COMPARACIÓN DE MODELOS ESTADÍSTICOS EN LA ESTIMACIÓN DE INDICADORES DE CALIDAD DE UVAS TINTAS A PARTIR DE INFORMACIÓN ESPECTRAL

Miguel Noguera, Borja Millan, Arturo Aquino, Antonio Javier Barragán, Miguel Angel Martínez-Bohórquez, y José Manuel Andújar.
Centro de Investigación en Tecnología, Energía y Sostenibilidad (CITES). Universidad de Huelva. La Rábida, Palos de la Frontera, 21819 Huelva, España.
{Miguel.noguera@diesia.uhu.es, Borja.millan@diesia.uhu.es, Arturo.aquino@diesia.uhu.es, Antonio.barragan@diesia.uhu.es, bohorquez@uhu.es, Andujar@diesia.uhu.es}

Resumen

Los métodos tradicionalmente empleados para la determinación del estado de calidad de frutas tienen una reducida resolución espacial y temporal derivada de sus limitaciones (elevado costo y amplia brecha temporal entre el muestreo y el acceso a la información). En las últimas décadas, se han publicado numerosos trabajos que destacan a los métodos basados en espectroscopia como una alternativa prometedora a estos. Además, recientemente el auge de la industria electrónica ha supuesto un abaratamiento de los componentes, generando interés por el desarrollo de nuevos dispositivos. Incentivado por este contexto, este trabajo presenta un dispositivo multiespectral de bajo costo basado en un sensor comercial (AS7265x, AMS) sensible a 18 bandas entre los 410 y los 940 nm. Marcando como objetivo la evaluación de 3 modelos de estimación no paramétricos (dos lineales (Regresión lineal múltiple y Regresión por mínimos cuadrados parciales) y uno no lineal (Red neuronal artificial)) en el modelado de parámetros indicadores de calidad en uva tinta (sólidos solubles totales y acidez). Entre los modelos explorados, la red neuronal demostró ser el más eficaz para ajustar la relación entre la información espectral adquirida con el sensor propuesto y los indicadores de calidad considerados.

Palabras clave: AS7265x, multiespectral, regresión lineal múltiple, regresión de mínimos cuadrados parciales, redes neuronales artificiales, maduración, uva.

1 INTRODUCCIÓN

El proceso de maduración de la fruta implica una serie de modificaciones morfológicas, fisiológicas, y bioquímicas. De este modo, una fruta verde e inmadura normalmente adquiere una potente coloración y una textura más suave a medida que altera su composición química, volviéndose más dulce y aromática. Existen factores (estrés biótico y abiótico) que afectan al proceso de maduración durante el desarrollo de la fruta en campo y también durante su almacenamiento postcosecha [1]. Dichos factores pueden ser modulados externamente mediante la toma de decisiones por parte del gestor del cultivo. Un estado de maduración óptimo va a determinar el estándar de calidad la fruta, con lo cual contar con medios que permitan monitorizar estos cambios es fundamental para los diferentes sectores de la industria agroalimentaria.

Existen parámetros objetivos ampliamente aceptados como indicadores del estado de maduración de la fruta. Dependiendo del cultivo concreto con el que se esté trabajando y la finalidad última que se le pretenda dar a la cosecha, los indicadores de maduración tenidos en cuenta van a variar. Entre otros podemos mencionar: la firmeza del fruto, su contenido en sólidos solubles (SSC), almidón, azúcares, ácidos, aceites, concentración interna de etileno, concentración de flavonoides, etc [2]. Tradicionalmente, la determinación de estos indicadores de calidad se ha realizado mediante métodos químicos y físicos. Estas metodologías requieren instalaciones de laboratorio y personal con un elevado grado de capacitación, lo cual supone un elevado coste. Además, requieren el transporte de las muestras al laboratorio y por lo general tediosos procesos. Todo ello ensancha la brecha temporal entre la recolección de las muestras y el acceso a la información, limitando la capacidad de toma de decisiones. Toda esta serie de limitaciones otorgan a las metodologías tradicionales una reducida resolución espacial y temporal en la monitorización, ya que limitan tanto el número de puntos que se pueden tener en cuenta en cada muestreo, como el número de muestreos que se pueden llevar a cabo a lo largo de la campaña [1]. Estos factores ponen de manifiesto la necesidad de implementar alternativas a los métodos tradicionales más rápidas, económicas y no-destructivas.

