

Optimización hiperheurística guiada por datos experimentales

Juan I. Ardenghi¹[0000-0003-3248-5965], Constanza Genovese^{1,3}[0000-0002-1138-5626], Paola P. Oteiza^{1,2}[0000-0001-7265-225X], y Nélica B. Brignole^{1,3}[0000-0002-4795-2872]

¹ Laboratorio de Investigación y Desarrollo en Computación Científica (LIDECC)-
Departamento de Ciencias e Ing. de la Computación (DCIC), Universidad Nacional del Sur
(UNS). 8000 Bahía Blanca, Argentina

² Departamento de Ingeniería Química (DIQ), UNS. B8000 Bahía Blanca. Argentina

³ Planta Piloto de Ingeniería Química (UNS - CONICET), Camino La Carrindanga km. 7 -
8000 Bahía Blanca – Argentina
dybrigno@criba.edu.ar

Resumen. La digitalización es uno de los principales motores de la innovación pues cada vez hay más datos disponibles. Su explotación eficiente es imprescindible. En particular, para la optimización de sistemas de procesos es un desafío combinar métodos cimentados en ecuaciones con estrategias basadas en datos. Presentamos un enfoque que consiste en la integración de modelos físicos con el aprendizaje automático. Se propone una metodología basada en la física para derivar un metamodelo utilizando boosting. Para el ajuste paramétrico se emplea el Optimizador Hiperheurístico Cooperativo Paralelo (PCHO). PCHO comprende las metaheurísticas clásicas denominadas Algoritmos Genéticos, Recocido Simulado y Optimización por Enjambre de Partículas, cuyos parámetros se eligen de forma adaptativa durante las ejecuciones. Algunos mecanismos complejos pueden requerir varias instancias para representar la realidad. El objetivo de los metamodelos es unificar varias estructuras posibles para obtener resultados generales. Se describe el diseño conceptual de este enfoque, cuyo marco general se concibió para encontrar un metamodelo de la velocidad de reacción inversa del desplazamiento de gas en agua (RWGS) a 600°C. Para este caso se generaron datos experimentales para varios escenarios reales. En consecuencia, fue necesario ajustar varias correlaciones potencialmente aplicables que se combinaron en un metamodelo final.

Palabras clave: Inteligencia Artificial, Optimización. Aprendizaje Automático, RWGS, Metamodelos.

1 Introducción

En la era de la Industria 4.0, la digitalización se ha convertido en uno de los principales impulsores de la innovación. Los datos están cada vez más disponibles y su explotación eficiente es necesaria cuando se buscan mejoras en varios dominios. Además, resolver problemas del mundo real a menudo es excesivamente costoso, requiriendo formulaciones diferentes.

El aprendizaje automático (ML: Machine Learning) ha surgido como una alternativa prometedora, lo que hace posible implementar algoritmos especializados para lograr predicciones precisas [1]. En primer lugar, se utiliza una partición del conjunto de datos original en subconjuntos de entrenamiento y de prueba. Luego, se lleva a cabo una estimación de errores de generalización de $(m + 1)$ pliegues de validación cruzada. Esta estimación es el error de generalización medio sobre todos los modelos [2]. Una vez calculadas las estimaciones, los modelos se entrenan en el nuevo conjunto de entrenamiento. Para garantizar la diversidad, cada modelo individual se entrena utilizando un algoritmo diferente o un conjunto de hiperparámetros.

Finalmente, se construyen metamodelos combinando modelos base individuales. En [3] se afirmó que los conjuntos suelen ser mucho más precisos que los clasificadores individuales que los componen. En [4] se introdujo la generalización apilada que es un esquema que minimiza la tasa de error de generalización de uno o más generalizadores. En cuanto al enfoque de apilamiento (stacking), en [5] se presentó el marco de apilamiento. Como conclusión de su evaluación experimental, los autores afirmaron que seleccionar el mejor en un conjunto mediante la aplicación de varios algoritmos de aprendizaje funciona igual de bien en comparación con una metodología de validación cruzada. Además, en [6] se construyeron conjuntos de modelos de aprendizaje automático y, por lo tanto, pudo mejorarse el rendimiento predictivo.

