

Combinación de modelos basados en mecanismos y basados en datos para la predicción de la evolución de calidad en pescado fresco

Miguel Novoa^a, Míriam R. García^a, Silvia Muñoz^b, Juan R. Herrera^b,
Marta López-Cabo^b, Carlos Vilas^{a,*}

^aGrupo de Bioingeniería y Bioprocesos (Bio2Eng), IIM-CSIC, Eduardo Cabello, 6, 36208 Vigo, España

^bGrupo de Microbiología y Tecnología de Productos Marinos, IIM-CSIC, Eduardo Cabello, 6, 36208 Vigo, España

To cite this article: Novoa, M., García, M.R., Muñoz, S., Herrera, J.R., López-Cabo, M., Vilas, C. 2023. Mathematical model to predict fresh fish quality. XLIV Jornadas de Automática, 435-440. <https://doi.org/10.17979/spudc.9788497498609.435>

Resumen

El desperdicio alimentario supone un gran problema de sostenibilidad. El pescado, en concreto, es un producto especialmente perecedero, degradándose con mucha rapidez cuando la temperatura de almacenamiento no se controla adecuadamente. Los modelos matemáticos son herramientas útiles para predecir la evolución de la calidad del pescado durante el transporte y almacenamiento. Conocer la dinámica de la degradación con antelación también nos permitirá definir estrategias para reducir el desperdicio alimentario y aumentar el valor añadido del pescado. En este trabajo se presenta una nueva metodología para la descripción de la evolución de calidad de filetes de merluza (*Merluccius merluccius*) envasada en atmósferas modificadas durante el transporte y almacenamiento. Dicha metodología consta de dos partes: un modelo basado en los mecanismos del proceso que describe la evolución de dos de los indicadores de calidad más utilizados (contenido bacteriano y bases volátiles), y un modelo de aprendizaje automático que permite correlacionar dichos indicadores con un indicador sensorial, el QIM (del inglés, *Quality Index Method*). Para la evolución de las bases volátiles se propone un nuevo modelo basado en la ley de potencias. Además, en este trabajo, los parámetros desconocidos del modelo se estiman utilizando datos experimentales. En el caso de los modelos de aprendizaje automático, se comparan técnicas de redes neuronales y árboles de decisión utilizando diferentes indicadores del error. El modelo combinado supone un avance en la predicción de calidad, ya que permite predecir el QIM a partir, exclusivamente, de datos de temperatura y obtener resultados satisfactorios.

Palabras clave: Calidad de pescado, Modelado matemático, Estimación de parámetros, Indicadores de calidad, Microbiología predictiva

Combination of mechanism-based and data-based models for the prediction of quality in fresh fish

Abstract

Food waste is a significant sustainability problem. Fish, in particular, is a very perishable product. In fact, spoilage increases rapidly when the storage temperature is not adequately controlled. Mathematical models are useful tools to predict fish quality evolution during transport and storage. This knowledge will help to develop strategies to reduce food waste and to increase the product value. In this work, we propose a new methodology to predict the evolution of quality, during transport and storage, of hake fillets packed in modified atmospheres. It consists of two parts: a model based on the process mechanisms to describe the evolution of two highly used quality indicators (bacterial content and volatile bases), and a machine learning model to correlate such indicators with one of the most employed sensory indexes: the Quality Index Method (QIM). Moreover, we propose a *Power Law* model to describe volatile bases evolution. We estimate the unknown parameters using experimental data. In the machine learning part, we compare random forest models and neural networks using several statistical techniques. Overall, this methodology is an improvement in fish quality prediction because it makes QIM prediction using just temperature data with promising results.

Keywords: Fish quality, Mathematical modelling, Parameter estimation, Quality Indicators, Predictive microbiology

*Autor para correspondencia: carlosvf@iim.csic.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

1. Introducción

La rápida degradación de la calidad de alimentos como el pescado supone que cerca de un 15 % de la captura se desperdicie (Tavares et al., 2021). La temperatura es el factor externo que más afecta a la frescura del pescado, de hecho, si las condiciones de transporte y almacenamiento no son adecuadas, el desperdicio puede llegar al 35 % (Corradini, 2018). Los modelos matemáticos permiten predecir la evolución de la calidad y la vida útil del pescado y conocer los efectos de los abusos de temperatura durante el transporte y almacenamiento. Esto facilita el diseño de estrategias para reducir el desperdicio alimentario, aportando, además, información de interés al consumidor.