 https://doi.org/10.17979/spudc.9788497498418.0568

Bajo este contexto, en las últimas décadas los enfoques basados en espectroscopia han cobrado una gran preponderancia como alternativa a los métodos tradicionales de caracterización del estado de calidad de frutas [3]. El fundamento teórico de este tipo de métodos reside en la interacción entre la luz y los frutos. Básicamente, los compuestos químicos que contiene la fruta se componen de varios átomos unidos por diferentes tipos de enlaces. Estos enlaces son excitados por la luz de longitudes de onda características dependientes de la naturaleza del enlace y los átomos que intervienen en el mismo. De este modo, cuando un haz de luz interacciona con un fruto, una parte de la radiación es absorbida, otra reflejada y el resto transmitida. Los sensores espectrales nos permiten caracterizar haces de luz. Por lo tanto, el uso de estos en combinación con una fuente de luz calibrada, en una disposición concreta, hace posible observar patrones de correlación entre la interacción de la luz con el fruto y la composición química del mismo [4].

De hecho, en las últimas décadas han sido numerosos los trabajos centrados en la caracterización de productos de la industria agroalimentaria y especialmente del sector agrícola [5]. Sin embargo, los equipos necesarios para realizar mediciones espectrales tradicionalmente han presentado un elevado coste, resultando este prohibitivo para la mayoría de los usuarios. No obstante, el avance reciente de la industria microelectrónica ha supuesto un abaratamiento del coste y una mejora de las prestaciones en los componentes electrónicos. Esto ha generado un interés creciente por el desarrollo de nuevos dispositivos espectrales con aplicaciones en agricultura de precisión. El abaratamiento de estos dispositivos permitiría transferir el uso de estas metodologías de caracterización no destructivas a lo largo de la cadena de valor de la cosecha, siendo estas asequibles no solo para centros de procesamiento sino también para agricultores y consumidores finales. Bajo este contexto, este trabajo presenta la evaluación de un dispositivo multiespectral de bajo costo para la estimación de parámetros indicadores del estado de maduración, usando en este caso como modelo el cultivo de la vid.

El uso de sensores espectrales como método de caracterización de frutos presenta un carácter indirecto, ya que la determinación del parámetro de interés se realiza a partir de la correlación de este con la huella espectral del fruto. Este método requiere, por tanto, de la construcción de ecuaciones de calibración que estimen los rasgos de calidad en función de las variables espectrales adquiridas. Esto se suele hacer mediante un modelo de datos obtenidos de un conjunto de muestras en las que se han medido tanto las variables espectrales como los rasgos de calidad. Esto entrama una serie de dificultades. Debido a los amplios picos de absorción o de emisión que se superponen, no suele haber ninguna variable espectral que solo esté influida por el parámetro de interés. Por ello es necesario extraer la información relevante contenida en múltiples variables y combinarla mediante el uso de técnicas de análisis de datos multivariantes. En este sentido, han sido numerosos los enfoques estadísticos propuestos para la modelización de parámetros biofísicos de interés agronómico a partir de datos espectrales [6]. Los primeros enfoques propuestos se basaban en métodos paramétricos, los cuales consisten en la determinación de expresiones parametrizadas que relacionen un número limitado de bandas espectrales con la variable biofísica de interés (índices de vegetación) [7]. Estos métodos requieren de un conocimiento previo de la relación entre el parámetro objetivo y las bandas empleadas para su caracterización. Posteriormente a estos surgieron los llamados métodos no paramétricos, los cuales, a diferencia de los métodos paramétricos, optimizan el algoritmo de regresión mediante una fase de aprendizaje autónomo basada en los datos de entrenamiento. Esto ofrece la posibilidad de entrenar con el espectro completo sin la necesidad de poseer un conocimiento previo de que parte de este está influenciado por el parámetro objetivo [8]. A grandes rasgos podemos dividir los métodos no paramétricos en dos categorías, lineales y no lineales. La diferencia entre ambos tipos reside en la naturaleza del modelo de estimación generado. Mientras que los métodos lineales asumen una relación lineal entre las variables espectrales y el parámetro objetivo, los modelos no lineales asumen una relación no lineal. Por tanto, la eficacia de un modelo para la estimación de un parámetro en concreto va a depender de la relación matemática que se dé entre las variables espectrales medidas y el parámetro biofísico de interés. En ocasiones esta relación es directa, con lo cual un modelo lineal puede presentar un buen desempeño. Sin embargo, fenómenos como la dispersión sufrida por la luz al atravesar los tejidos, pueden complicar la relación entre el espectro captado y el parámetro objetivo, haciendo más eficaces a los modelos no lineales.

