitudes de onda en este caso), más o menos correlacionadas, son transformadas en un número inferior de variables ortogonales que contienen la información principal del sistema.

Se parte de una matriz de datos X , donde cada fila es una muestra u objeto, y cada columna una variable, que puede ser la absorbancia a una longitud de onda, como en este caso. La dimensionalidad de esta matriz, si se han registrado m muestras y k variables, será $m \times k$.

Geométricamente, el PCA es un cambio en la dirección y dimensionalidad de los ejes (Figura 1). Busca las direcciones ortogonales que explican la máxima variabilidad de las muestras y las utiliza como nuevos ejes de coordenadas llamados componentes principales (PCs) para representarlas. El primer componente principal es la dirección que explica la máxima variabilidad; el segundo se escoge de tal forma que sea perpendicular al primero y que explique la máxima variabilidad una vez extraída la explicada por el primer componente principal, y así sucesivamente. Para poder definir estos nuevos ejes se utilizan sus *loadings* que son los cosenos de los ángulos que forman con los ejes antiguos. Los *scores* son las coordenadas de las muestras en estos *nuevos ejes*. Muestras con coordenadas muy similares, es decir espacialmente muy próximas, indican una alta similitud espectral, o lo que es lo mismo, una muy cercana identidad química. Esta es la base de los sistemas de clasificación por grupos (*cluster*).

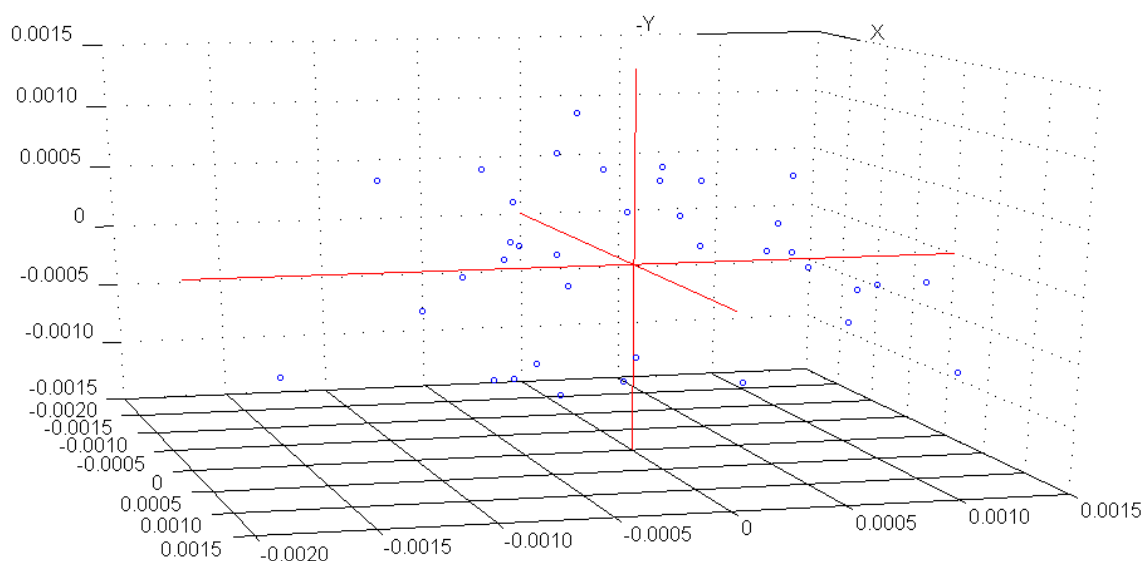


Figura 1- Representación geométrica de los componentes principales (PCs)

LA REGRESIÓN POR CUADRADOS MÍNIMOS PARCIALES (PLS)

La regresión por mínimos cuadrados parciales (Partial Least Squares, PLS) fue desarrollado por H. Wold en 1975 y hoy es un método quimiométrico ampliamente utilizado para realizar análisis cuantitativo.

Resulta especialmente útil para el estudio de relaciones complejas entre grandes cantidades de variables que presentan colinearidad. Mediante su aplicación se crea una ecuación de calibración o “modelo” que aplicada a los espectros de muestras “desconocidas”, predice un valor del parámetro de interés.

Para calcular estas ecuaciones, se colectan muestras con propiedades fisicoquímicas lo más cercana posible a las futuras muestras que pretendemos evaluar. Determinada la propiedad de interés por métodos de referencia establecidos, este conjunto de muestra compone el *set* de calibración.

La diferencia del PLS1 con otros métodos de regresión multivariantes previamente utilizados, por ejemplo PCR (*Principal Components Regression*), es que se intenta que los primeros componentes contengan la mayor información posible para la predicción de la matriz **Y**. Para ello, durante la etapa de calibración, el algoritmo PLS utiliza tanto la información contenida en la matriz de datos **X** (*p. ej.* datos espectrofotométricos) como la información contenida en la matriz de la propiedad a determinar, matriz **Y** (*p. ej.* concentraciones), obteniéndose unas variables auxiliares llamadas variables latentes, factores o componentes que tienen gran parecido a los componentes principales que se hallan a partir de un PCA.

Las muestras de calibración deben comprender todo el rango esperado de composición de las muestras futuras que se desea predecir y se deben medir bajo idénticas condiciones experimentales (método de muestreo, longitud de camino óptico, instrumento, temperatura, etc.). A partir de este conjunto de espectros y de los resultados obtenidos de los parámetros de interés por un método de referencia, se construyen las ecuaciones de calibración o modelos que correlacionan la matriz espectral con la variable a predecir.

La ecuación multivariable obtenida de este modo se utiliza para predecir la propiedad de interés en futuras muestras a partir de su espectro NIR.

Un aspecto importante del proceso de modelado es decidir el número de factores o componentes principales (PCs) que será utilizado en el modelo final, es decir determinar el tamaño óptimo del modelo. Para esta elección se considera la calidad del ajuste de los datos, la varianza explicada del espacio muestral, el valor de sesgo o *bias*, la eficiencia en los test de datos anómalos y el valor de los estimadores SEP (*standard error of prediction*) y RMSEP.

