Diseño y formulación de mezclas óptimas de biodiésel mediante aprendizaje automático y algoritmos evolutivos

Guillermo Díez-Valbuena[1], Alejandro García-Tuero[1], Eduardo Rodríguez[1], Antolin Hernández Battez[1], and Jorge Díez[2]

[1] Departamento de Construcción e Ingeniería de Fabricación, Universidad de Oviedo
{gdvalbuena,garciatalejandro,eduardo,aehernandez}@uniovi.es
[2] Centro de Inteligencia Artificial, Universidad de Oviedo
jdiez@uniovi.es

Resumen El diseño de un biodiésel óptimo exige equilibrar propiedades fisicoquímicas, como el Número de Cetano (CN), el Índice de Yodo (IV) y el Punto de Obstrucción del Filtro en Frío (CFPP). Aunque los enfoques computacionales recientes han abordado este problema mediante algoritmos de optimización, a menudo proponen perfiles de *ésteres metílicos de ácidos grasos* (FAMEs) no existentes en la naturaleza y carecen de aplicabilidad práctica para los productores. Para superar estas limitaciones, este artículo propone un marco metodológico estructurado en tres fases. Primero, se generan modelos predictivos de aprendizaje automático para estimar las propiedades del biodiésel a partir de su composición. Segundo, se emplea un algoritmo genético multiobjetivo (NSGA-II), con estrictas restricciones de dominio, que genera un frente de Pareto de soluciones teóricas óptimas compuesto por perfiles de FAMEs posibles en la naturaleza. Finalmente, se implementa un algoritmo de optimización de variables mixtas que, a partir del individuo objetivo seleccionado por un experto, calcula las proporciones de mezcla necesarias para replicar dicho perfil a partir de otros tipos de biodiésel conocidos. Los resultados demuestran que es posible alcanzar un biodiésel con un equilibrio sobresaliente (CN alto, IV y CFPP bajos) formulando una mezcla de tan solo dos materias primas reales (aceite de *Litsea* y *Chlorella*).

Keywords: Aprendizaje supervisado · Optimización multiobjetivo · Biodiésel · Perfil de FAMEs · Human-in-the-loop

1. Introducción

Gracias a su viabilidad como combustible de sustitución en los motores diésel y a su capacidad para operar bajo un ciclo de carbono neutro, el biodiésel se ha posicionado como una alternativa renovable clave para la descarbonización del sector del transporte. El biodiésel, compuesto por ésteres metílicos de ácidos grasos (FAMEs, por sus siglas en inglés), presenta propiedades fisicoquímicas que dependen directamente de la naturaleza de la materia prima utilizada en su

producción. Dado que el perfil de FAMEs varía drásticamente entre diferentes aceites vegetales y grasas animales, la calidad y el rendimiento del biocombustible resultante también fluctúan de manera significativa [9,18].

El diseño de un biodiésel óptimo se enfrenta a un desafío fundamental: el compromiso o *trade-off* inherente entre sus propiedades críticas. Por un lado, una alta proporción de ácidos grasos saturados eleva el Número de Cetano (CN, Cetane Number), mejorando la calidad de la ignición, y reduce el Índice de Yodo (IV, Iodine Value), aumentando la estabilidad a la oxidación [9,11]. Sin embargo, estas mismas moléculas saturadas cristalizan a temperaturas más altas, lo que empeora drásticamente las propiedades de flujo en frío, evidenciado por un aumento en el Punto de Obstrucción del Filtro en Frío (CFPP, Cold Filter Plugging Point) [9]. Por lo tanto, maximizar el CN y minimizar simultáneamente el IV y el CFPP requiere un equilibrio composicional complejo que ninguna materia prima pura puede ofrecer por sí sola [17].

Para abordar este problema, la investigación reciente se ha apoyado en herramientas computacionales basadas en Inteligencia Artificial. Diversos estudios han demostrado que los algoritmos de Aprendizaje Automático son eficaces para predecir las propiedades del biodiésel a partir de su perfil de FAMEs [7,8,10], reduciendo la dependencia de costosos ensayos empíricos. Asimismo, se han aplicado técnicas de optimización multiobjetivo para identificar perfiles teóricos ideales [2,15]. No obstante, la literatura actual presenta dos lagunas críticas. En primer lugar, los algoritmos evolutivos frecuentemente exploran el espacio de búsqueda sin restricciones composicionales estrictas, convergiendo hacia frentes de Pareto compuestos por perfiles de FAMEs no factibles, es decir, que no se corresponden con proporciones presentes en la naturaleza. En segundo lugar, los estudios suelen detenerse tras la identificación del individuo teórico óptimo, sin proporcionar una metodología que permita a la industria materializar dicho perfil mediante la mezcla (*blending*) de materias primas físicas disponibles.