Con base en lo expuesto, se marca como objetivo del presente trabajo la evaluación comparativa de tres métodos no paramétricos, dos lineales (Regresión lineal múltiple (LMR) y Regresión por mínimos cuadrados parciales (PLSR)) y uno no lineal (Red neuronal artificial (ANN)) para la generación de modelos matemáticos de estimación de parámetros indicadores del estado de maduración de uvas a partir de datos espectrales adquiridos con el sensor multiespectral propuesto.

https://doi.org/10.17979/spudc.9788497498418.0568

2 MATERIALES Y METODOS

2.1 DESCRIPCIÓN DEL DISPOSITIVO PROPUESTO

El componente principal del dispositivo propuesto es una placa de desarrollo AS7265x, basada en la familia de sensores inteligentes AS7265x (AMS, AG, Austria). Este sensor se compone de 3 chips, cada uno de los cuales presenta 6 filtros ópticos independientes. Todo esto le otorga sensibilidad en 18 banda entre los 410 y los 940 nm, con un ancho total a la mitad del máximo (FWHM) de 20 nm por banda. El dispositivo equipa como fuente de luz un componente formado por un conjunto de tres emisores IR de banda ancha (OSLON P1616 SFH 4737, OSRAM, Germany). Este emisor es apropiado para aplicaciones de espectroscopia por su amplio rango de emisión y su intensidad estable a lo largo de todo este. El control del dispositivo es llevado a cabo por una plataforma de desarrollo de bajo costo, concretamente una placa Arduino MKR-Zero (Arduino LLC, Italy). Una vez encendido, el dispositivo genera un nuevo archivo donde almacenar mediciones y queda en estado de reposo, a la espera de ser activado por parte del usuario. En este estado, una activación del pulsador hace que la placa controladora coordine una emisión de luz por parte del emisor led (con una corriente seleccionable por software), con una captura por parte del sensor. Los datos adquiridos y la información de interés, como los nombres de los archivos, se almacenan en una tarjeta SD para su posterior análisis. La placa controladora puede conectarse a un PC para configurar los parámetros internos del dispositivo (tiempo de exposición, ganancia, tiempo de iluminación y corriente de los leds) mediante un software personalizado. Para ayudar y guiar al usuario durante la medición, el dispositivo incluye una pantalla, concretamente un panel OLED de 0,96 pulgadas con una resolución de 128 por 64 píxeles. La disponibilidad de una pantalla integrada permite verificar el correcto funcionamiento del dispositivo en campo, y acceder a información en tiempo real sin la dependencia de un ordenador. La alimentación de todo el sistema se realiza mediante una batería LiPo 2s (polímero de iones litio). La carcasa del dispositivo se diseñó con Freecad 0.16 y se fabricó con una impresora 3D utilizando filamento de impresora 3D de ácido poliláctico (PLA) biodegradable. Se diseñaron tres partes diferentes que se pueden observar en la figura 1. Entre ellas encontramos un mango que alberga el pulsador, una cúpula que aloja la fuente de luz e integra una lámina difusora (OptSaver L-9960, Kimoto LDT, Switzerland) colocada delante del sensor con el objetivo de homogeneizar la radiación recibida por este, y finalmente el cuerpo del dispositivo, que incluye dos partes (delantera y trasera) generando un recipiente apropiado para el resto de los componentes que integra el dispositivo. Todos los elementos del dispositivo están alojados dentro de la carcasa, excepto la batería, que se instala en el exterior para evitar interferencias de temperatura con el sensor que puedan afectar a la precisión de la medición, así como para simplificar su sustitución en campo.

Figura 1: Imagen del sensor multiespectral desarrollado.