Los modelos basados en datos se centran en la Inteligencia Computacional y los métodos de ML se pueden emplear para construir modelos con el fin de complementar o reemplazar modelos basados en la física [7]. En [8] se proporcionó un estudio exhaustivo de trabajos basados en ML. Los autores destacaron la máxima importancia de forzar la plausibilidad física en el ML para modelos del mundo real. En particular para la optimización de procesos, es un desafío combinar métodos basados en ecuaciones con métodos basados en datos. En este trabajo, se presenta un procedimiento de ML para generación de metamodelos, el cual trabaja en paralelo empleando un optimizador hiperheurístico cooperativo.

2 Estrategia basada en datos para el Aprendizaje Automático

Nuestro enfoque de modelado basado en datos se enfoca en técnicas de aprendizaje automático que se aplican a la construcción de modelos con el fin de complementar y/o reemplazar modelos físicos. Los algoritmos de aprendizaje automático se utilizan para determinar la relación entre las entradas X y las salidas Y_{obs} de un sistema mediante un conjunto de datos de entrenamiento que es representativo de todo el comportamiento encontrado experimentalmente en el sistema (Fig. 1). El entrenamiento tiene como objetivo minimizar el error del modelo $\|Y_{obs} - Y_{mod}\|$. Una vez que el modelo ha sido entrenado, se evalúa utilizando un conjunto de datos independientes para determinar su capacidad predictiva, es decir, si puede generalizarse a cualquier dato.

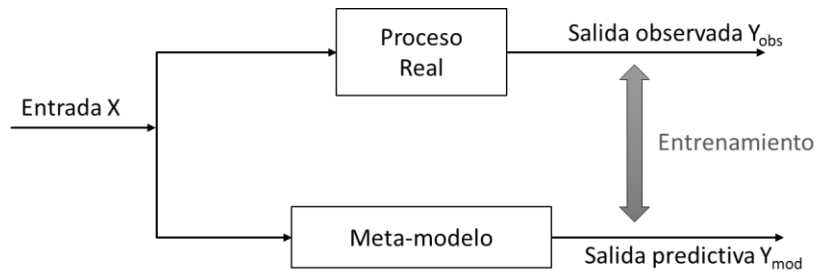


Fig. 1: Enfoque general para nuestro enfoque de modelado basado en datos.

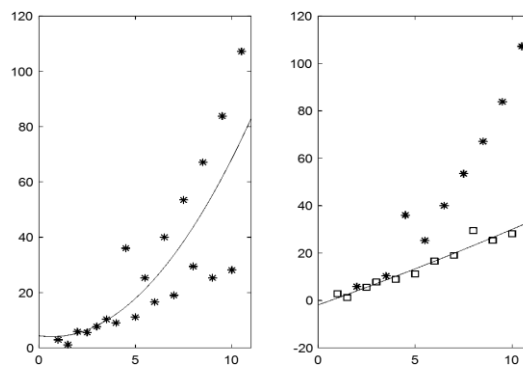


Fig. 2: El dilema del ajuste.

¿Por qué es necesario ajustar varios modelos base? Debido a las características del problema, diferentes escenarios físicos en la experimentación pueden generar datos que pueden no coincidir en su totalidad con un solo modelo teórico. En otras palabras, los datos necesitan ser clasificados en varios subconjuntos. Por lo tanto, se debe considerar la posibilidad de que pertenezcan a diferentes modelos base. Esto implica que todos los datos deben agruparse y luego, cada subconjunto debe ajustarse en consecuencia. La Fig. 2 ilustra dicho concepto. El gráfico izquierdo muestra el ajuste para el conjunto de datos completo. Aunque es el ajuste óptimo, exhibe un grado de dispersión que hace necesaria la filtración. En contraste, el gráfico de la derecha muestra los puntos seleccionados por el filtro (cuadrados) y su respectivo ajuste de curva. Naturalmente, los puntos restantes (asteriscos) se ajustarán a modelos alternativos.

La metodología propuesta consta de 3 pasos principales: Impulso, Generación del Metamodelo y su Validación.