Para superar estas limitaciones, este trabajo propone un marco metodológico integral para el diseño y la formulación práctica de mezclas de biodiésel. Las principales contribuciones de esta investigación son las siguientes:

1. **Modelos predictivos para las propiedades de los biodiésel**. Entrenamiento y validación de tres modelos de aprendizaje automático capaces de predecir el CN, el IV y el CFPP a partir del perfil de FAMEs, demostrando un alto rendimiento predictivo y capacidad de generalización mediante validación cruzada.
2. **Optimización multiobjetivo con restricciones de origen natural**. Implementación de un algoritmo genético para maximizar el CN y simultáneamente minimizar el IV y el CFPP. El algoritmo incorpora restricciones de dominio específicas para garantizar que los individuos generados en el frente de Pareto correspondan a perfiles de FAMEs posibles en la naturaleza.
3. **Formulación automatizada de mezclas**. Desarrollo de un algoritmo que calcula las proporciones necesarias para crear una mezcla (a partir de biodiésel reales) con el fin de replicar un perfil de FAMEs objetivo. Este perfil es seleccionado previamente en el frente de Pareto por un experto.

En la Sección 2 se analiza la literatura relacionada con esta investigación, sentando las bases para describir la metodología propuesta en la Sección 3. Seguidamente, los resultados obtenidos se discuten en la Sección 4, finalizando el artículo con las conclusiones más relevantes y las líneas de trabajo futuro.

2. Trabajo relacionado

La predicción de las propiedades fisicoquímicas del biodiésel a partir de su perfil de FAMEs ha sido ampliamente estudiada para minimizar la dependencia de costosos ensayos experimentales. Mientras que los primeros enfoques se basaban en modelos empíricos y correlaciones lineales [12,13,18,9], en los últimos años la literatura ha adoptado técnicas de Aprendizaje Automático. Se han empleado con éxito diversos algoritmos predictivos como Redes Neuronales Artificiales, Máquinas de Vectores de Soporte o Bosques Aleatorios, entre otros, para estimar propiedades clave con alta precisión [1,6,7,8,10]. Estos modelos permiten explorar el espacio de diseño de biocombustibles de manera computacional, eludiendo la síntesis física previa que encarecería y ralentizaría enormemente la fase de exploración.

De forma complementaria, algunos estudios muestran que los modelos predictivos pueden integrarse con algoritmos evolutivos para la optimización composicional [2,15], permitiendo orientar el diseño hacia perfiles de FAME más robustos frente a normativas y restricciones operativas. Sin embargo, Menon et al. [15] emplearon únicamente 13 biodiésel de partida (utilizados para entrenar la red neuronal que actúa como función de *fitness*), mientras que Bukkarapu et al. [2], aunque trabajaron con 103 muestras, las generaron a partir de la mezcla de únicamente 8 biodiésel base. Esta escasa variabilidad composicional puede inducir a un sesgo en el espacio de representación o a una fuga de datos (*data leakage*) durante la validación del modelo, limitando su capacidad de generalización.

Si bien las técnicas heurísticas han mejorado significativamente el diseño de biocombustibles, los enfoques existentes siguen condicionados por dos limitaciones metodológicas principales, introducidas previamente en la Introducción:

- Los algoritmos de optimización a menudo exploran un espacio de búsqueda sin incorporar restricciones composicionales. Esto resulta en la generación de perfiles de FAME puramente teóricos que resultan imposibles de obtener a partir de materias primas naturales [9].
- Los estudios suelen concluir al identificar un biodiésel teórico óptimo tras la optimización heurística, dejando un vacío metodológico sobre cómo trasladar este perfil ideal a la práctica industrial. Aunque la mezcla de biocombustibles (*biodiesel blending*) es una práctica común para mejorar las propiedades de flujo y combustión [16,19], rara vez se integra como la etapa resolutiva tras la optimización. Falta en la literatura un marco unificado que, partiendo de modelos predictivos, no solo garantice la viabilidad natural del perfil de FAMEs durante la evolución genética, sino que también aplique ingeniería inversa para determinar las proporciones exactas de mezcla a partir de inventarios de materias primas físicas disponibles.

3. Método propuesto

La metodología propuesta en este estudio se divide en tres fases secuenciales: i) el modelado predictivo de las propiedades del biodiésel basado en su perfil de FAMEs, ii) la optimización multiobjetivo para identificar perfiles teóricos ideales, y iii) la formulación de mezclas para alcanzar dichos perfiles.