Sin embargo, la calidad del pescado es un concepto complejo en el que intervienen diferentes atributos nutricionales y organolépticos. Algunos de los indicadores más utilizados para su determinación son: la oxidación lipídica, la degradación de nucleótidos, el contenido microbiológico, la concentración de nitrógeno básico volátil total (N-BVT), o ciertos indicadores sensoriales como el QIM (del inglés *Quality Index Method*). En la literatura se han desarrollado diferentes modelos matemáticos para describir dichos indicadores (García et al., 2022), aunque algunos de esos modelos presentan ciertas limitaciones.

El modelo más utilizado para describir el crecimiento de bacterias es el modelo logístico (Baranyi and Roberts, 1994). En cambio, los modelos para describir el comportamiento del N-BVT tienen, generalmente, forma exponencial y utilizan la ecuación de Arrhenius para establecer la relación entre el ratio de crecimiento y la temperatura. García et al. (2017) proponen un modelo que incluye los organismos psicrotrofos como productores de N-BVT. Además, introducen un término para describir la fase de latencia inicial observada. Sin embargo, dicho término depende explícitamente del tiempo. Por lo tanto, requiere conocer el momento de la muerte del pescado, dato que no siempre está disponible. El QIM es un indicador sensorial. Su evaluación se realiza de forma subjetiva mediante un panel de expertos que valoran distintos atributos sensoriales del pescado (Escribano et al., 2015). En el trabajo de García et al. (2017) se propone una ecuación empírica para calcular el QIM a partir del contenido en psicrotrofos y bases volátiles. Como se sugiere en García et al. (2022), se requiere desarrollar otras metodologías que permitan comparar los resultados obtenidos.

El objetivo de este trabajo es desarrollar una metodología que permita describir la evolución de la calidad de pescado a lo largo de toda la cadena de valor, es decir, desde la pesca hasta su venta al consumidor final. Como caso de estudio se utilizarán filetes de merluza (*Merluccius merluccius*) almacenados en atmósferas modificadas. Además, como indicadores de calidad se han seleccionado tres de los más utilizados en la literatura: contenido microbiológico, N-BVT y QIM. Para la descripción de la evolución de los dos primeros se utilizarán modelos basados en los mecanismos del proceso. En particular, se utilizará el modelo logístico para representar el contenido bacteriano, ya que ha sido ampliamente testado en la literatura. Para la evolución de N-BVT, se proponen dos modelos: exponencial y ley de potencias. Este último permite describir el retraso inicial observado en los datos experimentales sin introducir una dependencia explícita del tiempo transcurrido desde la muerte del pescado. En lo que se refiere al modelado del QIM,

dado que se trata de un indicador subjetivo, se propone el uso de técnicas de aprendizaje automático. En particular, se considerarán dos opciones: una red neuronal y un modelo de bosque aleatorio. Dichos modelos correlacionarán el contenido bacteriano y de N-BVT con el QIM. Se considerarán las bacterias *Pseudomonas spp.* y *Shewanella spp.*, ya que son dos de las principales responsables de la degradación de pescado fresco. El resultado final de la metodología es un modelo combinado que, en una primera fase, genera datos de contenido bacteriano y N-BVT utilizando modelos basados en los mecanismos del proceso. En una segunda fase, se alimenta el modelo de aprendizaje automático con los datos de la fase anterior para calcular un valor del QIM. La temperatura a lo largo del proceso se introduce como una entrada en los modelos mecanísticos.

2. Materiales y métodos

2.1. Métodos experimentales

2.1.1. Descripción de filetes y pesca

La merluza (*Merluccius merluccius*) es capturada en la zona FAO 27.9, posteriormente es fileteada y envasada en atmósferas modificadas (47 % CO_2 / 0.8 % O_2) en bandejas de aproximadamente 300 g. El grupo Profand S.L. se encarga de suministrar los filetes frescos de merluza con piel. Se llevaron a cabo tres ensayos, fijándose la temperatura de refrigeración a 1, 2.5 y 5 °C. Además, para cada tiempo de muestreo, se seleccionaron y analizaron 3 bandejas (tres repeticiones). Las muestras se emplearon para el análisis sensorial, microbiológico y bioquímico. El número de tiempos de muestreo en cada experimento fue 6, 8 y 8, respectivamente.

2.1.2. Análisis microbiológico

Dos de las bacterias más influyentes en la degradación del pescado fresco son la *Shewanella spp.* y las *Pseudomonas spp.* Es por ello que en este trabajo se han considerado ambas. Para medir la concentración de bacterias se mezclaron, primero, 25 g de músculo de pescado de cada bandeja y con 225 ml de agua de peptona tamponada previamente esterilizada donde se homogeneizó con un Stomacher durante 1 minuto. De los homogeneizados se sembraron alícuotas (0.1 ml) de diluciones adecuadas en agar rojo fenol de almidón glutamato (GSP), agar hierro Lyngby (IAL) y agar de recuento en placa (PCA).