2.1 Paso 1. Impulso (Boosting)

El "Boosting" [11] es un método de conjunto que combina varios aprendices débiles en un aprendiz fuerte. Entrena modelos secuencialmente con diferentes característi-

cas, cada uno intentando corregir a su antecesor (Fig. 3). Un modelo M_1 se construye a partir del conjunto de datos de entrenamiento para optimizar los coeficientes de una correlación bien conocida. Luego, todos los datos se clasifican para individualizar los datos que no coinciden con el modelo resultante. La discriminación de experimentos se realiza mediante técnicas de aprendizaje automático, tal como se explica en la sección de Resultados. Finalmente, los datos pertenecientes a la categoría de desajuste se utilizan para construir el segundo modelo M_2 . Este procedimiento continúa, y se agregan modelos hasta que se prediga correctamente todo el conjunto de datos de entrenamiento, o se hayan agregado el número máximo de modelos propuestos M_m . En este punto, los datos que permanecen en la categoría de desajuste se guardan en un conjunto de reserva R. Finalmente, podrían modelarse mediante un método general de aprendizaje automático, como una técnica de árbol de decisión [12], para obtener $M_{(m+1)}$. Este modelo adicional tiene más libertad para mejorar las propiedades de generalización cuando los fenómenos intrínsecos no pueden representarse adecuadamente a través de una fórmula específica.

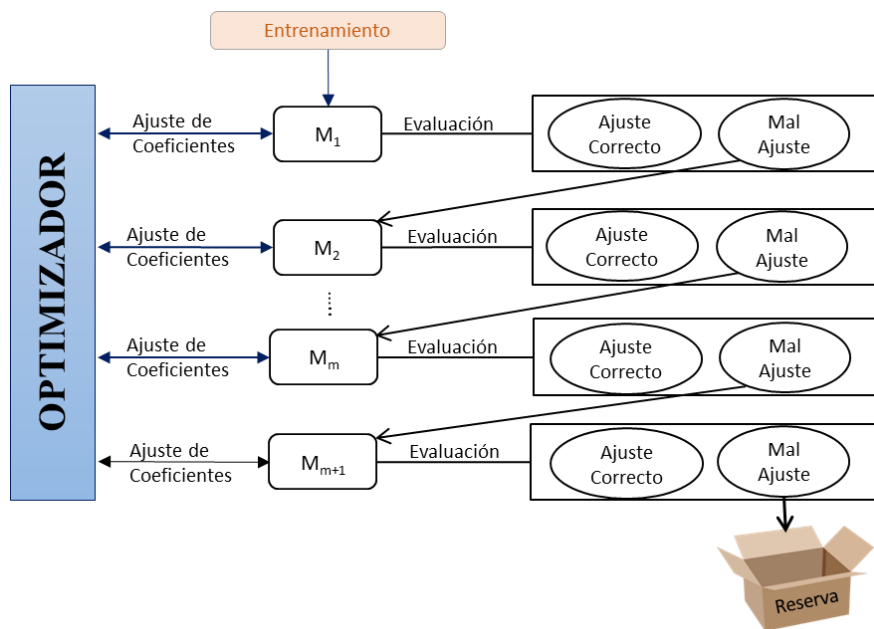


Fig. 3: Esquema del método Boosting para modelado basado en la interpretación física.

Hoy en día, los algoritmos de optimización metaheurísticos resultan muy atractivos por sus claras ventajas frente a los algoritmos tradicionales. Dado que los metaheurísticos pueden resolver múltiples objetivos, soluciones múltiples y formulaciones no lineales, se emplean para encontrar soluciones de alta calidad a un número cada vez mayor de problemas complejos del mundo real, como los combinatorios.

Por esta razón, en este trabajo hemos elegido al Optimizador Hiperheurístico Cooperativo Paralelo (PCHO) [9] para generar las correlaciones. Es un hiperheurístico de alto nivel que fue desarrollado originalmente para la optimización de ingeniería de procesos. El algoritmo, esquematizado en la Fig. 4, comprende las siguientes metaheurísticas clásicas: Algoritmos Genéticos (GA), Recocido Simulado (SA) y Optimización por Enjambre de Partículas (PSO), cuyos parámetros se eligen de forma adaptativa durante las ejecuciones. En [9] se experimentó PCHO en colecciones que contienen exclusivamente problemas de Ingeniería Química. Luego, en [10] se analizó el rendimiento de PCHO en una variedad de experimentos para validar su comportamiento.