Observando las expresiones matemáticas de ambos estimadores se notará que en el caso del SEP se incluye la contribución del *bias*. En una situación en que el sesgo tienda a cero ambos estimadores tenderán a coincidir en su valor.

$$SEP = \sqrt{\frac{1}{N-1} \sum (x_i - y_i - \text{bias})^2}$$

$$RMSEP = \sqrt{\frac{1}{N} \sum (x_i - y_i)^2}$$

Donde $x_i - y_i$ = diferencia entre el valor predicho (x_i) y el método de referencia ASTM (y_i) de la muestra i .

$$\text{bias} = \frac{1}{N} \sum (x_i - y_i) = \bar{x} - \bar{y}$$

y N = número total de muestras

La incertidumbre de las predicciones futuras de los modelos se evalúa por el error estándar de predicción (SEP) de los datos, obtenido mediante validación cruzada completa (*full cross validation*). Este procedimiento construye la ecuación de calibración sobre $n-1$ muestras y predice la muestra excluida. La muestra removida se reintroduce en el *set* de muestras de calibración y se repite el mismo proceso para cada muestra. Habiéndose obtenido las predicciones individuales para el *set* total de muestras, se calcula la diferencia promedio entre los valores predichos y los obtenidos por el método de referencia.

Se han propuesto diferentes técnicas para deducir el número de PCs óptimos. Se basan, en general, en el análisis del error de predicción para distinto número de factores. Algunas de ellas han sido descritas en la bibliografía [Malinowsky, 1991].

Puesto que el método más habitual para la construcción del modelo es el de la validación cruzada (*cross validation*), una forma muy popular de seleccionar el número de factores óptimo es la sugerida por Wold [Wold, 1978], que consiste en representar el valor de RMSEP frente al número de PCs y buscar el mínimo. Se parte de la idea de que el error disminuye al aumentar el número de PCs, ya que se modela cada vez mejor el sistema, hasta que llega un momento en que los nuevos PCs introducidos únicamente explican ruido y el RMSEP vuelve a aumentar debido al sobre ajuste del modelo.

Aunque es una idea razonable, el hecho de utilizar únicamente un número limitado de muestras (como máximo todas las presentes en el conjunto de calibración) hace que el método esté sujeto a un cierto error y tenga el peligro de generar algún sobre ajuste de los datos.

Otros autores prefieren utilizar el primer mínimo local que aparece en la representación del RMSEP frente al número de PCs [Martes, 1989], aunque puede ser que en este caso se produzca un sub-ajuste de los datos. En este trabajo hemos adoptado el criterio propuesto por Harald Martens del primer mínimo local. Un sobre ajuste de los datos por excesivo número de PCs implica el riesgo de modelar ruido y provocar inestabilidad del modelo en futuras predicciones.

Cuando se trabaja con métodos analíticos normalizados, donde los estimadores estadísticos de precisión y sesgo están determinados en la propia norma, un modo de evaluar la consistencia de las predicciones es mediante la comparación de las respectivas expresiones de incertidumbre.

El error estándar de predicción SEP nos permite estimar la incertidumbre promedio de las predicciones de nuevas muestras, expresado en las mismas unidades que las del parámetro modelado.

Entonces, se puede representar el resultado de predicciones con su intervalo de confianza al 95% como:

$$\text{valor predicho} \pm 2*SEP$$

Esta medida es válida a condición de que las muestras nuevas sean similares a las utilizadas para la calibración del modelo, en caso contrario el error de la predicción puede ser superior. Esta condición no representa una complicación práctica ya que los equipos NIR modernos contienen algoritmos que calculan parámetros tales como el *leverage* o la *distancia de Mahalanobis*, que son una medida de *similitud* de una muestra desconocida con relación a las utilizadas en la construcción del modelo. Estos parámetros se informan simultáneamente con la predicción y permiten establecer rápidamente la confiabilidad de la predicción.

Para el control de las especificaciones de gasoil se aplican métodos de referencia ASTM y la incertidumbre de un método ASTM se evalúa por la reproducibilidad R:

$$R' = 2.77 \times \sigma_{R'}$$

y el desvío estándar $\sigma_{R'}$ de la reproducibilidad es:

$$\sigma_{R'} = \sqrt{\frac{\sum_{i=1}^n (I_i - \bar{I})^2}{n-1}}$$

Por lo tanto cuando el error de predicción (SEP) resulte inferior o igual al desvío estándar de la reproducibilidad σ_R del método ASTM de referencia podemos aseverar que la predicción es confiable.

Adicionalmente ASTM ha generado la guía estándar ASTM D6122 (*Standard Practice for Validation of Multivariate Process Infrared Spectrophotometers*) que describe un tratamiento estadístico rigurosos para evaluar la performance de las predicciones de analizadores NIR con relación al método de referencia ASTM.

La guía establece dos ensayos críticos para evaluar la calidad de las predicciones:

Test de sesgo: computa la diferencia promedio entre el valor del método de referencia y el dado por el modelo. Luego mediante un ensayo de tipo **t** establece la significancia del valor obtenido. Si el valor de *t* calculado es inferior al *t* tabulado para el dado número de muestras, se concluye que la diferencia promedio no es estadísticamente significativa (ausencia de sesgo).

Test de correlación: computa la pendiente **m** de la recta de regresión *y_{ref}* vs *y_{pred.}*, y el error estándar de la pendiente **S_m**, luego se calcula el cociente **m/S_m** y se compara con el percentil 95 de la distribución **t** de Student para n-2 grados de libertad. Si el valor obtenido supera al tabulado se concluye que existe correlación significativa entre valores de referencia y valores predichos por el modelo.

Si se superan ambos ensayos, el modelo es apto para ser instalado en el analizador y sus predicciones sobre nuevas muestras se pueden considerar estadísticamente indistinguibles de las obtenidas por el método primario.