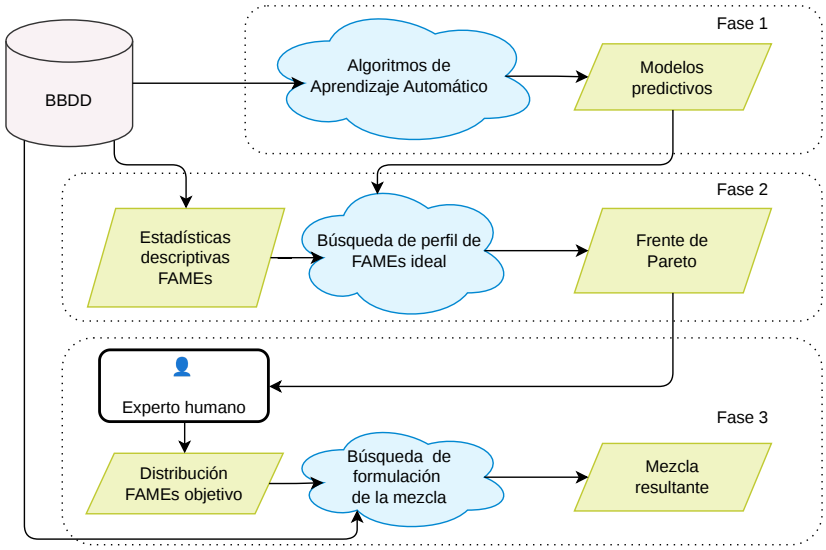


Figura 1: Diagrama de la metodología propuesta.

En la Figura 1 se muestra un diagrama de la metodología propuesta. El proceso toma como punto de partida la base de datos de biodiésel disponible, de la cual se extraen los conjuntos de datos necesarios para entrenar un modelo predictivo independiente para cada una de las propiedades de interés. Posteriormente, estos modelos actúan como *funciones de fitness* dentro del algoritmo genético multiobjetivo, el cual aborda la búsqueda de un biodiésel con las propiedades optimizadas. Este algoritmo también utiliza como entrada las estadísticas descriptivas de los FAMEs presentes en la base de datos, garantizando así la generación de individuos (distribuciones de FAMEs) que sean factibles en la naturaleza. El resultado de esta fase es la obtención de un frente de Pareto. Llegados a este punto, un experto humano analiza el frente y selecciona la distribución de FAMEs óptima en función de los requerimientos específicos [14]. Esta decisión humana da pie a la última fase del método, donde un algoritmo genético monobjetivo calcula una mezcla a partir de varios biodiésel de la base

de datos, logrando una composición lo más cercana posible a la seleccionada por el experto y replicando, por consiguiente, las propiedades deseadas.

En las siguientes secciones, además de describir la base de datos, se detallan estas tres fases en profundidad.

3.1. Base de datos de biodiésel

Se cuenta con una base de datos compuesta por 544 muestras de biodiésel, donde cada biodiésel se describe mediante su perfil de 20 FAMEs. En la Figura 2 se muestra la distribución y variabilidad de las concentraciones de los FAMEs presentes en el estudio. El análisis revela una marcada asimetría en la composición, donde ácidos grasos como el C18:1 y el C18:2 presentan los valores medianos más altos y los rangos intercuartílicos más amplios, indicando su predominancia en los biocombustibles. En contraste, componentes como el C15:0 y el C17:1 se detectan únicamente a nivel de trazas. Adicionalmente, la figura evidencia una gran cantidad de valores atípicos en compuestos como el C12:0, C16:0 y C18:3, lo que sugiere una considerable heterogeneidad en ciertas muestras específicas del conjunto de datos.

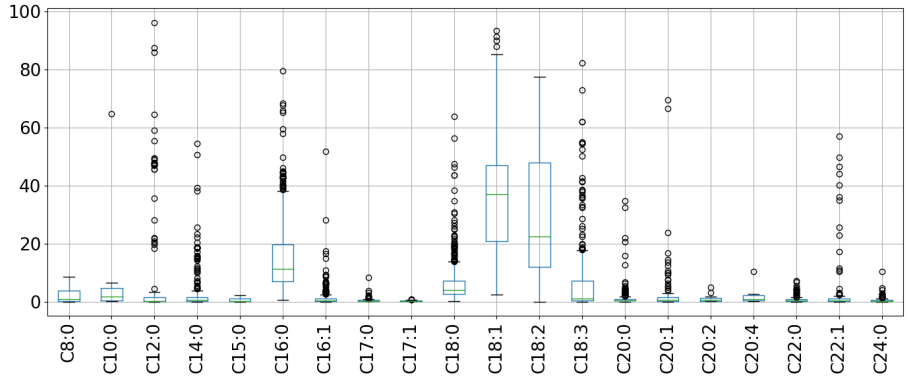


Figura 2: Diagrama de cajas de los 20 FAMEs presentes en el estudio.