Las *Pseudomonas spp.* se contaron como colonias crecidas en placas GSP y las colonias negras formadas en IAL se contaron para enumerar las bacterias productoras de H_2S identificadas como *Shewanella spp.* Tras 48 horas de incubación a 25 °C, se contaron las colonias crecidas en medio GSP identificadas como *Pseudomonas spp.*, y las colonias negras crecidas en IAL, como *Shewanella spp.* El conteo de los microorganismos psicrotrofos aerobios crecidos en PCA se produjo tras la incubación a 17 °C durante 3 días.

2.1.3. Análisis de bases volátiles

El contenido de nitrógeno básico volátil total (N-BVT) se determinó mediante el método de Lücke and Geidel (1935), siguiendo las modificaciones de Antonacopoulos (1960). Brevemente, se mezclaron aproximadamente 10 g de músculo de merluza y se transfirieron a una unidad de destilación Kjeldahl. A continuación, se añadieron unas gotas de antiespumante de

silicona y 2 g de MgO y se recogieron alrededor de 75 ml de destilado al vapor sobre 5 ml de una disolución de Ácido Bórico 1 % (p/v) con indicador verde de bromocresol y rojo de metilo. A continuación, se determinó el contenido de N-BVT por titulación del destilado con HCl 0,1 N.

2.1.4. Análisis sensorial

Para el análisis sensorial se utilizó el QIM como indicador, el cual se adaptó a los filetes de merluza. La evaluación de los diferentes atributos en los filetes de pescado (apariencia, color, textura y olor) se llevó a cabo de acuerdo con los requisitos de las normas ISO (ISO 8586:2012). Para ello, un panel de 7 expertos asigna una puntuación de demérito de 0 a 3 en la evaluación de la firmeza muscular, olor y sequedad, y una puntuación de demérito de 0 a 2 para las marcas o hematomas. El valor del QIM es el resultado de sumar las puntuaciones de todos los atributos. A mayor puntuación, menor calidad.

2.2. Modelos matemáticos

2.2.1. Modelos basados en los mecanismos del proceso

Para describir la evolución de poblaciones bacterianas se ha seleccionado el modelo logístico (Baranyi and Roberts, 1994) por ser uno de los más utilizados. Además, para describir el aumento del ratio de crecimiento con la temperatura de almacenamiento (modelo secundario) se ha elegido la ecuación de Ratkowsky (Ratkowsky et al., 1982). Matemáticamente, el modelo se representa de la forma (García et al., 2015):

$$\frac{dn}{dt} = \mu(1 - 10^{n-n^*}), \text{ con } \mu = b^2(T - T_0)^2, \quad (1)$$

siendo n la concentración de bacterias (expresada en $\log(CFU/g)$, donde CFU son unidades formadoras de colonias) en el músculo de pescado a un tiempo t . n^* indica la concentración máxima de bacterias. La variable μ representa la velocidad de crecimiento de las bacterias. Dicha variable depende de la temperatura de almacenamiento (T), de una temperatura conceptual mínima para el crecimiento bacteriano (T_0) y de una constante (b). Aunque la estructura del modelo es la misma tanto para *Shewanella spp.* como para *Pseudomonas spp.*, los parámetros n^* , b , y T_0 variarán según la bacteria considerada.

Como se menciona en la introducción, el nitrógeno básico volátil total (N-BVT) se produce por la acción bacteriana en el alimento. A concentraciones bajas, se puede observar experimentalmente una fase de latencia. A concentraciones intermedias y altas, el comportamiento es exponencial. En este trabajo se propone el uso de la ley de potencias para describir la evolución de N-BVT, ya que permite describir la fase de latencia a concentraciones bajas y la fase exponencial a concentraciones altas. En particular, se considera una reacción de orden N con dependencia de la concentración de bacterias (n_s), de la forma:

$$\frac{dN-BVT}{dt} = k \cdot n_s^N \quad (2)$$

donde n_s es la suma de las concentraciones de *Shewanella spp.* y *Pseudomonas spp.* Este modelo se comparará con el modelo exponencial, propuesto en la literatura (Tsironi et al., 2009):

$$\frac{dN-BVT}{dt} = k \cdot N-BVT, \quad (3)$$

En ambos modelos se utiliza la ecuación de Arrhenius para describir la dependencia del ratio de crecimiento (k) con la temperatura (modelo secundario)(Ratkowsky et al., 1982):

$$k = k_{ref} \exp \left[\frac{E_a}{R} \left(\frac{1}{T_{ref} + 273,15} - \frac{1}{T + 273,15} \right) \right] \quad (4)$$

donde k_{ref} es el valor del coeficiente de crecimiento a una temperatura de referencia $T_{ref} = 4^\circ C$, E_a es la energía de activación y $R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$ es la constante universal de los gases. Los parámetros desconocidos son: n^* , T_0 y b para cada una de las bacterias, y k_{ref} , E_a , y N .