La forma en que se comparte la información establece el grado de cooperación entre los agentes. La política de actualización depende de cada metaheurística. En el caso de PSO y GA, la población se actualiza convenientemente, mientras que en SA el punto de inicio cambia. Dependiendo de su entrada, cada metaheurística adopta un comportamiento diferente para incluir el mejor descubrimiento hasta el momento. La información compartida proporciona una idea sobre la ubicación espacial de áreas prometedoras. Por lo tanto, se puede generar un nuevo conjunto de muestras. Por ejemplo, dado que el potencial progenitor (es decir, la solución candidata recientemente informada) aparentemente es mejor, el conjunto enriquecido agregará diversidad a la población inicial de GA, favoreciendo así la convergencia.

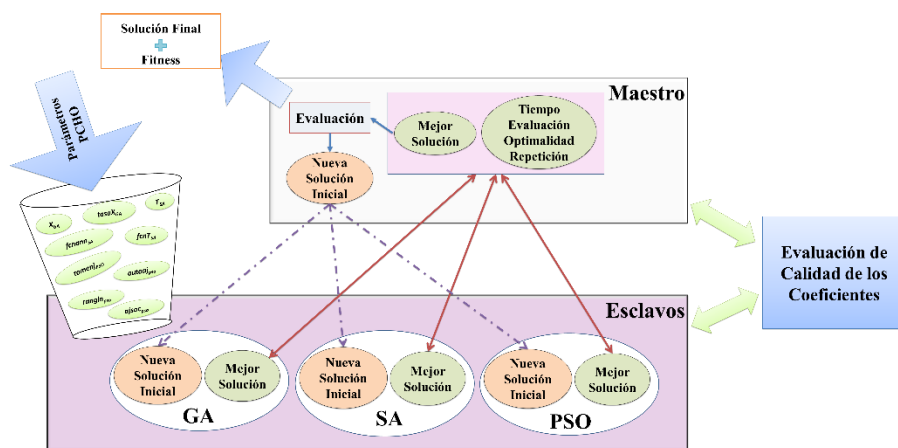


Fig. 4: Esquema del funcionamiento del optimizador PCHO.

2.2 Paso 2. Generación del metamodelo

En nuestro enfoque, el conjunto de modelos base $B=\{M_1, \dots, M_m\}$ es un conjunto ordenado que contiene modelos basados en física que son interpretables a nivel modular. El objetivo es crear un único metamodelo predictor a partir de las predicciones de este conjunto de modelos base, cuyo índice corresponde a la jerarquía que detentan dentro del conjunto en relación con la preferencia en el orden a ser utilizados. Es decir, como se intentará clasificar los experimentos recorriendo el conjunto B , se nom-

bra M_1 al modelo elegido como primero para intentar clasificar los experimentos, luego con aquellos experimentos no clasificados se intentará con M_2 y así sucesivamente. Por otro lado, si la característica físicamente más relevante del proceso está presente en s modelos, dichos modelos formarán parte de un subconjunto de modelos $S_c = \{M_{ind(1)}, \dots, M_{ind(s)}\} \subseteq B$. La indización $ind(j)$ indica el índice que ocupa el modelo en el conjunto B . Como ejemplo, si la característica física f reúne a los modelos M_3, M_4 y M_7 , entonces $ind(1)=3, ind(2)=4$ e $ind(3)=7$. En este subconjunto también los modelos ocupan un orden jerárquico.

Por cada modelo M_i y con el conjunto de experimentos que se busca clasificar se establece un perceptrón P_i , es decir, un algoritmo para el aprendizaje supervisado de clasificadores binarios [13]. Estos se utilizan para clasificar datos y en nuestro caso, dado un experimento, el binarismo del perceptrón P_i refiere a la predicción acerca de si el modelo M_i será o no el adecuado para dicho experimento. La elección de perceptrones lineales responde a que ofrecen una serie de ventajas en términos de simplicidad, interpretabilidad, eficiencia computacional y prevención de sobreajuste. [14]

Para cada experimento, se consideran las variables lógicas $h, \{p_1, \dots, p_m\}$ que representan respectivamente la presencia de la característica más relevante de los experimentos y la calidad/pertinencia de los ajustes. Estas variables booleanas sirven para almacenar los siguientes valores lógicos: 1=verdadero; 0=falso.

