

GENERACIÓN Y VALIDACIÓN DE MODELOS PREDICTIVOS DE PROTEÍNA EN LECHE DESCREMADA EN POLVO USANDO ESPECTROSCOPIA DE LUZ INFRARROJA CERCANA

Alfonso Pérez Gallardo
0000-0002-3571-763X

Alejandra González Altamirano
0000-0001-6342-7181

Emilio Méndez Merino

Ángela Suárez Jacobo
0000-0003-4143-9557

Jorge García Fajardo
0000-0002-8454-2661

All content in this magazine is licensed under a Creative Commons Attribution License. Attribution-Non-Commercial-Non-Derivatives 4.0 International (CC BY-NC-ND 4.0).



Resumen: El contenido de proteína es el parámetro más importante de la leche descremada en polvo (LDP), debido a su alta demanda en la industria alimentaria. No obstante, los análisis de referencia son costosos, poca rapidez y complejos. Una solución es la medición mediante espectroscopía de luz infrarroja cercana por la rapidez, bajo costo, facilidad y porque no es necesario el uso de reactivos. Para implementar esta metodología es necesaria la generación de modelos predictivos confiables que sean equivalentes a la medición de referencia. Para ello se realizó la construcción de curvas de calibración. El principal problema, para la elaboración de las curvas de calibración se debe al contenido de lotes y variabilidad baja entre las muestras. Por ello es necesario, combinar (robustecer) los modelos existentes o incluir muestras que aporten mayor rango de medición. El objetivo de este trabajo es la comparación de modelos matemáticos generados con muestras de LDP, la generación de curvas de calibración al combinar modelos (robustecimiento), así como una metodología para validar dichos modelos mediante pruebas de contraste de valores t . Como resultados se muestra un modelo predictivo construido con las muestras internas, con una R^2 de 62.92, modelo del fabricante con una R^2 de 99.78 y modelo combinado con una R^2 de 99.27. Las curvas validadas por este método son útiles para su uso de rutina, con la confiabilidad de que será equivalente a los valores de referencia.

Palabras clave: Leche descremada en polvo, NIR, proteína.

INTRODUCCIÓN

La espectroscopia de luz infrarroja cercana (NIR) es una metodología de análisis que ofrece una alternativa directa, higiénica, rentable y no destructiva (Mohamed cols.,

2021). Esta metodología es aplicable a productos lácteos, adoptada para mediciones de rutina debido a su rapidez, bajo costo y facilidad de aplicación (Dos Santos cols., 2020; Wang cols., 2021). El empleo de esta técnica involucra la realización de análisis quimiométricos que sirven para construir curvas de calibración, que posteriormente serán utilizadas para las mediciones de rutina (Miller cols., 2002). El primer paso para la construcción de una curva de calibración es crear una base de datos de espectros de las muestras y obtener los valores de referencia. En este paso la calidad espectral y la precisión de la medición son fundamentales para obtener una buena curva de calibración. También, es necesario que las muestras estén distribuidas de manera uniforme en el rango de estudio y que este sea lo suficientemente amplio para que se realice un buen ajuste (Strani cols., 2021). No obstante, no siempre es posible tener rangos amplios con muestras obtenidas de productos de línea, por la propia conformancia y control de calidad entre lotes de producción. Por ello, se vuelve necesario buscar alternativas que permitan ampliar dichos rangos (Dos Santos cols., 2020). Entre las soluciones están el combinar productos similares. Un ejemplo es la combinación de leches de distinto contenido de proteína y grasa como la leche descremada, leche entera y leche concentrada (Risoluti cols., 2020). Otra opción es trabajar sobre el robustecimiento de curvas de calibración previamente desarrolladas de manera interna o por el fabricante del equipo, que mediante la adición de muestras puedan generar modelos predictivos útiles para la medición de rutina de las muestras internas (Temizkan cols., 2020). El objetivo del trabajo es implicaciones de trabajar con muestras de rangos estrechos (8,000 – 4,000 cm^{-1} aproximadamente) y como impacta los modelos generados. Se ilustra un caso de

robustecimiento de un modelo previamente construido para obtener un mejor ajuste. Enseguida se muestran las implicaciones que tienen la exactitud de las mediciones de referencia en las curvas de calibración y finalmente se presenta una metodología para realizar una revisión y validación de dichos modelos.