2.2.2. Modelos de aprendizaje automático

Dado que el QIM es un indicador de calidad subjetivo, en el que se combinan distintos atributos (ver Sección 2.1.4), se ha optado por utilizar modelos de aprendizaje automático. Se considerarán dos enfoques: redes neuronales y bosques aleatorios.

Las redes neuronales artificiales (McCulloch and Pitts, 1943) consisten en estructuras de capas formadas por nodos que se conectan entre sí. Para introducir no linealidades en la salida de los nodos se utilizan funciones de activación (Sharma et al., 2017). En nuestro caso, ReLU (*Rectified Linear Unit*) para la primera capa (compuesta por 3 nodos) y tangente hiperbólica para la capa interna (compuesta por 3 nodos). Para elegir el número de nodos de la capa oculta hemos seguido las reglas empíricas propuestas por Heaton (2008).

Los modelos de bosque aleatorio (Breiman, 2001) se basan en la construcción de árboles de decisión que calculan predicciones independientes de la variable a estimar. Después, se calcula la media de estas predicciones para dar un resultado final.

Para entrenar el modelo de bosque aleatorio se utilizan los datos experimentales de los indicadores de calidad. Las redes neuronales, en cambio, necesitan una gran cantidad de datos para entrenarse. Para aumentar el número de muestras, se utilizan datos sintéticos generados mediante modelos de aprendizaje automático, a partir de los datos experimentales disponibles (<https://gretel.ai/>). Cabe destacar que el proceso de generación de datos sintéticos fiables depende de la cantidad y calidad de los datos experimentales usados en el mismo.

Para evaluar la calidad de los modelos, se usa validación cruzada de k iteraciones y validación cruzada repetida de k iteraciones (Hastie et al., 2009). Estas técnicas consisten en dividir el conjunto de datos en n subgrupos. Parte de esos subgrupos se utilizan para estimar los parámetros del modelo y la otra parte para la validación del mismo. En la validación cruzada repetida se realiza la validación para diferentes agrupaciones de los datos con el objetivo de mejorar la estimación del modelo. Para seleccionar los hiperparámetros óptimos (por ejemplo, el número de árboles de decisión en el caso de bosque aleatorio o el número de épocas en la red neuronal) se utilizan técnicas de validación cruzada anidada. Estas técnicas constan de dos bucles anidados, el bucle interno selecciona el mejor conjunto de hiperparámetros y el bucle externo calcula el error del modelo.

Para la implementación de los modelos de aprendizaje automático se han seleccionado las librerías Scikit-learn (Pedregosa et al., 2011) y Keras (Chollet et al., 2015) de Python.

Durante el entrenamiento y validación de los modelos de aprendizaje automático se utilizan los datos obtenidos experimentalmente de todos los indicadores. Sin embargo, una vez

validados, se utilizarán los resultados de los modelos basados en los mecanismos del proceso (N-BVT, *Pseudomonas spp.*, y *Shewanella spp.*) para obtener un valor estimado del QIM.

2.3. Calibración de los modelos basados en mecanismos

El valor de los parámetros desconocidos del modelo (1)-(4) se calcula como el que minimiza la distancia entre la predicción del modelo y los datos experimentales. Dado que se cuenta con tres repeticiones de los datos experimentales en cada punto de muestreo, se puede estimar la variabilidad entre muestras y el error experimental. Para dotar de mayor importancia aquellos puntos con menor variabilidad, se usa la verosimilitud (*log-likelihood*) para el cálculo de dicha distancia (función objetivo):

$$J_{llk}(\theta) = -\frac{1}{2} \sum_{k=1}^{n_c} \sum_{i=1}^{n_{sk}} \left[\log(2\pi) + \log(\sigma_{x_{k,i}}^2) + \left(\frac{x_{k,i}(\theta) - x_{k,i}^{exp}}{\sigma_{x_{k,i}}} \right)^2 \right], \quad (5)$$

donde θ se corresponde con el vector de parámetros a estimar, n_c es el número de variables medidas experimentalmente (observables), n_{sk} indica el número de puntos de muestreo de cada observable, y $x_{k,i}(\theta)$ y $x_{k,i}^{exp}$ representan, ambos, el valor del observable x_k a tiempo t_i . El primero ($x_{k,i}$) se refiere al valor calculado por el modelo, mientras que el segundo ($x_{k,i}^{exp}$) es el valor experimental. $\sigma_{x_{k,i}}(\theta)$ es la desviación estándar de la muestra del observable x_k en el tiempo t_i . En este trabajo se consideran tres observables, i.e., las concentraciones de bases volátiles, *Pseudomonas spp.* y *Shewanella spp.*

Es habitual que la función objetivo (5) sea multimodal (contiene varios mínimos en el espacio de búsqueda). Por lo tanto, para encontrar el mínimo global, se utiliza un método híbrido que combina búsqueda global con local. En particular, se utiliza el algoritmo de *Scatter Search*, *ESS* (Egea et al., 2009), un método basado en poblaciones e implementado en la toolbox de Matlab, *AMIGO* (Balsa-Canto et al., 2016).