Entonces, comenzando con todas sus variables booleanas inicializadas en 0, para una única característica relevante h , y agregando un conjunto R como reserva para aquellos experimentos que no puedan ser clasificados, el metamodelo $Meta(e_i)$ formulado en la Ec. 1 representa el proceso de clasificación en función de sus características del experimento i -ésimo en donde las variables binarias van modificando sus valores en función de los criterios de selección, y de esa forma cada experimento queda incluido en el conjunto correspondiente al modelo que mejor lo ajusta.

$$\begin{aligned}
 Meta(e_i) = & h \left(p_{ind(1)} M_{ind(1)} + \sum_{j=2}^s \left(p_{ind(j)} M_{ind(j)} \prod_{k=ind(1)}^{ind(j)-1} (1 - p_{ind(k)}) \right) \right) + \\
 & + (1 - h) \left(p_1 M_1 + \sum_{j=2}^m (p_j M_j \prod_{k=1}^{j-1} (1 - p_k)) \right) + \\
 & (1 - h) \prod_{i=1}^m (1 - p_i) R
 \end{aligned} \tag{1}$$

En el primer término de la Ec. 1 se recorren los modelos del subconjunto S_c correspondiente a la presencia de la característica física que setea un 1 en la variable h , anulando el segundo y tercer término. El segundo término considera la posibilidad de la ausencia de la característica física, lo que deja a la variable h en su valor original igual a cero. Esto anula el primer término, y los modelos del conjunto B son recorridos uno a uno en su orden de jerarquía hasta que alguna de las variables binarias p_i tome valor 1, lo que anularía el tercer término. Si al recorrer todos los modelos, ninguna de las variables binarias modifica su valor inicial cero, entonces se anulan el primero y el segundo término quedando el experimento ubicado en el conjunto de reserva R que contiene todos los experimentos no clasificados.

2.3 Paso 3. Validación del metamodelo

El metamodelo puede validarse simultáneamente desde dos puntos de vista: numérico y fisicoquímico. Meckesheimer et al. [15] investigaron métodos de evaluación computacionalmente baratos para la validación de metamodelos basados en la validación cruzada *leave-k-out* y desarrollaron directrices con el objeto de seleccionar *k* para diferentes tipos de metamodelos. Además de emplear esta técnica, la evaluación numérica del rendimiento del conjunto de modelos se efectuará comparando las predicciones de dicho conjunto con los valores reales de la colección de datos de prueba, usando como métrica el porcentaje de exactitud en la predicción de cada perceptrón. Estos valores de legitimación son los expuestos en la sección de Resultados.

3 Caso de Estudio: Modelado cinético de la reacción RWGS

El ajuste de coeficientes para el diseño riguroso de reactores constituye el caso de estudio principal que inspiró el desarrollo de esta estrategia. La estimación de coeficientes cinéticos a menudo da lugar a formulaciones muy complejas y realizar experimentos es costoso, consume mucho tiempo y es muy engorroso debido a mecanismos de reacción complejos. Los datos son escasos porque estos experimentos físicos son muy costosos y su obtención generalmente lleva mucho tiempo. Por lo tanto, las condiciones de operación se optimizan solo en una cantidad limitada de datos experimentales. Siempre que se generen nuevos datos, es importante incorporar esta información de manera eficiente para mejorar los resultados de optimización.

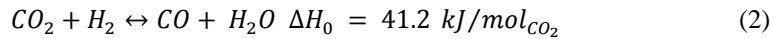
Predecir la velocidad de la Reacción Inversa de Desplazamiento del Gas de Agua (RWGS) es de gran importancia en la industria de procesos. Hay muy pocas correlaciones disponibles en la literatura para prever esta velocidad, las cuales han sido diseñadas exclusivamente para los mecanismos propios de catalizadores específicos tales como los basados en Ni. Más aún, en su mayoría son aplicables a un rango estrecho de bajas temperaturas. Esta condición operativa lleva a la predicción de velocidades de reacción excesivamente altas, lo que hace imposible usarlas para un diseño de reactor riguroso. Además, se desea ajustar modelos para emplear el catalizador de Ce-Pd sintetizado en nuestro laboratorio. Por lo tanto, es un desafío desarrollar un modelo representativo que imite la realidad.