2.2 PROTOCOLO EXPERIMENTAL

La generación de modelos matemáticos de estimación de parámetros biofísicos de plantas a partir de datos espectrales requiere una fase previa de recolección de datos. El conjunto de datos obtenido debe incluir el parámetro de interés medido mediante métodos estándar y, por otro lado, la firma espectral de cada una de las muestras. Con este objetivo se realizó un experimento en un viñedo comercial de la variedad Syrah (uva tinta) con la denominación de origen "Condado de Huelva" (Bodegas Contreras Ruiz, S.L, Rociana del Condado, Huelva). La finca experimental presenta una distribución heterogénea en cuanto a composición y estructura del suelo. Esto se ve reflejado en la fisiología de las plantas, encontrándose zonas con diferente ritmo de maduración. Esto nos permitió acceder a un rango de estados de maduración lo suficientemente amplio en una única jornada de muestreo. Dicho muestreo se realizó, por tanto, de manera aleatoria, abarcando toda la superficie de la finca, en una fecha próxima al momento óptimo para la vendimia, según el criterio del gestor de la finca.

Se recogieron un total de 80 racimos de uva que fueron inmediatamente envasados y etiquetados para su transporte al laboratorio. Una vez en laboratorio se procedió a la segunda fase del experimento centrada en la adquisición de la huella espectral de cada una de las muestras usando el dispositivo propuesto. Las mediciones espectrales se realizaron usando una cámara de adquisición, la cual consistió en una estancia de paredes opacas, de modo que el proceso de medición quedara totalmente aislado de contaminación con luz externa. El proceso de

https://doi.org/10.17979/spudc.9788497498418.0568

medición se realizó enfrentando la cúpula del dispositivo con la cara superior de los racimos y realizando dos capturas por muestra, considerándose la reflectancia media de los dos espectros como dato representativo de cada muestra. Una vez cada 15 muestras se tomó una captura de un patrón de reflectancia conocida (53%) (Labsphere, Inc, North Sutton, NH, USA), para calibrar la reflectancia de las muestras y evitar así eventuales errores debidos a variaciones de la fuente de luz. Las 18 señales de reflectancia de la superficie de reflectancia conocida se utilizaron como referencia para normalizar la reflectancia de las muestras en las 18 bandas adquiridas por el dispositivo propuesto. Dicha normalización se realizó según la siguiente fórmula matemática:

$$Rcal_{wl} = \frac{R_{wl} * 0.53}{Rref_{wl}} \quad (1)$$

Donde R_{wl} es el valor de reflectancia medido para una banda espectral dada en una captura de una muestra, $Rref_{wl}$ es el valor de reflectancia medido para esa banda espectral en la captura anterior del patrón, y $Rcal_{wl}$ es el valor corregido de reflectancia en la muestra para la banda dada.

2.3 ANALISIS DE REFERENCIA

Una vez adquirida la huella espectral de cada una de las muestras, estas fueron sometidas a métodos químicos destructivos para caracterizarlas mediante indicadores de su estado real de maduración. Se consideraron como parámetros objetivo el contenido en sólidos solubles totales y la acidez, por ser estos indicadores de madurez ampliamente aceptados en la industria vinícola. El procesado de las muestras consistió en lo siguiente. Cada uno de los racimos fue desgranado, tras esto se seleccionaron al azar 50 uvas por racimo, las cuales fueron licuadas para extraer el mosto. Finalmente, a partir del mosto se determinó el contenido en sólidos solubles totales (Brix°) usando un refractómetro digital (HI96801, Hanna instruments, Spain) y la acidez (g/l ácido sulfúrico) usando un titulador automático (LDS1155500, Dujardin-Salleron, France).

2.4 METODOS DE MODELIZACIÓN

Una vez culminada la fase de recolección de datos, se exploraron diferentes métodos de modelización matemática con el objetivo de averiguar cuál ajustaba mejor la relación entre la información espectral adquirida con el dispositivo propuesto y los parámetros objetivo (SSC y acidez). Se usó el software Orange 3 para el entrenamiento y validación de los modelos. Los cuales fueron entrenados utilizando los 18 datos de reflectancia normalizada adquiridos por el sensor como datos de entrada y un indicador de calidad como objetivo (un modelo por parámetro objetivo). Se empleó el método de validación *leave one out* (LOOCV). Este es un tipo particular de validación cruzada, que consiste en considerar como subconjunto de validación una única muestra, tomando el resto como subconjunto de entrenamiento, lo que obliga a entrenar tantos modelos como número de muestras existan. Tras cada entrenamiento se calcula el error en la estimación de la muestra de validación, considerándose la media de los errores en los sucesivos entrenamientos como el error estimado por el LOOCV. El método LOOCV permite reducir la variabilidad que se origina al dividir aleatoriamente el conjunto de datos en subgrupos de entrenamiento y validación externa (test). Esto se debe a que, por su forma de funcionar, una vez completado el proceso LOOCV todos los datos disponibles se acaban empleando tanto en el entrenamiento como en el test. De este modo, al no haber una disgregación aleatoria de los datos, los resultados del LOOCV son totalmente reproducibles. Por estas características, y dado el limitado volumen del conjunto de datos disponible (n = 80), se consideró el método LOOCV como el más apropiado para establecer una comparación entre diferentes modelos de estimación. Ya que este permite que la aleatoriedad que se genera de la división del conjunto de datos deje de ser un factor con peso en la comparación.