Una de las formas de seleccionar el mejor modelo cuando hay varios disponibles consiste en comparar el valor de la función objetivo de cada uno de ellos. Generalmente, el modelo con mayor número de parámetros será el que proporcione el menor valor de la función objetivo. Sin embargo, aunque dicho modelo resulte en un mejor ajuste, podría llevar a una peor predicción cuando las condiciones experimentales son distintas a las utilizadas para el ajuste (sobrestimación). En este trabajo, para evitar este problema se utilizarán dos criterios que penalizan el número de parámetros de cada modelo: el Criterio de Información de Akaike corregido (AIC, por sus siglas en inglés) (Akaike, 1974) y el Criterio de Información Bayesiana (BIC, por sus siglas en inglés) (Raftery, 1995).

El AIC corregido se calcula de la siguiente forma:

$$AIC_c = J_{llk} + \frac{2n_p(n_d + 2)}{n_d - n_p + 1} \quad (6)$$

siendo n_d el número de medidas experimentales y n_p el número de parámetros.

El BIC se obtiene mediante la siguiente expresión:

$$BIC = -2J_{llk} + \log(n_d) \cdot k \quad (7)$$

La penalización por el número de parámetros en el BIC es mayor que en el AIC. Cuanto más pequeño sea el valor de ambos criterios, mejor será el modelo.

3. Resultados y discusión

3.1. Modelos basados en los mecanismos del proceso

El procedimiento descrito en la sección anterior se ha seguido para calcular el valor de los parámetros desconocidos del modelo (1)-(3). De las tres series de datos experimentales disponibles, dos se utilizaron en dicho procedimiento y la tercera se reservó para validación. Los valores óptimos de los parámetros se muestran en la Tabla 1.

Modelo N-BVT	Parámetro	Valor estimado	Unidades
Ley de potencias	E_a	$2,4 \times 10^5$	J/mol
	N	3,9	
	k_{ref}	$1,0 \times 10^{-3}$	d^{-1}
	b_{ps}	$2,3 \times 10^{-1}$	$d^{-1} \cdot ^\circ C^{-2}$
	b_{sh}	$9,6 \times 10^{-2}$	$d^{-1} \cdot ^\circ C^{-2}$
	$T_{0,ps}$	$3,2 \times 10^{-1}$	$^\circ C$
	$T_{0,sh}$	-5,0	$^\circ C$
	n_{ps}^*	6,0	log(CFU/g)
	n_{sh}^*	5,5	log(CFU/g)
	Exponencial	E_a	$5,5 \times 10^4$
k_{ref}		$8,3 \times 10^{-2}$	d^{-1}
b_{ps}		$9,1 \times 10^{-2}$	$d^{-1} \cdot ^\circ C^{-2}$
b_{sh}		$9,5 \times 10^{-2}$	$d^{-1} \cdot ^\circ C^{-2}$
$T_{0,ps}$		$-3,4 \times 10^{-1}$	$^\circ C$
$T_{0,sh}$		-5,0	$^\circ C$
n_{ps}^*		3,6	log(CFU/g)
n_{sh}^*		5,4	log(CFU/g)

Tabla 1: Valores estimados de los parámetros de los modelos de N-BVT, *Pseudomonas spp.* y *Shewanella spp.*. d son días.

La primera fila de gráficas en la Figura 1 muestra los perfiles de temperatura medidos en cada ensayo. En el primer ensayo se observan varias oscilaciones de temperatura relativamente grandes (hasta $4 \text{ }^\circ C$) causadas por la abertura de la puerta de la cámara durante varios minutos. En los otros dos ensayos las oscilaciones son mucho menores. Las otras tres filas de gráficas muestran la comparación entre los datos experimentales (puntos), con sus errores (barras), y los resultados de simulación del modelo (1)-(3) utilizando los valores óptimos de los parámetros (líneas). Las dos columnas de la izquierda muestran los resultados en los experimentos utilizados para el ajuste, mientras que la columna de la derecha se refiere a los resultados de validación. La segunda fila de gráficas muestra los resultados para N-BVT. Las líneas continuas representan los resultados de simulación obtenidos utilizando el modelo exponencial, mientras que las líneas discontinuas se refieren al modelo de la ley de potencias. Como se aprecia en la figura, este último representa mejor la evolución de N-BVT en los tres experimentos, particularmente durante la fase inicial (latencia).