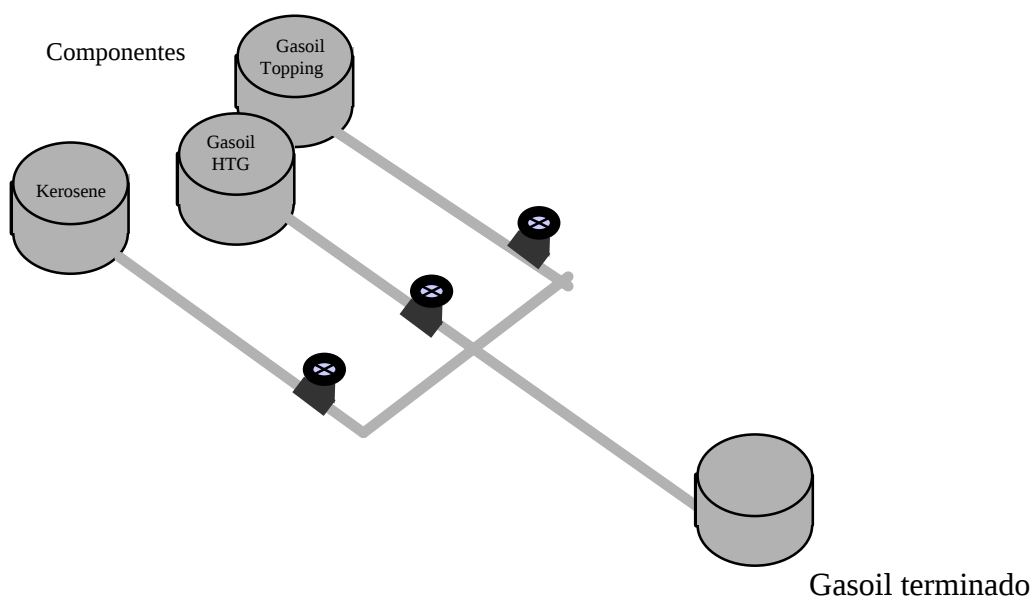
El cálculo de los modelo de calibración se realizó por regresión por mínimos cuadrados parciales (PLS1) mediante el uso del programa estadístico *Unscrambler*. Se construyeron modelos de los siguientes parámetros: densidad, destilación, punto de inflamación, punto de enturbiamiento, POFF y azufre.

CONDICIONES EXPERIMENTALES

Las medidas fueron realizadas en un analizador *NIR Luminar 3000 de Brimrose*, con tecnología *AOTF* (dispositivo óptico acústico sintonizado en frecuencia) en un intervalo de longitudes de onda de 1050 y 1650 nm con resolución de 1 nm.

Las muestras y espectros han sido provistos por el laboratorio del Complejo Industrial La Plata. Se estudió un conjunto de 61 muestras de destilados medios procedentes de Topping, HTG y de la colectora de gasoil.

El esquema general de la instalación es el siguiente:



ANÁLISIS PCA DE CORRIENTES DE GASOIL

Se realizó una descomposición en *componentes principales* del conjunto total de espectros de gasoil procesados con derivada primera. Este conjunto comprende 21 muestras de gasoil de HTG, 20 muestras de gasoil de Topping y 20 muestras de gasoil de colectora Y59.

Debido a la diferente composición del gasoil proveniente de HTG y Topping, los espectros de dichas muestras se agrupan según su proveniencia y las clases observadas se corroboran en la representación de los dos primeros componentes principales que se presenta en la siguiente figura:

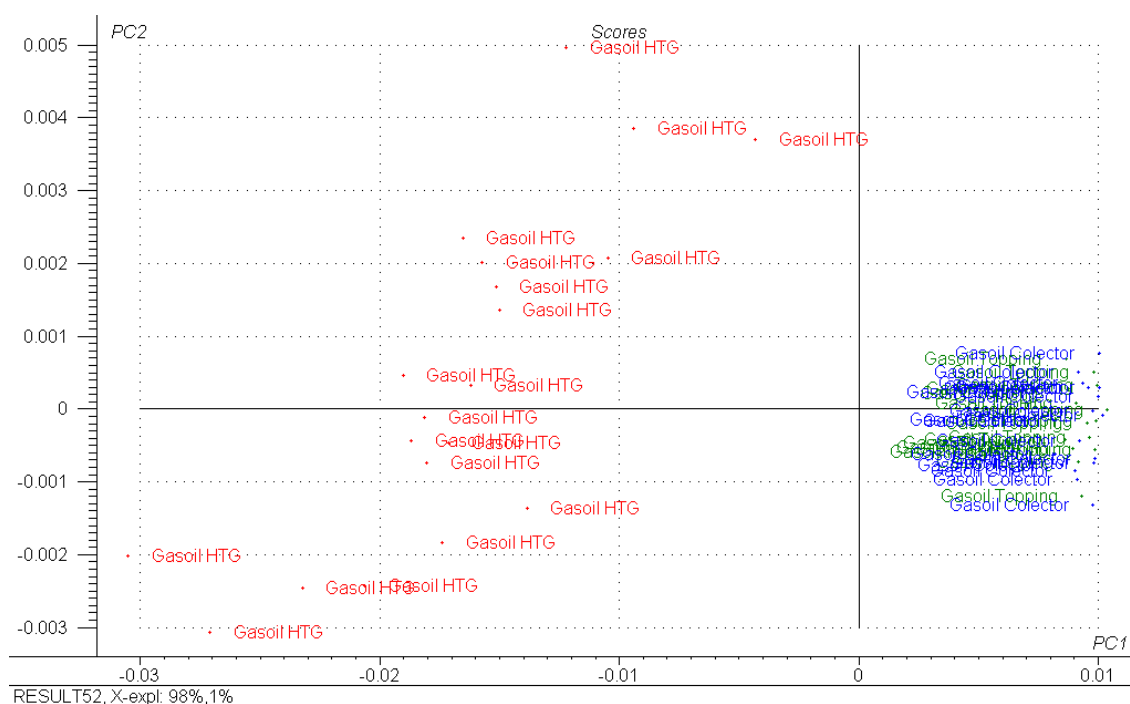


Fig. 2- Representación del primer y segundo PC del conjunto de muestras

En la siguiente figura se observa la representación geométrica para los tres primeros factores o PCS. Se podrá observar al pie del gráfico los porcentajes de varianza explicada en cada caso. Se comprobará que el 99 % se explica con dos PCs.