Para cada muestra se dispone, además, de los valores empíricos correspondientes a las propiedades fisicoquímicas de interés. Es importante destacar que, en la mayoría de los casos, cada muestra incluye el valor de una única propiedad, siendo infrecuente disponer de las tres medidas simultáneamente para un mismo individuo. En total, el conjunto de datos cuenta con registros del CN para 389 muestras, del IV para 265 y del CFPP para 219. Como se observa en los histogramas de la Figura 3, el rango de distribución de las tres variables es bastante amplio, reflejando adecuadamente la gran diversidad tipológica de los tipos de biodiésel analizados.

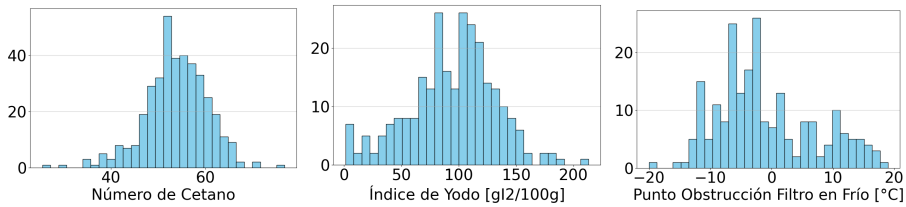


Figura 3: Histogramas de distribución de las diferentes propiedades en los biodiésel de la base de datos.

3.2. Modelado Predictivo de Propiedades

A partir de la base de datos descrita anteriormente, se generan tres conjuntos de datos, uno para cada propiedad. El perfil de FAMEs de cada biodiésel se representa como un vector de composición $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$, donde x_i representa el porcentaje en masa del i-ésimo FAME y donde n representa el número de FAMEs considerados en el estudio.

A partir de estos conjuntos de datos, se entrenan tres modelos predictivos independientes para mapear el perfil de FAMEs hacia las tres propiedades fisicoquímicas de interés: Número de Cetano (CN), Índice de Yodo (IV) y Punto de Obstrucción del Filtro en Frío (CFPP). Estos modelos actuarán posteriormente como funciones de evaluación en la etapa de optimización evolutiva y se denotarán formalmente como: $f_{CN}(\boldsymbol{x})$, $f_{IV}(\boldsymbol{x})$ y $f_{CFPP}(\boldsymbol{x})$.

Para garantizar la robustez y la capacidad de generalización de los modelos, el rendimiento se calcula mediante una validación cruzada de 10 folds, donde cada entrenamiento individual incorpora un ajuste de hiperparámetros utilizando a su vez otra validación cruzada de 5 folds.

3.3. Módulo para encontrar distribuciones ideales de FAMEs

Este módulo tiene como objetivo identificar distribuciones de FAMEs que den lugar a unos determinados valores óptimos en las propiedades estudiadas. El algoritmo de optimización genera individuos (perfiles composicionales) que se evalúan mediante los modelos predictivos descritos en la sección anterior. De esta manera, es posible estimar el rendimiento esperado de cada individuo para cada una de las propiedades, eludiendo la síntesis física y la caracterización fisicoquímica del biodiésel, lo que supondría un elevado coste temporal y económico.

Como resultado, este algoritmo genera un frente de Pareto (compuesto por el conjunto de soluciones no dominadas). En esta etapa, un experto humano (*Decision Maker*) analiza el frente y selecciona la distribución de FAMEs que mejor se ajusta a los requerimientos específicos del biodiésel, un procedimiento formalmente conocido como *toma de decisiones a posteriori* [14].

Formulación de la Optimización Multiobjetivo Dado que los objetivos de diseño (maximizar el CN y minimizar el IV y el CFPP) presentan un conflicto

inherente, la búsqueda del biodiésel óptimo se formula como un problema de optimización multiobjetivo. El propósito del motor de búsqueda, basado en algoritmos genéticos, es encontrar un frente de Pareto de vectores $\boldsymbol{x}$ que satisfagan simultáneamente las siguientes funciones objetivo:

$$(\text{máx}\, f_{CN}(\boldsymbol{x}), \text{mín}\, f_{IV}(\boldsymbol{x}), \text{mín}\, f_{CFPP}(\boldsymbol{x})).$$

En la práctica, para adaptar el problema a los algoritmos de minimización estándar, la primera función se reformula como $\text{mín} - f_{CN}(\boldsymbol{x})$.