Se evaluaron dos métodos lineales (regresión lineal múltiple (LMR) y regresión por mínimos cuadrados parciales (PLSR)) y un método no lineal (red neuronal artificial (ANN)).

- **Regresión lineal múltiple (LMR):** la LMR permite generar modelos lineales en los que el valor del parámetro objetivo se determina a partir de un conjunto de variables independientes (reflectancia espectral, en este caso). Los coeficientes que determinan el peso de cada variable explicativa son elegidos de forma que la suma de cuadrados entre los valores observados y los estimados sea mínima, reduciendo así la varianza residual. La ecuación resultante recibe el nombre de hiperplano, el cual tendrá tantas dimensiones como variables explicativas se tengan en cuenta [9]. La principal limitación de este método se da cuando existe multicolinealidad entre las variables explicativas. De este modo, los modelos LMR siguen la siguiente ecuación:

$$Y_i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_{ni}) + e_i \quad (2)$$

Donde β_0 es la ordenada en el origen, el valor de la variable Y cuando todos los predictores son cero. B_i es el efecto promedio que tiene el incremento en una

unidad de la variable predictora X_i sobre la variable dependiente Y, manteniéndose constantes el resto de las variables. Por último, e_i es el residuo (diferencia entre el valor observado y el estimado por el modelo).

- **Regresión por mínimos cuadrados parciales (PLSR):** la PLSR se ha utilizado ampliamente en aplicaciones de agricultura de precisión. Este enfoque basado en datos usa métodos matemáticos y estadísticos para extraer información de datos complejos. Concretamente, la PLSR es un enfoque de calibración bilineal que mediante el uso de la compresión de datos reduce la colinealidad de las variables independientes combinando estas para generar nuevas variables ortogonales no correlacionadas entre sí. Estas nuevas variables representan la información estructural relevante que puede utilizarse para predecir la variante dependiente [10]. Matemáticamente, esto se obtiene mediante una descomposición del valor singular de la matriz de producto cruzado. Se han propuesto diferentes algoritmos para obtener esta descomposición. En el caso del algoritmo NIPALS (usado en el presente trabajo), se obtiene la siguiente:

$$X = TP^T + E \quad (3)$$
$$Y = UQ^T + F \quad (4)$$

Donde T, P y E se refieren respectivamente a las puntuaciones, las cargas y las matrices residuales obtenidas de la descomposición bilineal de la matriz X por el modelo PLSR; Q y F son las cargas Y, y las matrices residuales de la descomposición bilineal de la matriz Y por el modelo PLSR. Cabe señalar que en la Ecuación (4) Y se define como una matriz, porque el modelo PLSR puede manejar múltiples variables Y al mismo tiempo y definir un conjunto de variables latentes que capturan la máxima covarianza entre la matriz de las variables X y la matriz de las variables Y. Sin embargo, en este trabajo se generaron modelos independientes para la estimación de cada variable objetivo.

Las características mencionadas hacen de PLSR un método adecuado para la recuperación de parámetros biofísicos de la vegetación a través de información espectral, especialmente cuando los predictores presentan multicolinealidad.

- **Redes neuronales artificiales (ANN):** Una ANN es un método no lineal impulsado por aprendizaje automático. Este método consiste en una estructura de neuronas conectadas entre sí y organizadas en capas. Las neuronas de las distintas capas están interconectadas, y cada conexión tiene un peso específico. Cada neurona realiza fundamentalmente una regresión lineal seguida de una función no lineal. En resumen, la arquitectura de la ANN trabaja para minimizar la desviación media cuadrática mediante la regla de aprendizaje de corrección de errores. De este modo, el error se reduce ajustando el peso de cada capa de neuronas. Estas características dan lugar a una capacidad extraordinaria para vincular información espectral compleja con parámetros clave sin ninguna restricción en la distribución de la muestra. Esto hace que los enfoques basados en ANN sean adecuados para definir las intrincadas relaciones no lineales que normalmente existen entre las firmas espectrales y los parámetros biofísicos de plantas.