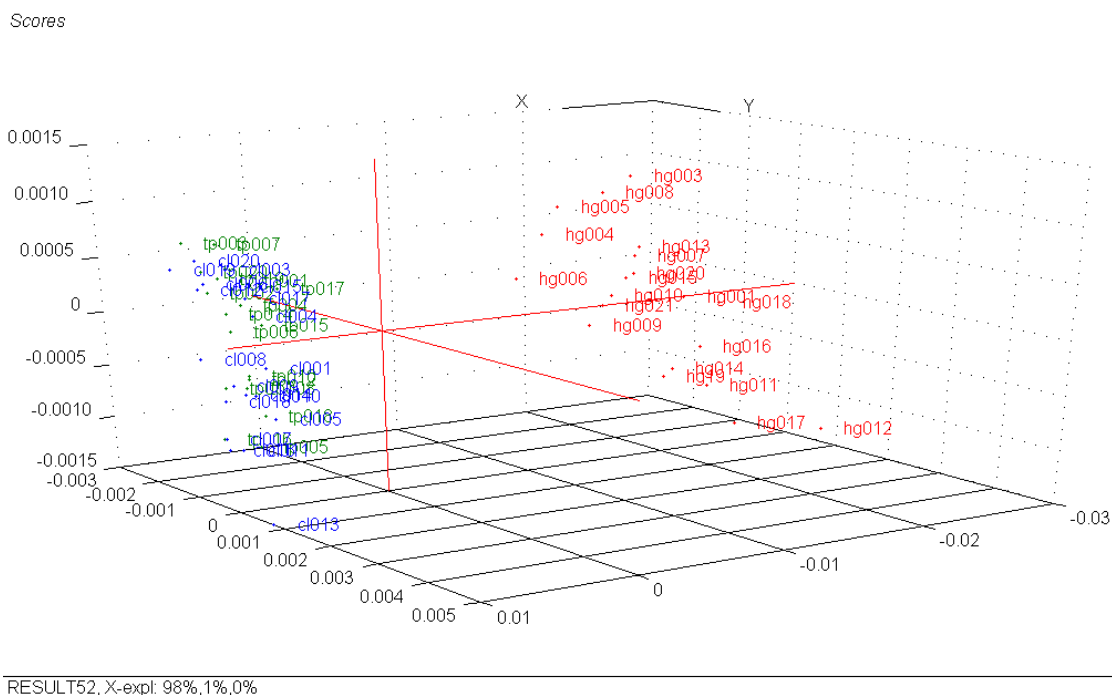


Figura 3

En ambos gráficos se evidencia claramente las dos consecuencias mencionadas del análisis PCA, reducción del número de variables y clasificación de muestras. La matriz espectral original tenía 600 variables (long. de ondas) y en la nueva dimensionalidad mediante PCA se explica mas del 99 % de la varianza total con sólo tres nuevas variables o PCs.

También se han puesto en evidencia las similitudes y disparidades de las tres corrientes analizadas, la corriente de HTG (hg) aparece alejada espacialmente del racimo que conforman topping (tp) y colectora (cl).

Otro aspecto de suma utilidad al evaluar corrientes de procesos es observar las dimensiones espaciales de ambos racimos, se deduce fácilmente que la corriente de HTG tiene una mayor variabilidad de su composición en el tiempo que las otras dos corrientes. Esto muestra la potencia del análisis PCA, un espectro NIR que se obtiene en minutos en condiciones *off-line*, o en menos tiempo en una instalación *on-line*, procesado mediante PCA, permite monitorear la variabilidad del proceso prácticamente en tiempo real.

CONSTRUCCION DE LOS MODELOS

Se construyeron modelos correspondientes al conjunto de muestras de corrientes bases de Topping y HTG junto a la corriente de salida, desarrollados a partir de datos de referencia de los métodos ASTM.

Se presentan a modo de ejemplo la representación grafica del modelo de predicción de Índice de Cetano