Para una comparación cuantitativa entre ambos modelos, se han calculado el AIC y el BIC de cada uno. Los valores alcanzados por ambos criterios y el valor de la función objetivo del problema de estimación de parámetros se muestran en la Tabla 2. El modelo de la ley de potencias mejora al exponencial en los tres criterios y, por lo tanto, representa mejor el comportamiento experimental sin incurrir en el problema de sobreestimación.

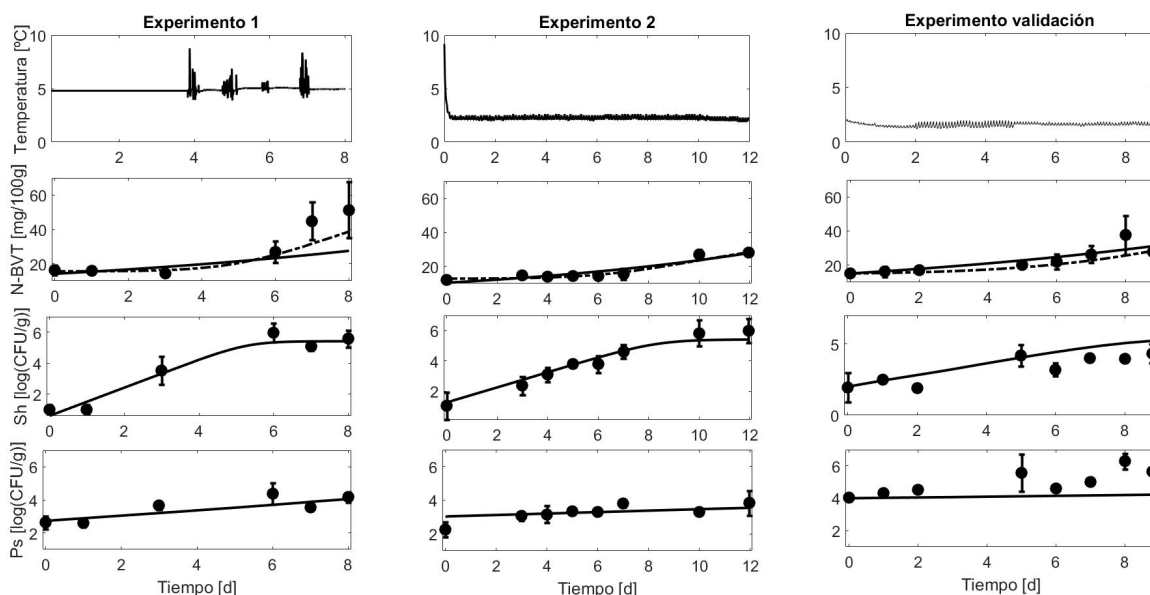


Figura 1: Perfiles de temperatura medidos en cada ensayo (primera fila de gráficas) y comparación entre los datos experimentales (puntos), con sus desviaciones (barras), y los resultados de simulación de los modelos calibrados (líneas).

Modelo	J_{llk}	AIC_c	BIC
Modelo exponencial	60.7	81.1	90.6
Modelo de la ley de potencias	42.6	66.3	76.3

Tabla 2: Valor de función objetivo, criterio de Akaike corregido y criterio Bayesiano para el modelo exponencial y modelo de la ley de potencias.

Las dos últimas filas de gráficas en la Figura 1 muestran los resultados para las dos poblaciones de bacterias consideradas. Se puede observar que el modelo seleccionado representa satisfactoriamente el comportamiento experimental. Ambas poblaciones crecen a distinta velocidad, mayor en el caso de *Shewanella spp.* Además, nótese que en el experimento a menor temperatura el contenido de *Pseudomonas spp.* es mayor que el de los otros dos experimentos. Esto se debe a que las muestras de pescado recepcionadas estaban más degradadas en el caso del experimento a menor temperatura.

3.2. Modelos de aprendizaje automático

Los datos de los experimentos 2 y 3 (14 datos) se usaron para entrenar los modelos de aprendizaje automático y el experimento 1 se reservó para su validación. La razón de esto es la conveniencia de usar el mayor número de datos para el entrenamiento de los algoritmos y el experimento 1 es el que tiene menos datos experimentales. La red neuronal fue entrenada con 1000 datos sintéticos generados con los modelos de aprendizaje profundo de la plataforma (<https://gretel.ai/>) a partir de los datos de los experimentos 2 y 3.