  En este trabajo concretamente se usó un modelo ANN basado en un algoritmo de perceptrón multicapa con retro propagación. La arquitectura de la red estaba compuesta por una capa oculta con seis neuronas, 18 entradas (18 datos de reflectancia normalizada) y una capa de salida. Se empleó la función de identidad como función de activación de la capa oculta. Y finalmente, como algoritmo de aprendizaje se usó una variante de la familia de los cuasi-Newton, denominado L-BFGS-B.

2.5 CRITERIOS DE EVALUACIÓN DEL REDIMIENTO DE LOS MODELOS

El rendimiento de los modelos de estimación se midió mediante el coeficiente de determinación (R^2), el error cuadrático medio (RMSE) y el error absoluto medio (MAE). Estos indicadores se calcularon a partir de los valores de referencia de SSC y acidez medidos mediante métodos estándar y los valores estimados por los diferentes modelos durante el proceso de entrenamiento y validación (*Leave one out*).

3 RESULTADOS Y DISCUSIÓN

La tabla 1 recoge diferentes indicadores del rendimiento de los modelos evaluados en la estimación de la concentración de SSC de uvas a partir de datos espectrales adquiridos con el dispositivo propuesto. En ella destaca el rendimiento de la ANN levemente por encima de la PLSR según los tres indicadores considerados (R^2, RMSE, y MAE), por otro lado, tendríamos la LMR con un rendimiento por debajo de los otros dos modelos

https://doi.org/10.17979/spudc.9788497498418.0568

considerados. La menor eficacia de la LMR con respecto a la PLSR se puede explicar a razón de una probable multicolinealidad entre los 18 datos de reflectancia captados por el sensor. La PLSR define nuevas variables como combinaciones lineales ortogonales (no correlacionadas) de las variables originales que capturan al máximo la covarianza entre las variables de entrada (datos espectrales) y el parámetro objetivo (SSC). Esto supone una reducción de la multicolinealidad de las variables independientes, lo cual podría explicar el mejor desempeño del modelo basado en PLSR con respecto a la LMR. Por otro lado, la ANN mostró los mejores resultados entre los tres modelos evaluados, con el mayor valor R^2 (0.59) entre los valores observados y estimados de SSC, y los valores más reducidos de RMSE y MAE (1.23 y 0.96).

Tabla 1: Valores de R^2, RMSE, y MAE entre los valores de SSC medidos por métodos estándar y los estimados por los modelos basados en LMR, PLSR, y ANN.

Solidos solubles totales (Brix)			
	R^2	**RMSE**	**MAE**
LMR	0.48	1.42	1.13
PLSR	0.58	1.27	1.01
ANN	0.59	1.23	0.96

Por otro lado, la Tabla 2 recoge los mismos indicadores de eficacia mostrados por los modelos en la estimación de la acidez de las uvas. La tendencia en cuanto al desempeño de los diferentes modelos fue similar a la mostrada en la estimación de SSC, siendo de nuevo la ANN el más preciso, seguido de PLSR, y LMR. Sin embargo, en este caso el ajuste mostrado por los cuatro modelos fue superior al mostrado en la estimación de SSC. Destaca de nuevo el desempeño mostrado por del modelo basado en ANN con el mayor coeficiente R^2 (0.64) y los menores errores en la estimación (RMSE = 0.87, y MAE = 0.69). La superioridad de los modelos basados en ANN en el ajuste de ambos parámetros podría deberse a una cierta no linealidad en la relación entre la firma espectral medida con el dispositivo propuesto y el SSC y especialmente la acidez de las uvas. En este sentido, su carácter no-lineal le otorga una mayor flexibilidad que le permite ajustar mejor el espacio de características.

Tabla 2: Valores de R^2, RMSE, y MAE entre los valores de acidez medidos por métodos estándar y los estimados por los modelos basados en LMR, PLSR, y ANN.