Como algoritmo de optimización se emplea el NSGA-II [3] con un tamaño de población de 300 individuos. La inicialización de la población se realiza mediante un muestreo gaussiano acotado, adaptado a los límites y a la distribución empírica de los FAMEs (ver Figura 2). Para la generación de descendencia, se utiliza el operador de Cruce Binario Simulado (SBX) [5] con una probabilidad $p_c = 0.9$ y un índice de distribución $\eta_c = 15$. Asimismo, se aplica el operador de Mutación Polinómica (PM) [4] con una probabilidad $p_m = 1/n$ y un índice de distribución $\eta_m = 20$. Finalmente, se integra un operador de reparación posterior a la mutación y el cruce para garantizar la viabilidad física de las soluciones generadas.

Representación de los individuos y restricciones Los individuos se representan mediante vectores de n números reales, donde cada componente corresponde a la fracción másica de un FAME.

Una de las contribuciones críticas del procedimiento presentado en este trabajo es la imposición de restricciones estrictas en el espacio de búsqueda del algoritmo evolutivo, evitando así que los operadores de cruce y mutación generen perfiles de FAMEs imposibles en la naturaleza.

Restricción de normalización: La suma de todas las fracciones de FAMEs debe representar la totalidad del perfil de ácidos grasos. Idealmente, esto implicaría que $\sum_{i=1}^{n} x_i = 100\,\%$. Sin embargo, dado que un biodiésel real puede contener trazas de ácidos grasos no contemplados en este estudio, se admite cierta tolerancia computacional, considerando factibles aquellos individuos cuya suma total se sitúe en el intervalo $[90, 100]$.

Restricciones de frontera (límites naturales): Cada FAME individual x_i debe acotarse entre un límite inferior L_i y un límite superior U_i determinados por las distribuciones empíricas observadas en la naturaleza (es decir, los límites teóricos documentados para aceites vegetales y grasas animales, tal que $L_i \leq x_i \leq U_i$).

3.4. Módulo para el cálculo de la mezcla

Tras la obtención del frente de Pareto, un experto selecciona el individuo deseado de acuerdo con las propiedades buscadas, el cual queda definido por su perfil de FAMEs y se denota como $\boldsymbol{b}_{obj}$ (biodiésel objetivo). Dado que este perfil ideal teórico rara vez corresponde a una única materia prima comercial pura,

se debe calcular la formulación de mezcla (*blending*) necesaria para replicarlo a partir de los perfiles de biodiésel físicos disponibles en la base de datos.

Sea $B \in \mathbb{R}^{n\times m}$ la matriz de biodiésel, donde cada una de las m columnas representa el perfil composicional de un biodiésel real disponible. Debido a restricciones logísticas y operativas de la industria, la formulación empírica se restringe a un máximo de K componentes (donde un menor valor de K facilita la viabilidad práctica de la mezcla). Este planteamiento transforma el caso en un problema de optimización con variables mixtas, ya que exige tomar simultáneamente la decisión discreta de qué materias primas seleccionar y la decisión continua de en qué proporciones combinarlas. Para formular matemáticamente este proceso, se definen:

- Un vector de variables discretas $\boldsymbol{i} = [i_1, i_2, \ldots, i_K] \in \{1, \ldots, m\}^K$, que representa los índices de los biodiésel seleccionados de la matriz B para participar en la mezcla.
- Un vector de variables continuas $\boldsymbol{v} = [v_1, v_2, \ldots, v_K] \in [0, 1]^K$, donde cada valor representa la proporción bruta (sin normalizar) de cada componente seleccionado. A partir de este vector, la fracción másica real w_j de cada biodiésel se obtiene aplicando una normalización para garantizar la conservación de la masa:
$$w_j = \frac{v_j}{\sum_{l=1}^{K} v_l} \quad \forall j \in \{1, \ldots, K\},$$
lo que da lugar al vector de pesos normalizado $\boldsymbol{w} = [w_1, w_2, \ldots, w_K] \in [0, 1]^K$, cumpliendo estrictamente que $\sum_{j=1}^{K} w_j = 1$.

De esta manera, el perfil de FAMEs resultante de la mezcla de los K componentes se obtiene mediante la combinación lineal:

$$\boldsymbol{b}_{mezcla} = \sum_{j=1}^{K} w_j \boldsymbol{b}_{i_j}$$

donde $\boldsymbol{b}_{i_j}$ es la columna de la matriz B referida al biodiésel con índice i_j.

Por consiguiente, el objetivo de esta fase de optimización es minimizar el Error Cuadrático Medio (MSE) entre el perfil de FAMEs de la mezcla propuesta y el perfil objetivo teórico $\boldsymbol{b}_{obj}$, formulado como:

$$\min_{\boldsymbol{i},\boldsymbol{w}} \frac{1}{n} \|\boldsymbol{b}_{mezcla} - \boldsymbol{b}_{obj}\|^2.$$

Esta optimización está sujeta a la restricción combinatoria de no seleccionar redundantemente una misma materia prima dentro de una misma solución, exigiendo que $i_p \neq i_q \quad \forall p, q \in \{1, \ldots, K\}$ con $p \neq q$.