Para la validación cruzada se ha dividido el conjunto de datos en 10 grupos. En cada iteración 9 grupos se utilizan para el entrenamiento y el restante para la validación. Los grupos de entrenamiento y validación cambian en cada iteración. En la validación cruzada repetida realizamos la validación cruzada para 3 agrupaciones diferentes del conjunto total de datos, aumentando el número de validaciones de los modelos.

En la tabla 3 se muestran diferentes métricas (coeficiente de determinación (R^2) y error cuadrático medio (MSE) en el experimento 1, y error absoluto medio (MAE) de la validación cruzada y la validación cruzada repetida) que permiten comparar la capacidad de los modelos para representar los datos experimentales. El modelo de bosque aleatorio es mejor en las cuatro métricas. Esto puede ser debido a que los datos sintéticos utilizados para entrenar la red neuronal fueron generados a partir de un conjunto pequeño de datos experimentales.

Modelo	R^2	MSE	MAE	
			$k - FOLD$	$Rep\ k - FOLD$
Red neuronal	0.80	1.92	2.02 ± 0.83	1.56 ± 0.32
Bosque aleatorio	0.96	0.39	1.14 ± 0.68	1.08 ± 0.64

Tabla 3: Valores del coeficiente de determinación (R^2), error cuadrático medio (MSE), y error absoluto medio (MAE) de los modelos de aprendizaje profundo utilizados en el trabajo.

En la figura 2 se muestra la comparación entre los datos experimentales de QIM y los resultados de los dos modelos de aprendizaje automático (red neuronal y bosque aleatorio) para el experimento 1. Los datos de contenido de N-BVT, *Pseudomonas spp.* y *Shewanella spp.*, necesarios para alimentar los modelos de red neuronal y bosque aleatorio, se han obtenido a partir de la simulación del modelo (1)-(3) con valores óptimos de los parámetros. Como se observa, el modelo de bosque aleatorio es capaz que predecir con mayor acierto el comportamiento observado experimentalmente para el QIM, especialmente en la primera fase de retraso que la red neuronal no es capaz de captar. Probablemente, el peor funcionamiento de la red neuronal se deba, en parte, al uso de datos sintéticos generados a partir de un conjunto pequeño de datos.

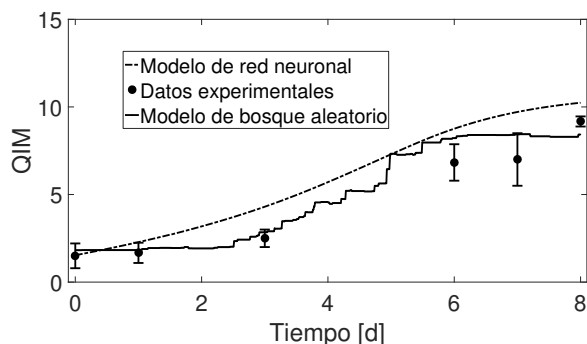


Figura 2: Comparación, en términos del QIM, entre los datos experimentales y los modelos de red neuronal y bosque aleatorio.

4. Conclusiones y trabajo futuro

El objetivo de este trabajo era desarrollar una metodología para predecir la evolución de la calidad de filetes de Merluza (*Merluccius merluccius*) envasados en atmósferas modificadas durante el almacenamiento en condiciones de refrigeración. Para ello, se han desarrollado diferentes modelos matemáticos para predecir la evolución de tres de los principales indicadores de calidad en pescado fresco: el QIM, la concentración de bases volátiles y la concentración de bacterias (*Pseudomonas spp.* y *Shewanella spp.*). Dado que el QIM es un indicador subjetivo, se ha optado por utilizar modelos de aprendizaje automático para la descripción de su evolución. En particular, se ha desarrollado un modelo de red neuronal y otro de bosque aleatorio, resultando el segundo el más adecuado para representar la evolución del QIM. En lo que se refiere a los otros tres indicadores, se han desarrollado distintos modelos basados en los mecanismos del proceso y se han calibrado utilizando los datos de dos experimentos. Un tercer experimento se dejó para la validación de los modelos. En el caso de crecimiento bacteriano, se ha utilizado un modelo logístico capaz de representar el comportamiento experimental, incluso en el experimento de validación. Para las bases volátiles se han considerado y comparado dos modelos: exponencial y ley de potencias, resultando el segundo modelo cuantitativamente mejor.