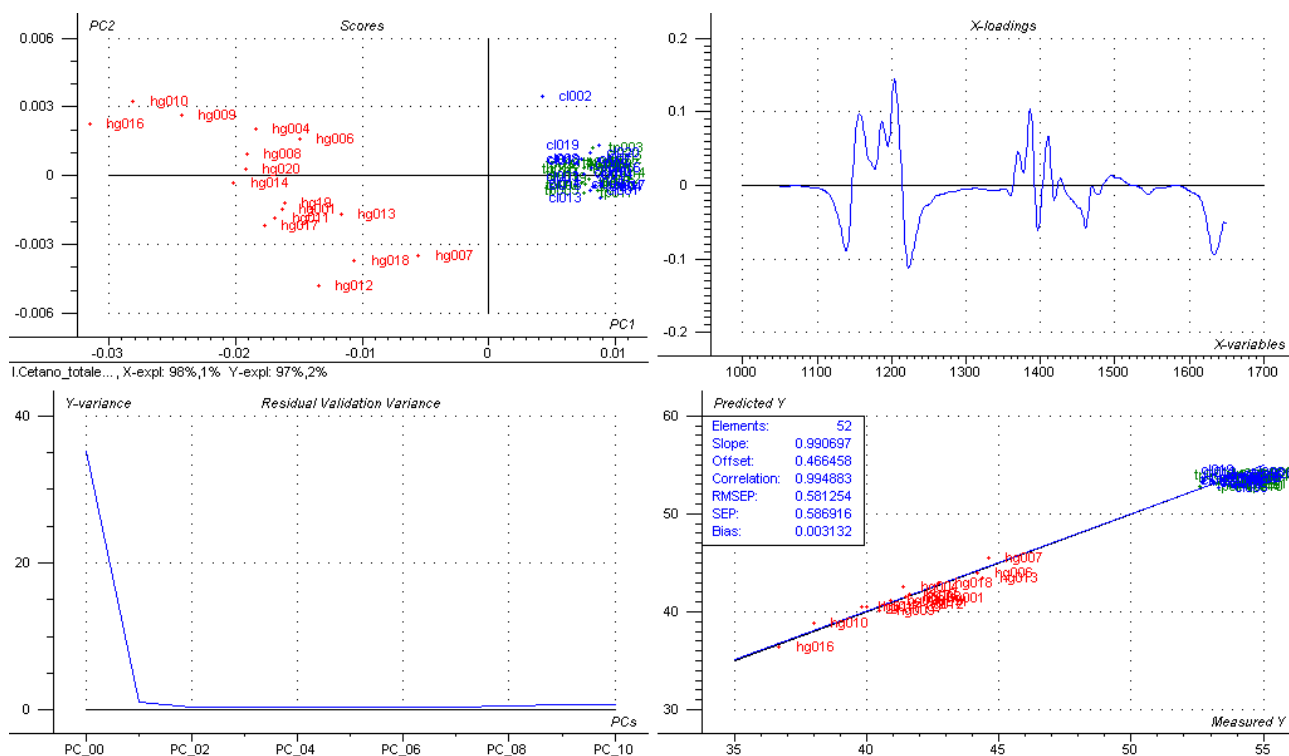


Figura 3

Obsérvese la relación entre el gráfico de *scores* y la recta de regresión. Se comprueba que el segundo PC clasifica a izquierda las muestras con IC menores 47-48 y a derecha las mayores a esos valores. Nótese también en el gráfico de *scores* los porcentajes de varianza explicada para X (datos espectrales) e Y (datos de Índice de Cetano ASTM). En ambos casos con dos factores se explica el 99 % de la varianza de los datos.

Para una mejor apreciación se representa a escala completa el modelo de regresión con sus estimadores estadísticos de precisión y sesgo.

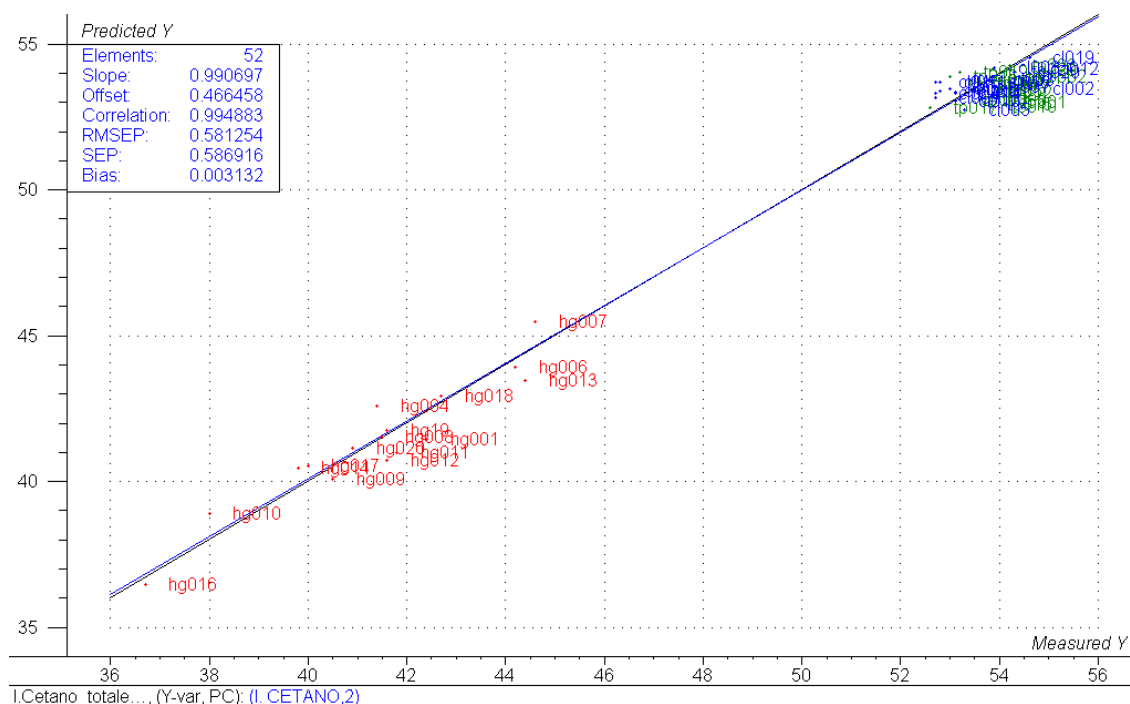


Figura 4

RESULTADOS DE VALIDACIÓN

En las tablas siguientes se muestran los resultados de validación según los criterios y estimadores definidos en ASTM D6122. Los errores de estándar de las predicciones (SEP) obtenidos por validación cruzada son inferiores en la mayoría de los casos a la desviación estándar de la reproducibilidad de los métodos de referencia. Sólo en el caso de la densidad el valor alcanza al de R del método, aquí es normal que el método primario compita favorablemente dada la alta precisión de los densímetros actuales. En cualquier caso es un indicador sólo para una instalación en línea ya que en laboratorio la medición por densímetro insume más o menos el mismo tiempo que la lectura por NIR. La diferencia radica en que en el mismo tiempo por NIR se obtienen todos los demás parámetros.