Para abordar esta búsqueda, se emplea un Algoritmo Genético de Variables Mixtas. Esta heurística permite la coexistencia y evolución simultánea de cromosomas de tipo discreto y continuo mediante los operadores de Cruce Binario Simulado ($p_c = 0.9$ y $\eta_c = 15$) y Mutación Polinómica ($p_m = 1/(2K)$). Los índices de distribución de la mutación se ajustan de forma asimétrica para favorecer la microexploración en el ajuste continuo de los pesos ($\eta_m = 30$) y los macro-saltos en el cambio discreto de las materias primas ($\eta_m = 10$).

Tabla 1: Resultados obtenidos por los modelos predictivos para cada una de las tres propiedades estudiadas.

	CN		IV		CFPP	
Sistema	MSE	R^2	MSE	R^2	MSE	R^2
Predictor media	43.8	-0.007	1445.9	-0.021	65.4	-0.011
Decision Tree	30.6	0.298	645.1	0.544	36.6	0.434
Random Forest	25.7	0.409	462.8	0.673	**29.4**	**0.545**
XGBoost	26.5	0.391	478.5	0.662	33.8	0.477
Regresión Ridge	23.1	0.469	484.2	0.658	39.7	0.387
SVR(lineal)	**21.8**	**0.499**	**387.4**	**0.726**	40.2	0.378
SVR(polinómico)	26.2	0.399	392.2	0.723	47.9	0.259
SVR(RBF)	21.9	0.497	388.1	**0.726**	48.4	0.252

4. Resultados

La Tabla 1 presenta los resultados alcanzados mediante validación cruzada, tal como se detalla en la Sección 3.2, para diferentes algoritmos de aprendizaje automático ampliamente utilizados. El rendimiento predictivo se evaluó utilizando el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R^2). Se observa que las Máquinas de Vectores de Soporte (SVR) obtienen el mejor rendimiento en la predicción del Número de Cetano (CN) y del Índice de Yodo (IV). Por su parte, para la predicción del Punto de Obstrucción del Filtro en Frío (CFPP), el modelo basado en Random Forest logra los mejores resultados. Los modelos que demostraron un rendimiento superior en esta validación se seleccionaron como funciones de evaluación para la etapa de optimización.

Al aplicar el algoritmo de optimización multiobjetivo descrito en la Sección 3.3 para buscar un biodiésel con un valor alto de CN y valores bajos de IV y CFPP, se generó el frente de Pareto ilustrado en la Figura 4. Se puede apreciar que las soluciones encontradas por el algoritmo (en comparación con la distribución empírica mostrada previamente en la Figura 3) logran situarse exitosamente en un rango alto de CN, un rango muy bajo de IV y un rango intermedio de CFPP. Ante este abanico de compromisos, un experto en la materia actuó como *Decision Maker* y seleccionó, siguiendo requisitos normativos, un biodiésel objetivo ($\boldsymbol{b}_{obj}$) con un CN e IV intermedios y un CFPP bajo (61.8, 40.7 $gI_2/100g$ y -8.7 °C, respectivamente), cuya composición teórica se detalla en la Tabla 2.

La fase final del procedimiento consistió en el cálculo operativo de la mezcla que permitiera materializar este perfil $\boldsymbol{b}_{obj}$. Para ello, se ejecutó el algoritmo de optimización de variables mixtas detallado en la Sección 3.4. Fijando $K = 2$ (es decir, limitando la mezcla a únicamente dos biodiésel), se logró reproducir una distribución de FAMEs altamente concordante con la del biodiésel objetivo, tal y como se observa en la Tabla 2. Las propiedades predichas para esta nueva formulación son 61.3 para el CN, 34.7 $gI_2/100g$ para el IV y -6.3 °C para el CFPP, mostrando una desviación mínima respecto al óptimo teórico.