A pesar de que los resultados obtenidos en este trabajo son satisfactorios, nos gustaría mejorarlos realizando en el futuro las siguientes tareas: (i) llevar a cabo un diseño óptimo de experimentos que nos permita mejorar la calidad de la información obtenida de los datos experimentales. (ii) Considerar otros modelos secundarios tanto para el crecimiento bacteriano como para el contenido en bases volátiles. (iii) Modelización del crecimiento de la bacteria *Photobacterium phosphoreum* ya que, aunque tanto *Pseudomonas spp.* como *Shewanella spp.* son dos de las bacterias más utilizadas en la degradación de la calidad de pescado, en condiciones de atmósferas modificadas podría ser más importante la primera. (iv) Modelización de la formación y degradación de nucleótidos dado que también son ampliamente utilizados como indicadores de calidad en pescado.

5. Agradecimientos

Este trabajo ha sido realizado en el marco de los proyectos: Pescando 4.0 (ref. 202178002), financiado por fondos Next-

GenerationEU -Plan de Recuperación, Transformación y Resiliencia (PRTR)- gestionado por el Ministerio de Agricultura, Pesca y Alimentación mediante el Real Decreto 685/2021; y del programa de Ciencias Marinas (ThinkInAzul) apoyado por el Ministerio de Ciencia e Innovación y Xunta de Galicia con financiación de la Unión Europea NextGenerationEU (PRTR-C17.I1) y el Fondo Europeo Marítimo y de Pesca.

Referencias

- Akaike, H., 1974. A new look at the statistical model identification. IEEE transactions on automatic control 19, 716–723.
- Antonacopoulos, N., 1960. Verbesserte apparatur zur quantitativen destillation wasserdampflichiger stoffe. Zeitschrift für Lebensmittel-Untersuchung und Forschung 113, 113–116.
- Balsa-Canto, E., Henriques, D., Gábor, A., Banga, J.R., 2016. Amigo2, a toolbox for dynamic modeling, optimization and control in systems biology. Bioinformatics 32, 3357–3359.
- Baranyi, J., Roberts, T.A., 1994. A dynamic approach to predicting bacterial growth in food. International journal of food microbiology 23, 277–294.
- Breiman, L., 2001. Random forests. Machine Learning , 5–32.
- Chollet, F., et al., 2015. Keras. URL: <https://github.com/fchollet/keras>.
- Corradini, M.G., 2018. Shelf life of food products: from open labeling to real-time measurements. Annual review of food science and technology 9, 251–269.
- Egea, J.A., Balsa-Canto, E., García, M.S.G., Banga, J.R., 2009. Dynamic optimization of nonlinear processes with an enhanced scatter search method. Industrial & Engineering Chemistry Research 48, 4388–4401.
- Escribano, E., Massa, A., Yeannes, M.I., 2015. Desarrollo del método del índice de calidad (qim) para evaluar la frescura de la merluza común (merluccius hubbsi) almacenada en hielo .
- García, M.R., Cabo, M.L., Herrera, J.R., Ramilo-Fernández, G., Alonso, A.A., Balsa-Canto, E., 2017. Smart sensor to predict retail fresh fish quality under ice storage. Journal of Food Engineering 197, 87–97.
- García, M.R., Ferez-Rubio, J.A., Vilas, C., 2022. Assessment and prediction of fish freshness using mathematical modelling: A review. Foods 11, 2312.
- García, M.R., Vilas, C., Herrera, J.R., Bernárdez, M., Balsa-Canto, E., Alonso, A.A., 2015. Quality and shelf-life prediction for retail fresh hake (merluccius merluccius). International journal of food microbiology 208, 65–74.
- Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. volume 2. Springer.
- Heaton, J., 2008. Introduction to neural networks with Java. Heaton Research, Inc.
- Lücke, F., Geidel, W., 1935. Bestimmung des flüchtigen basischen stickstoffs in fischen als maßstab für ihren frischezustand. Zeitschrift für Untersuchung der Lebensmittel 70, 441–458.
- McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics 5, 115–133.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. Journal of machine learning research 12, 2825–2830.
- Raftery, A.E., 1995. Bayesian model selection in social research. Sociological methodology , 111–163.
- Ratkowsky, D.A., Olley, J., McMeekin, T., Ball, A., 1982. Relationship between temperature and growth rate of bacterial cultures. Journal of bacteriology 149, 1–5.
- Sharma, S., Sharma, S., Athaiya, A., 2017. Activation functions in neural networks. Towards Data Sci 6, 310–316.
- Tavares, J., Martins, A., Fidalgo, L.G., Lima, V., Amaral, R.A., Pinto, C.A., Silva, A.M., Saraiva, J.A., 2021. Fresh fish degradation and advances in preservation using physical emerging technologies. Foods 10, 780.
- Tsironi, T., Dermesonlouoglou, E., Giannakourou, M., Taoukis, P., 2009. Shelf life modelling of frozen shrimp at variable temperature conditions. LWT-Food Science and Technology 42, 664–671.