DENSIDAD					
Método de Referencia ASTM D 4052		Característica del Modelo NIR		Test de correlación y Test de Bias	
R_{ASTM} (g/ml)	0.0012	RMSEP	0.0012	m (Slope(a-r))	1.004
s_R	0.0004	Bias	5.138+E9	Sm	0.021
72% R	0.0004	SEP	0.0012	m/Sm	48.164
DesvEst	0.01	m (Slope(r-a))	0.9799	t	0.103
N° Mtras	58	r	0.9899	S_d^2	0
		SEP/ s_R	2.8	S_d	0.001
				t S_d	0

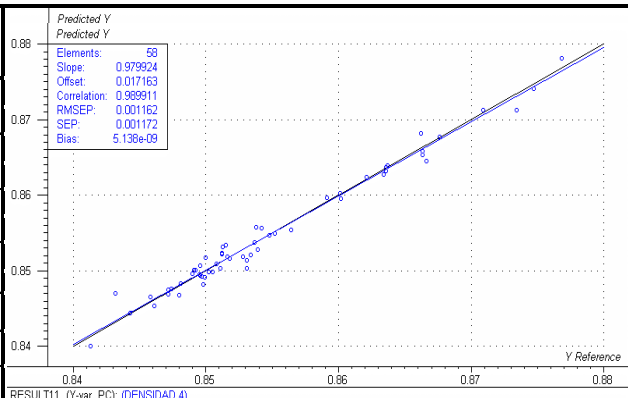


Tabla 1

10% Evaporado					
Método de Referencia ASTM D 87		Característica del Modelo NIR		Test de correlación y Test de Bias	
R_{ASTM} (°C)	7.1	RMSEP	3.7497	m (Slope(a-r))	1.0049
s_R	2.56	Bias	-0.5796	Sm	0.0256
72% R	5.1	SEP	3.7603	m/Sm	39.2313
DesvEst	25.94	m (Slope(r-a))	0.9749	t	0.8987
N° Mtras	34	r	0.9898	S_d^2	14.1402
		SEP/ s_R	0.5	S_d	3.7603
				t S_d	3.3794