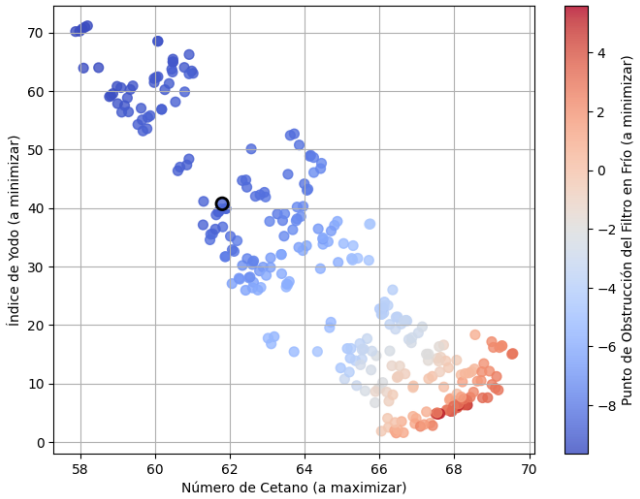


Figura 4: Representación gráfica de los perfiles de FAME en el frente de Pareto. Aparece resaltado con un borde el $\boldsymbol{b}_{obj}$ seleccionado por el experto en la materia.

Tabla 2: Composición de FAMEs en el biodiésel objetivo y en la mezcla obtenida

	C12:0	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1	C22:1	C24:0
$\boldsymbol{b}_{obj}$	69.0	0	4.8	0.3	2.1	0	6.9	8.6	2.4	1.5	4.0	0.5
$\boldsymbol{b}_{mezcla}$	69.9	0.5	9.0	1.6	1.7	2.7	6.2	6.9	0.5	0	0	0

Nota: Los FAMEs C8:0, C10:0, C15:0, C17:0, C17:1, C20:2, C20:4 y C22:0 tienen presencia 0 en la composición de ambos biodiésel.

Esta mezcla óptima es industrialmente realizable combinando únicamente dos tipos de biodiésel: i) un 72.6 % de biodiésel derivado de *Litsea* (un arbusto propio de zonas tropicales) y ii) un 27.4 % de biodiésel procedente de *Chlorella* (una microalga de agua dulce).

5. Conclusiones y trabajo futuro

En este trabajo se ha presentado una metodología integral que cubre el ciclo completo del diseño computacional de biodiésel, salvando la brecha existente entre la optimización teórica y la formulación práctica a escala industrial.

Se ha demostrado que el uso de técnicas de Aprendizaje Automático, específicamente Máquinas de Vectores de Soporte (SVR) para el CN y el IV, y Random Forest para el CFPP, permite predecir con alta precisión las propiedades fisicoquímicas analizadas basándose únicamente en el perfil de FAMEs. Esto da lugar a un modelado predictivo robusto que reduce drásticamente la dependencia de iteraciones experimentales costosas.

La integración de restricciones de frontera empíricas en el algoritmo evolutivo multiobjetivo (NSGA-II), junto con una inicialización adaptada de la población, garantiza que el espacio de búsqueda explore exclusivamente perfiles de FAMEs posibles en la naturaleza. Esto corrige un sesgo metodológico común en la literatura previa, dando lugar a una optimización realista.

Asimismo, el desarrollo del algoritmo de optimización de variables mixtas del último módulo demuestra que los perfiles teóricos óptimos del frente de Pareto pueden materializarse mediante la combinación de proporciones precisas de materias primas reales. El caso de estudio evidenció que, combinando únicamente dos biodiésel (derivados de *Litsea* y *Chlorella*), es posible replicar con gran exactitud el perfil de FAMEs objetivo y, por consiguiente, alcanzar las propiedades deseadas. Cabe señalar que la validación de la mezcla propuesta se ha realizado exclusivamente mediante los modelos predictivos, sin contrastar experimentalmente las propiedades estimadas. Una validación empírica futura permitiría confirmar la aplicabilidad industrial real de la formulación obtenida.

Como líneas de trabajo futuro, el módulo de mezcla podría refinarse incorporando funciones de penalización basadas en el coste económico, la huella de carbono o la disponibilidad logística de las materias primas, acercando aún más esta herramienta a las necesidades operativas del sector de los biocombustibles. Asimismo, sería interesante explorar el uso de NSGA-III con perfiles de referencia (reference points) definidos por expertos, orientando la búsqueda hacia las regiones del espacio de soluciones más alineadas con sus intereses.

Agradecimientos Este trabajo ha sido financiado mediante el proyecto PID2022-136656NB-I00 del Ministerio de Ciencia e Innovación de España. La colaboración de Guillermo Díez-Valbuena ha sido financiada por el Gobierno del Principado de Asturias mediante el programa de becas predoctorales *Severo Ochoa* (ref. BP22-153). La Agencia de Ciencia, Competitividad Empresarial e Innovación Asturiana (SEKUENS) también ha colaborado en la financiación de este trabajo a través de los proyectos IDE/2024/000686 e IDE/2024/000696.

Conflictos de intereses Los autores declaran no tener conflictos de intereses.