METODOLOGÍA

MUESTRAS DE LECHE

Para la construcción de los modelos se analizaron 67 muestras de leche descremada en polvo (LDP). Para la validación se emplearon 40 muestras de LDP. Las muestras fueron adquiridas y donadas por Sigma Alimentos Lácteos S.A. de C.V. Fueron resguardadas en un almacén, a temperatura ambiente de 25°C.

ANÁLISIS BROMATOLÓGICOS

La determinación del contenido de proteína total fue acorde con el método de Kjeldahl, según la norma NMX-F-608-NORMEX-2011. Se utilizó el factor de conversión de porcentaje de nitrógeno a porcentaje de proteína.

ESPECTROSCOPIA DE LUZ INFRARROJA CERCANA

Las mediciones espectrales de NIR se realizaron en un espectrofotómetro FTIR-NIR modelo MPA-LSM, (Bruker Optics, Alemania) en modo de reflexión empleando el compartimento de la esfera de integración. Los espectros fueron tomados de 12,000 a 4,000 cm^{-1} con 64 barridos y 16 cm^{-1} de resolución. Para ello se tomaron 300 g de cada muestra que se colocaron en la copa de medición. La medición se realizó a temperatura ambiente de 25°C, por triplicado.

ANÁLISIS QUIMIOMÉTRICO Y OBTENCIÓN DE LOS MODELOS.

Todo el procesamiento de datos espectrales su tratamiento y la regresión por PLS, fueron realizados empleando el software de quimiometría OPUS 7.5 (Bruker Optics). La revisión de modelos del fabricante del equipo también se realizó empleando dicho software, aunque no se pueden visualizar los espectros al estar protegidos, si fue posible trabajar con la optimización quimiométrica de los modelos. Para cada modelo se registraron los coeficientes de ajuste a la recta o R^2 , el error cuadrado medio del coeficiente de variación (RMSECV), la proporción entre el desempeño y la desviación (RPD), la desviación y la pendiente de la recta. Se eligieron los modelos con el menor RMSECV.

VALIDACIÓN DE CURVAS DE CALIBRACIÓN

Para validar los modelos se emplearon pruebas de contraste t para datos emparejados. Este método se basa en el contraste del valor t tabulado vs el valor t calculado a partir de las diferencias entre los grupos de valores predichos y los valores de referencia (Miller cols., 2002). Para este análisis se emplearon 40 muestras de LDP no incluidas en los modelos, con estas muestras se determinó la precisión, sensibilidad y especificidad del modelo.

RESULTADOS Y DISCUSIÓN

En la **Figura. 1** se muestran los espectros obtenidos de las 67 muestras LDP. En ella se puede apreciar una huella espectral de absorbancia sin presencia de ruido o picos extraños que indiquen una mala adquisición. Se aprecia dispersión o *scattering* típico de las muestras de LDP y que se debe a la reflexión en varias direcciones del haz de luz infrarroja. Esta dispersión es eliminada posteriormente por el procesamiento matemático como

la obtención de la primera derivada, normalización vectorial o corrección multiplicativa de dispersión (también conocida como MSC) (Miller cols., 2002). El tratamiento matemático fue evaluado y seleccionado automáticamente por el software, la elección se realizó de acuerdo con el que produjera menor RMSECV.

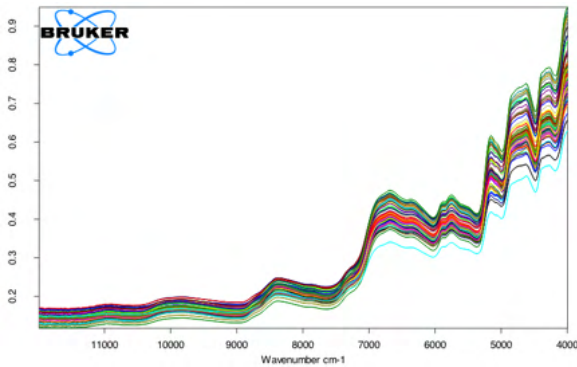


Figura 1. Espectros NIR de las 67 muestras de LDP.