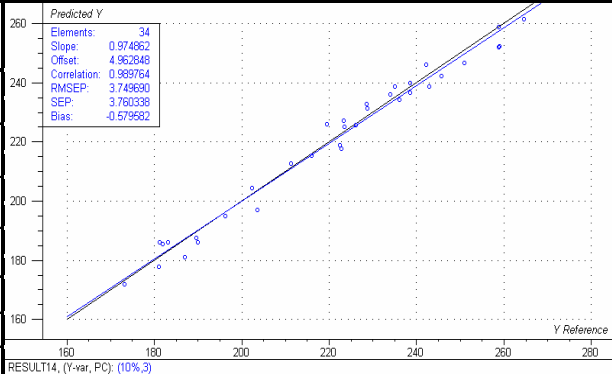


Tabla 2

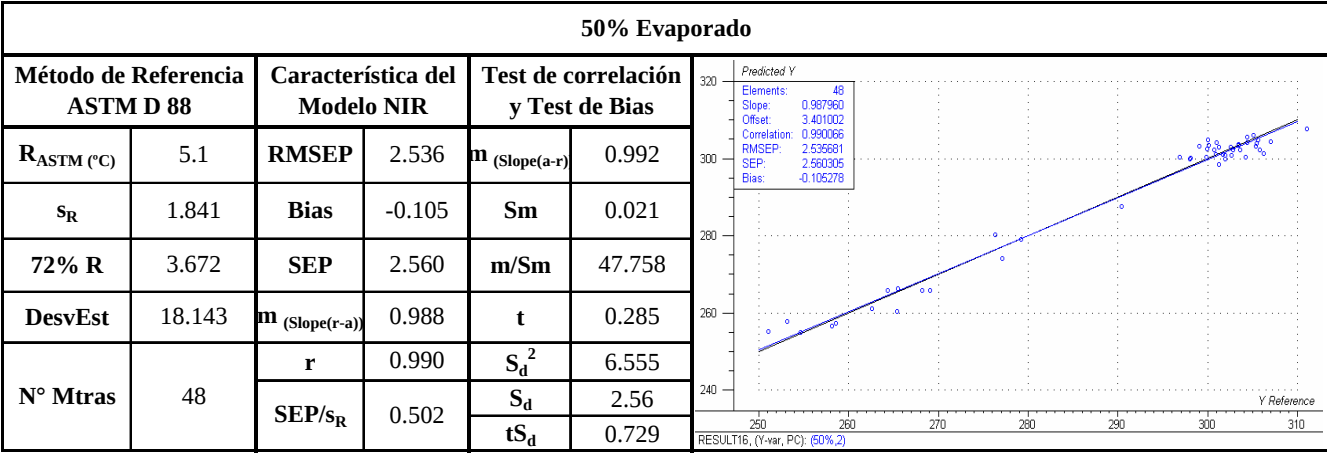


Tabla 3

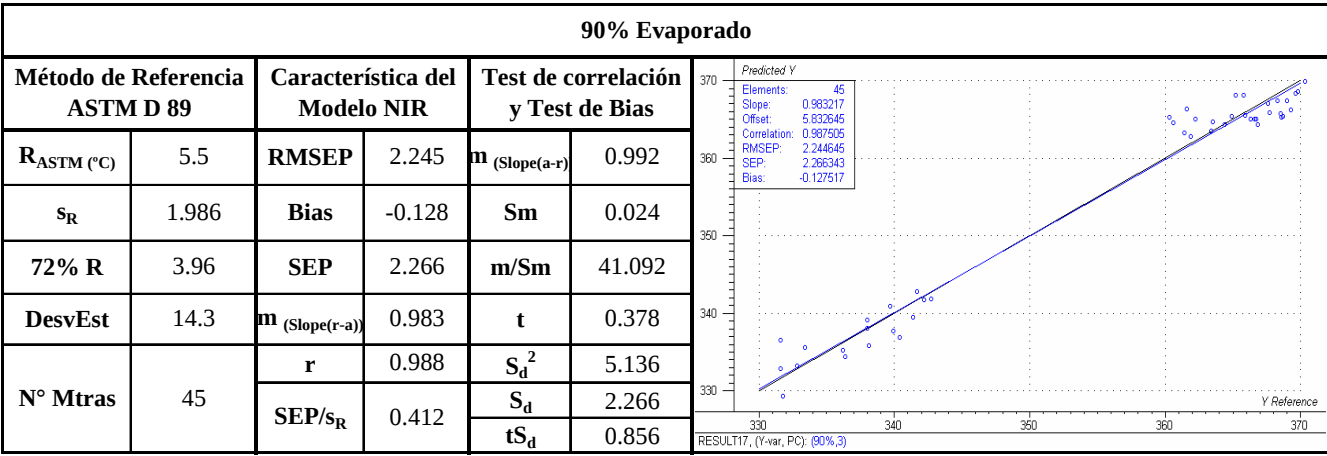


Tabla 4

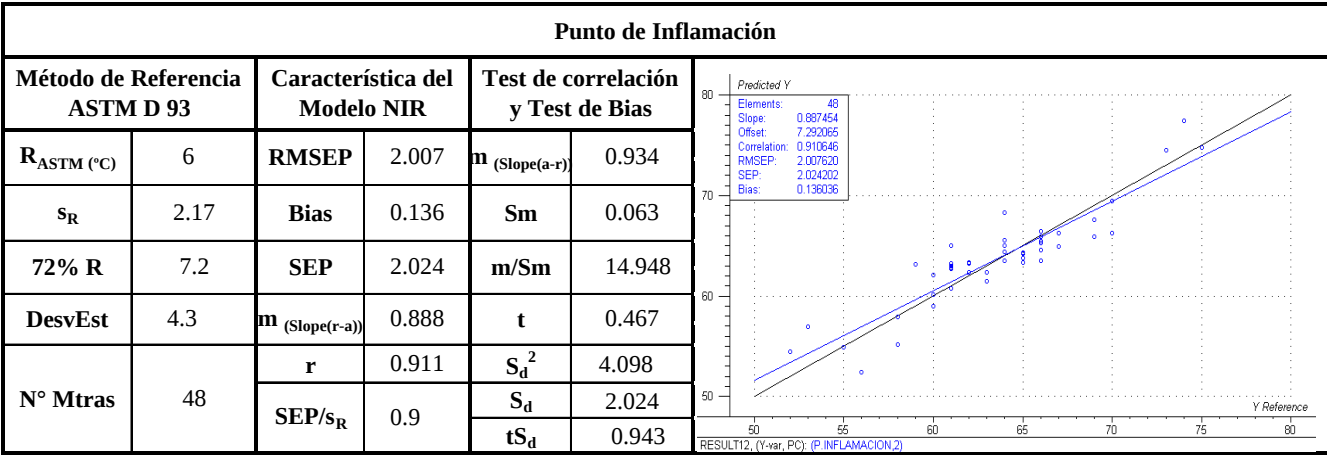


Tabla 5

Punto de Enturbiamiento					
Método de Referencia ASTM D 2500		Característica del Modelo NIR		Test de correlación y Test de Bias	
R_{ASTM} (°C)	3	RMSEP	1.112	m (Slope(a-r))	1.005
s_R	1.083	Bias	-0.005	Sm	0.028
72% R	2.16	SEP	1.124	m/Sm	35.454
DesvEst	5.91	m (Slope(r-a))	0.9606	t	0.032
N° Mtras	47	r	0.9826	S_d^2	1.264
		SEP/ s_R	0.375	S_d	1.124
				tS_d	0.036

RESULT21, (Y-var, PC): (PTO, ENTURBIAMIENTO,2)

Tabla 6

POFF					
Método de Referencia ASTM D 6371		Característica del Modelo NIR		Test de correlación y Test de Bias	
R_{ASTM} (°C)	1.9	RMSEP	0.8797	m (Slope(a-r))	0.995
s_R	0.686	Bias	0.037	Sm	0.022
72% R	1.368	SEP	0.888	m/Sm	44.369
DesvEst	5.772	m (Slope(r-a))	0.9814	t	0.2938
N° Mtras	49	r	0.9883	S_d^2	0.789
		SEP/ s_R	0.467	S_d	0.888
				tS_d	0.261

RESULT23, (Y-var, PC): (POFF,2)

Tabla 7

AZUFRE					
Método de Referencia ASTM D 4294		Característica del Modelo NIR		Test de correlación y Test de Bias	
R_{ASTM} (ppm)	142	RMSEP	72.83	m (Slope(a-r))	1.003
s_R	51.338	Bias	-13.016	Sm	0.029
72% R	102.388	SEP	72.594	m/Sm	35.029
DesvEst	416.635	m (Slope(r-a))	0.968	t	1.118
N° Mtras	39	r	0.985	S_d^2	5262.1
		SEP/ s_R	0.5105	S_d	72.54
				tS_d	81.129

RESULT37, (Y-var, PC): (AZUFRE,3)

Tabla 8

Los resultados del *test de Correlación* ($m/S_m > 2$) indican una alta correspondencia entre los datos de laboratorio y la predicción del analizador NIR en todos los parámetros, además se observa que los datos de referencia y las predicciones obtenidas no presentan un desvío estadísticamente significativo en el *Test de Bias* ($t < 2$).

Se debe tener en cuenta que mientras menor sea la incertidumbre en la determinación de la propiedad por el método primario (ASTM) mejor será la correlación con las predicciones de NIR y menor el estimador SEP. En el caso de propiedades críticas, se puede mejorar el set de datos de referencia mediante análisis replicados del set de calibración

CONCLUSIONES

Los espectros NIR del conjunto de muestras evaluadas, contiene la información química necesaria para modelar diversas propiedades fisicoquímicas tales como la densidad, la destilación, el punto de inflamación, el punto de enturbamiento, el Índice de Cetano y el azufre.

La aplicación de métodos quimiométricos tales como PCA y PLS1 a los espectros NIR de muestras del *blending* de gasoil permite la construcción de modelos de predicción de sus propiedades y permite evaluar la calidad del combustible en forma rápida y en una sola medida.