Referencias

1. Alviso, D., Artana, G., Duriez, T.: Prediction of biodiesel physico-chemica properties from its fatty acid composition using genetic programming **264**, 116844. https://doi.org/10.1016/j.fuel.2019.116844
2. Bukkarapu, K.R., Krishnasamy, A.: Investigations on the applicability of machine learning algorithms to optimize biodiesel composition for improved engine fuel properties. International Journal of Engine Research **25**(7), 1299–1314 (2024). https://doi.org/10.1177/14680874241227540
3. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Transactions on Evolutionary Computation **6**(2), 182–197 (2002). https://doi.org/10.1109/4235.996017
4. Deb, K.: Multi-objective optimization using evolutionary algorithms. John Wiley & Sons, New York, NY, USA (2001)

5. Deb, K., Agrawal, R.B., et al.: Simulated binary crossover for continuous search space. Complex systems **9**(2), 115–148 (1995)
6. Díez-Valbuena, G., García Tuero, A., Díez, J., Rodríguez, E., Hernández Battez, A.: Application of machine learning techniques to predict biodiesel iodine value. Energy **292**, 130638 (2024). https://doi.org/10.1016/j.energy.2024.130638
7. Díez-Valbuena, G., García Tuero, A., Díez, J., Rodríguez, E., Hernández Battez, A.: Prediction of the cold flow properties of biodiesel using the fame distribution and machine learning techniques. Journal of Molecular Liquids **400**, 124555 (2024). https://doi.org/10.1016/j.molliq.2024.124555
8. Díez-Valbuena, G., García-Tuero, A., Rodríguez, E., Hernández Battez, A., Díez, J.: Modeling biodiesel properties by preference learning: Case study of cetane number. Engineering Applications of Artificial Intelligence **144**, 110107 (2025). https://doi.org/10.1016/j.engappai.2025.110107
9. Hoekman, S.K., Broch, A., Robbins, C., Ceniceros, E., Natarajan, M.: Review of biodiesel composition, properties, and specifications. Renewable and Sustainable Energy Reviews **16**(1), 143–169 (2012). https://doi.org/10.1016/j.rser.2011.07.143
10. Huang, Y., Li, F., Bao, G., Xiao, Q., Wang, H.: Modeling the effects of biodiesel chemical composition on iodine value using novel machine learning algorithm. Fuel **316**, 123348 (May 2022). https://doi.org/10.1016/j.fuel.2022.123348
11. Knothe, G.: Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. Fuel Processing Technology **86**(10), 1059–1070 (2005). https://doi.org/10.1016/j.fuproc.2004.11.002, biodiesel Processing and Production
12. Knothe, G.: A comprehensive evaluation of the cetane numbers of fatty acid methyl esters. Fuel **119**, 6–13 (2014). https://doi.org/10.1016/j.fuel.2013.11.020
13. Krisnangkura, K., Yimsuwan, T., Pairintra, R.: An empirical approach in predicting biodiesel viscosity at various temperatures. Fuel **85**(1), 107–113 (2006). https://doi.org/10.1016/j.fuel.2005.05.010
14. Meignan, D., Knust, S., Frayret, J.M., Pesant, G., Gaud, N.: A review and taxonomy of interactive optimization methods in operations research. ACM Trans. Interact. Intell. Syst. **5**(3) (Sep 2015). https://doi.org/10.1145/2808234
15. Menon, P.R., Krishnasamy, A.: A composition-based model to predict and optimize biodiesel-fuelled engine characteristics using artificial neural networks and genetic algorithms. Energy & Fuels **32**(11), 11607–11618 (2018). https://doi.org/10.1021/acs.energyfuels.8b02846
16. Monirul, I.M., Masjuki, H.H., Kalam, M.A., Zulkifli, N.W.M., Rashedul, H.K., Rashed, M.M., Imdadul, H.K., Mosarof, M.H.: A comprehensive review on biodiesel cold flow properties and oxidation stability along with their improvement processes. RSC Advances **5**, 86631–86655 (2015). https://doi.org/10.1039/C5RA09555G
17. Moser, B.R.: Influence of blending canola, palm, soybean, and sunflower oil methyl esters on fuel properties of biodiesel. Energy & Fuels **22**(6), 4301–4306 (2008). https://doi.org/10.1021/ef800588x
18. Ramos, M.J., Fernández, C.M., Casas, A., Rodríguez, L., Ángel Pérez: Influence of fatty acid composition of raw materials on biodiesel properties. Bioresource Technology **100**(1), 261–268 (2009). https://doi.org/10.1016/j.biortech.2008.06.039
19. Vasileiadis, V., Kyriklidis, C., Karayannis, V., Tsanaktsidis, C.: Application of evolutionary computation to the optimization of biodiesel mixtures using a nature-inspired adaptive genetic algorithm. Algorithms **17**(5) (2024). https://doi.org/10.3390/a17050181