MODELOS DE CALIBRACIÓN Y VALIDACIÓN MEDIANTE CONTRASTE DE VALORES T

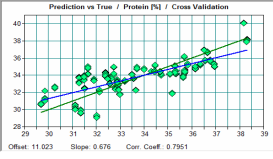
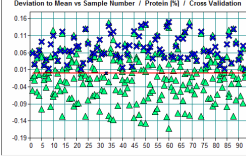
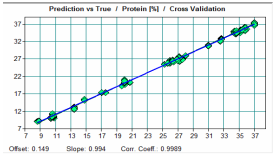
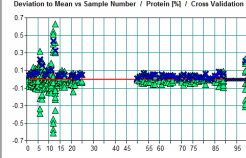
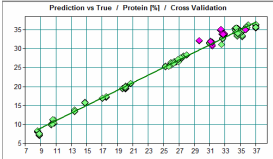
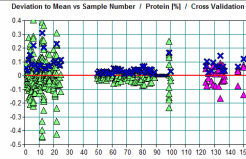
En el **Cuadro 1**, se muestra la comparación entre los tres modelos predictivos generados: (1) el interno con las 67 muestras analizadas de LDP, (2) el del fabricante y, (3) el combinado. La adición de las muestras internas al modelo del fabricante disminuye el RPD y la R^2 , mientras aumenta el RMSECV (Dos Santos cols., 2020). Esto es esperado ya que las muestras que se añadieron al modelo muestran una desviación respecto a la media mayor a las que existían en ese rango en el modelo del fabricante. Sin embargo, la combinación de modelos es necesaria debido a que el modelo del fabricante no es capaz de predecir las muestras internas de manera exacta (ver graficas de mediciones predichas vs valores de referencia en el **Cuadro 2**).

Los parámetros de R^2 , RMSECV y RPD, indican el ajuste de los datos a una curva de

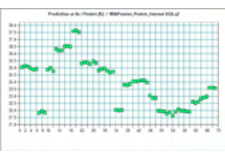
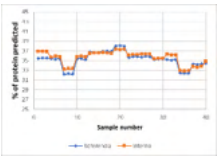
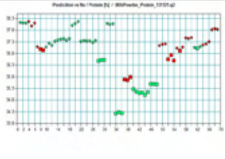
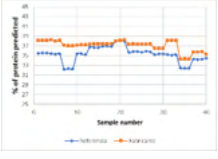
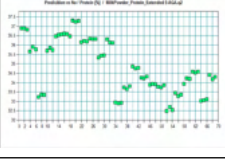
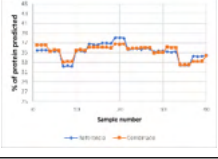
calibración, pero hace falta validar la curva mediante contraste de valor t para datos emparejados, con la finalidad de revisar si la curva arroja mediciones equivalentes a la medición de referencia (Miller cols., 2002). Como se puede observar en el **Cuadro 2** el modelo del fabricante no resultó adecuado para medir las muestras internas, esto es atribuible a la composición de las muestras, esto se debe a que las muestras eran de diferentes lotes y la composición cambia por temporada (Lin cols. 2017), o también se ve afectado por el tipo de proveedor. Al no contener las muestras LDP internas, el modelo del fabricante considera sus huellas espectrales como puntos fuera de la curva u *outliers* (Temizkan cols., 2020). Por esta razón es necesario el robustecimiento de las curvas (modelo combinado). Por otro lado, el modelo interno, aunque es capaz de medir las muestras internas, tiene una baja R^2 , un RMSECV más alto y presenta una alta dispersión sobre todo en los valores de las muestras internas. Finalmente, se puede apreciar como el modelo combinado ya es capaz de predecir las muestras internas de manera correcta, sin una diferencia significativa vs el método de referencia al realizar el contraste de valores t . Esto es porque se ha aumentado tanto rango como variabilidad de muestras.