Las predicciones presentan una alta correlación en todos los parámetros, pero es recomendable enriquecer los modelos de manera sistemática y permanente con el mayor número de muestras representativas posible, de manera de cubrir todo el rango de valores de las propiedades medidas y capturar toda la variabilidad del proceso. De esa manera se robustecerán los modelos y se asegurarán las predicciones.

Cuando se observan en análisis PCA grupos claramente diferenciados por su procedencia; una opción a considerar es la de construir modelos para cada corriente de componentes y otro para la corriente de salida de blending.

Este trabajo fue restringido a algunas propiedades fisicoquímicas, otros parámetros, tales como hidrocarburos aromáticos totales, monoaromáticos/diaromáticos/poli-aromáticos, y número de cetano también han sido modelizados con éxito mediante NIR y PLS1.

La experiencia internacional indica que la mejora de productividad de los laboratorios que instalan esta metodología compensa la inversión inicial en equipamiento. El muy bajo costo de mantenimiento de un equipo NIR frente al de la suma de analizadores necesarios para cubrir la totalidad de análisis requeridos, constituye un atractivo adicional.

Del mismo, la oportunidad de optimizar en tiempo real el uso de las corrientes al *blending* sin sacrificar calidad permite rentabilizar adicionalmente las operaciones de blending. Un pequeño aumento de margen, en la magnitud de volumen de combustible producido, lleva habitualmente los períodos de repago de una inversión en línea a valores no mayores a los dos años.

BIBLIOGRAFIA

ASTM E 1655 “Standard Practices for Infrared, Multivariate, Quantitative Analysis.

ASTM D 6122 Standard Practice for Validation of Multivariate Process Infrared Spectrophotometers

Control of Refining Processes on Mid-Distillates by Near Infrared Spectroscopy. N. Zanier-Szydlowski¹, A. Quignard², F. Baco², H. Biguerd², L. Carpot² and F. Wahl². Oil & Gas Science and Technology – Rev. IFP, Vol. 54 (1999), No. 4, pp. 463-472

Apport du bootstrap à la régression PLS :application à la prédiction de la qualité des gazoles. S. Aji¹, S. Tavoraro¹, F. Lantz², A. Faraj Oil & Gas Science and Technology – Rev. IFP, Vol. 58 (2003), No. 5, pp. 599-608.

Partial Least Square Modeling for the Control of Refining Processes on Mid-Distillates by Near Infrared Spectroscopy S. Aji¹, N. Schildknecht-Szydlowski¹ and A. Faraj

Malinowsky, E. R.; Factor Analysis in Chemistry, 2nd ed., Wiley, New York, **1991**.

Martens, H.; Naes, T.; Multivariate Calibration, John Wiley & Sons, Chichester, **1989**.

Wold, H.; Multivariate Analysis, Ed. Krishnaiah, P. R., Academic Press, New York, **1966**.

Wold, H.; Soft Modeling by Latent Variables; the Non-linear Iterative Partial Least Squares Approach, en Perspectives in Probability and Statistics, ed. J. Gani, Academic Press, London, **1975**.
Wold, S.; Esbensen, K.; Geladi, P.; Chemom. Intell. Lab. Syst.; **2**, 37, **1987**.

Wold, S.; Technometrics, **20**, 397, **1978**.

Jackson, J. E.; User's Guide to Principal Components, Wiley, New York, **1991**.

Héctor R. Menez. Es Consultor de Investigación en la Unidad de Tecnología de Repsol YPF. Se graduó en Ciencias Químicas en la Universidad de Córdoba y obtuvo un post-grado en Refinación de Petróleo en la UBA. Ha desarrollado la mayor parte de su carrera profesional en YPF y luego en Repsol YPF. Actualmente desarrolla su actividad en el centro tecnológico de Ensenada, CTA. Sus principales áreas de interés actuales son la tecnología NIR y sus aplicaciones en análisis y optimización de proceso, estadística multivariable y quimiometría, estudios sobre impacto olfativo en unidades de producción, modelos de dispersión de contaminantes y olores, caracterización analítica de derrames, prospectiva tecnológica, etc. Es instructor interno en estadística aplicada al control de procesos y calidad en laboratorios, y consultor en temas de analítica y medio ambiente.
hmenezv@repsolypf.com

Laura G. Markiewicz. Es Licenciada en Ciencias Químicas, graduada en la Universidad de Buenos Aires, completó un postgrado en Gestión Ambiental en el Instituto Tecnológico de Buenos Aires (ITBA). Integró el staff del Laboratorio de Medio Ambiente e Higiene y Seguridad de la Refinería La Plata, posteriormente se incorporó al centro tecnológico de Ensenada, CTA como analista de laboratorio en cromatografía gaseosa. Actualmente forma parte del Programa de Refino donde participa en proyectos de sistemas de control avanzado previstos para las operaciones de blending de naftas (NIR) y modelos de gestión integral de olores en refinerías de petróleo, siendo también miembro del Equipo Motor de la Comunidad de Práctica de Normalización Analítica de Repsol YPF.
[lmarkiewiczzt@repsolypf.com](mailto:markiewiczzt@repsolypf.com)

Gustavo Marchetti: Título Técnico Químico, Asistente Técnico de Laboratorio. Se desempeña como Administrador del Sistema Gestión de Laboratorio (SGL). Referente de este sistema para todas las Refinerías de Argentina. Representante del Comité Permanente (España- Argentina) del SGL. Participación continúa en trabajos para ASTM e IRAM. Integra el Comité De Blending para revisión y actualización de especificaciones de Combustibles de Exportación. Participación en la puesta en servicio de NIR para Laboratorio.
gmarchettia@repsolypf.com

Horacio Fernandez: Título Técnico Químico, Asistente Técnico de Laboratorio. Se desempeña como Referente de Combustibles por refinería en Comité del IRAM, Participación continua en trabajos para ASTM. Integra el Comité de Blending para revisión y actualización de especificaciones de Combustibles de Exportación. Participación en la puesta en servicio de NIR para Laboratorio. Referente en motores WAUKESHA para determinación de Octanajes en Gasolinas.
hfernandezs@repsolypf.com