Acidez (g/l ácido clorhídrico)			
	R^2	**RMSE**	**MAE**
LMR	0.52	1.02	0.82
PLSR	0.60	0.92	0.73
ANN	0.64	0.87	0.69

4 CONCLUSIONES

El objetivo del presente trabajo era la evaluación comparativa de tres modelos matemáticos de estimación en el ajuste de datos espectrales adquiridos con un sensor multiespectral de bajo costo y el SSC y la acidez de uvas tintas medidos por métodos estándar. Entre los cuatro modelos evaluados, el enfoque basado en ANN demostró ser el más eficaz. El rendimiento alcanzado por los cuatro modelos pone de manifiesto la relación existente entre la información adquirida por el sensor y los indicadores de estado de maduración considerados. Los resultados alcanzados son prometedores de cara al establecimiento de un método de caracterización del estado de calidad de uvas rápido, eficaz, y barato. La reducción de costes posibilitaría el acceso a estas tecnologías a otras fases de la cadena de valor de la cosecha, haciéndolas asequibles para agricultores y consumidores, además de centros de procesamiento. Todo esto redundaría en una mayor seguridad alimentaria y un incremento en la calidad del producto final, ya que posibilitaría una monitorización continua del estado de calidad del fruto.

Agradecimientos

Este trabajo fue apoyado por la subvención PID2020-119217RA-I00 financiado por MCIN/AEI/ 10.13039/ 501100011033, y la beca IJC2019-040114-I financiada por MCIN/AEI/ 10.13039/501100011033.

English summary

COMPARISON OF STATISTICAL MODELS IN THE ESTIMATION OF QUALITY INDICATORS OF RED GRAPES FROM SPECTRAL INFORMATION

Abstract

The methods traditionally used for the determination of fruit quality status have a low spatial and temporal resolution due to their limitations (high cost and wide time gap between sampling and access to information). In the last decades, numerous research has informed about the potential of spectroscopy based methods for estimate plant biophysical parameters. In addition, the recent boom in the electronics industry has led to cheaper components, generating interest in the development of new devices. Encouraged by this context, this work presents a low-cost multispectral device based on a commercial sensor (AS7265x, AMS) sensitive to 18

bands between 410 and 940 nm. Aiming at the comparative evaluation of 3 non-parametric estimation models (two linear (Multiple Linear Regression and Partial Least Squares Regression) and one non-linear (Artificial Neural Networks) in the modelling of quality indicators of red grapes (total soluble solids and acidity). Among the models explored, the neural network proved to be the most effective in adjusting the relationship between the spectral information acquired with the proposed sensor and the quality indicators considered.

Keywords: AS7265x, multispectral, multiple linear regression, partial least square regression, artificial neural networks, ripening, grapes.

Referencias

[1] Li, B.; Lecourt, J.; Bishop, G (2018) Advances in Non-Destructive Early Assessment of Fruit Ripeness towards Defining Optimal Time of Harvest and Yield Prediction—A Review. *Plants*, *7*, 3.

[2] Vanoli, Maristella; Buccheri, M (2012) Overview of the methods for assessing harvest maturity. *Stewart Postharvest Rev. 8*, 1–11.

[3] Cattaneo, T.M.P.; Stellari, A (2019) Review: NIR Spectroscopy as a Suitable Tool for the Investigation of the Horticultural Field. *Agronomy*, *9*, 503.

[4] Lu, R.; Van Beers, R.; Saeys, W.; Li, C.; Cen, H (2020) Measurement of optical properties of fruits and vegetables: A review. *Postharvest Biol. Technol. 159*, 111003.

[5] Nicolaï, B.M.; Beullens, K.; Bobelyn, E.; Peirs, A.; Saeys, W.; Theron, K.I.; Lammertyn, J (2007) Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol. Technol. 46*, 99–118.

[6] Verrelst, J.; Malenovský, Z.; Van der Tol, C.; Camps-Valls, G.; Gastellu-Etchegorry, J.P.; Lewis, P.; North, P.; Moreno, J (2019) Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surv. Geophys. 40*, 589–629.

[7] Clevers, J.G.P.W (2014) Beyond NDVI: Extraction of biophysical variables from remote sensing imagery. *Remote Sens. Digit. Image Process. 18*, 363–381.

[8] Hastie, T.; Friedman, J.; Tibshirani, R (2001) *The Elements of Statistical Learning*; Springer Series in Statistics; Springer New York: New York, NY; ISBN 978-1-4899-0519-2.

[9] Saeys, W.; Nguyen Do Trong, N.; Van Beers, R.; Nicolaï, B.M (2019) Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. *Postharvest Biol. Technol. 158*, 110981.

[10] Wold, S.; Sjöström, M.; Eriksson, L (2001) PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst. 58*, 109–130.

 https://doi.org/10.17979/spudc.9788497498418.0568