CONCLUSIONES

Se encontró que en la generación de curvas de calibración es necesario trabajar con rangos amplios y que el set de muestras de LDP, incluya muestras similares a las que van a ser analizadas. Esto tiene implicaciones en la validez del método para reconocer las muestras y arrojar valores equivalentes a la medición de referencia. También se expuso una metodología para validar curvas de calibración, que se sustenta en pruebas de contraste de valores t . Las curvas validadas por

	Predicción vs referencia / validación cruzada	Desviación de la media vs número de muestra	R ²	RMSECV	RPD	Pendiente	Coefficiente de correlación
Modelo construido con muestras internas			62.92	1.33	1.64	0.68	0.79
Modelo del fabricante			99.78	0.42	21.2	0.99	0.99
Modelo combinado			99.27	0.751	11.8	1.00	0.99

Cuadro 1. Comparación de modelos predictivos construidos con las muestras internas contra el modelo del fabricante o la combinación de ambos. *En el modelo combinado se muestran en color rosa las muestras internas añadidas al modelo del fabricante.*

	Valores predichos	Valores predichos vs valores de referencia	t tabulada	t calculada	Diferencia significativa
Modelo construido con muestras internas			1.684	2.705	Si
Modelo del fabricante¹			1.684	10.498	Si
Modelo combinado			1.684	-0.335	No

Cuadro 2. Validación de curvas de calibración mediante contraste t para datos emparejados. ¹ los cuadros de color verde son muestras medidas correctamente. Los cuadros de color rojo son muestras fuera del rango. Las estrellas rojas son muestras detectadas como outliers.

este método son útiles para su uso de rutina, con la confiabilidad de que será equivalente a valores de referencia.

REFERENCIAS

1. Dos Santos Pereira, E. V., de Sousa Fernandes, D. D., de Araújo, M. C. U., Diniz, P. H. G. D., & Maciel, M. I. S. (2020). Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms. *LWT*, 127, 109427.
2. Lin, Y., O'Mahony, J. A., Kelly, A. L., & Guinee, T. P. (2017). Seasonal variation in the composition and processing characteristics of herd milk with varying proportions of milk from spring-calving and autumn-calving cows. *Journal of Dairy Research*, 84(4), 444-452.
3. Miller, J. N., & Miller, J. C. (2002). *Estadística y quimiometría para química analítica*, 4ta. Edición, Prentice Hall, págs. (49-51, 146-151, 221-242), ISBN: 0-13-022888-5.
4. Mohamed, H., Nagy, P., Agbaba, J., & Kamal-Eldin, A. (2021). Use of near and mid infra-red spectroscopy for analysis of protein, fat, lactose and total solids in raw cow and camel milk. *Food Chemistry*, 334, 127436.
5. Musa, M. A., & Yang, S. (2021). Detection and quantification of cow milk adulteration using portable near-infrared spectroscopy combined with chemometrics. *African Journal of Agricultural Research*, 17(2), 198-207.
6. Risoluti, R., Gullifa, G., & Materazi, S. (2020). Assessing the Quality of Milk Using a Multicomponent Analytical Platform MicroNIR/Chemometric. *Frontiers in Chemistry*, 8.
7. Strani, L., Grassi, S., Alamprese, C., Casiraghi, E., Ghiglietti, R., Locci, F., Pricca, N. & De Juan, A. (2021). Effect of physicochemical factors and use of milk powder on milk rennet-coagulation: Process understanding by near infrared spectroscopy and chemometrics. *Food Control*, 119, 107494.
8. Temizkan, R., Can, A., Dogan, M. A., Mortas, M., & Ayvaz, H. (2020). Rapid detection of milk fat adulteration in yoghurts using near and mid-infrared spectroscopy. *International Dairy Journal*, 110, 104795.
9. Wang, C., Reis, M. G., Waterhouse, G. I., Hemar, Y., & Reis, M. M. (2021). Prediction of dairy powder functionality attributes using diffuse reflectance in the visible and near infrared (Vis-NIR) region. *International Dairy Journal*, 117, 104981.