La reacción RWGS se ha vuelto muy importante en la actualidad y se han realizado varios estudios cinéticos en rangos de temperatura entre 200 °C y 350 °C en diferentes tipos de catalizadores [16]. En contraste, en [17] trabajaron con un catalizador de Ni fijando la temperatura en 477 °C. Sin embargo, no hay información disponible en la literatura sobre los mecanismos de reacción de los catalizadores de Ce-Pd a altas temperaturas.

3.1 Descripción del problema

Es necesario ajustar un modelo cinético preciso para la velocidad de reacción RWGS sobre catalizadores de Ce-Pd/ α -Al₂O₃ para el diseño riguroso de un reactor

RWGS para la producción de metanol [18]. No existen modelos disponibles para las condiciones de operación del proceso propuesto (600 °C). En este caso, derivar ecuaciones cinéticas adecuadas para altas temperaturas constituye una brecha en la investigación. La reacción RWGS [19] es una reacción química de gran relevancia (Ec. 2). En [20] se ha determinado que la RWGS es una reacción de equilibrio favorecida a temperaturas más altas debido a su carácter moderadamente endotérmico y tiempos de contacto más bajos.



3.2 Formulación del modelo y enfoque experimental

La Ec. 3 es el balance de masa para cada componente i de la reacción en estado estacionario, donde W es la masa del catalizador; X_i la conversión; F_i^0 , es el flujo de alimentación y r_i , la velocidad de reacción.

$$\frac{dX_i}{d(W/F_i^0)} = r_i \quad (3)$$

En [21] se reportaron detalladamente sobre la cinética de la síntesis de metanol, incluyendo algunas correlaciones especialmente desarrolladas para la tasa de reacción RWGS. Aunque existen modelos propuestos para la reacción RWGS en catalizadores comerciales, el mecanismo de la reacción en catalizadores de Ce-Pd es desconocido.

En primer lugar, se seleccionó un modelo cinético M_1 (Ec. 4), denominado Modelo de Langmuir-Hinshelwood-Hougen-Watson (LH-HW) [22], donde las variables de optimización son K_1, K_{CO_2}, K_0 y k_f . Este enfoque es frecuentemente usado para conseguir predicciones útiles, pero no siempre tiene la validez teórica que se le suele atribuir.

$$r_i^{LH-HW} = k_f \left(p_{CO_2} p_{H_2} - \frac{p_{CO} p_{H_2O}}{K_{RWGS}} \right) / \left((1 + K_1 p_{H_2}) \left(1 + K_{CO_2} p_{CO_2} + \frac{p_{CO} p_{H_2O}}{K_0} \right) \right) \quad (4)$$

Una desventaja del modelo LH-HW es que su empleo nos conduce a un nivel de complejidad matemática innecesario, siendo su formulación poco amigable desde el punto de vista numérico. En contraste, las correlaciones simples basadas en funciones exponenciales se consideran representativas para velocidades que reflejan cualquier mecanismo de reacción [23]. Por esta causa, este estilo de formulación es muy conveniente. Más aún, hay estudios cinéticos para la reacción WGS a bajas temperaturas

que proponen modelos empíricos de ley de potencia con una sólida base en datos experimentales [24]. Por lo tanto, esta formulación parecería ser prometedora para realizar el ajuste de coeficientes cinéticos de la reacción RWGS.

Por ende, como una alternativa viable (M_2), se adoptó el modelo cinético clásico (Ec. 5) denominado Ley de las Potencias (PL: Power-Law). En la Ec. 5 las presiones parciales se denotan como p_i ; K_{RWGS} es la constante de equilibrio y las variables de optimización son $\alpha, \varepsilon, \gamma, \delta$ y k .

$$r_i^{PL} = k p_{CO_2}^\alpha p_{H_2}^\varepsilon p_{CO}^\gamma p_{H_2O}^\delta \left(1 - \frac{p_{CO} p_{H_2O}}{K_{RWGS} p_{CO_2} p_{H_2}}\right) \quad (5)$$

4 Resultados y discusión

A partir de un conjunto de 113 experimentos, se seleccionaron al azar 90 de estos para entrenamiento y 23 para la validación de los predictores. Para el entrenamiento, los experimentos se dividen en tres grupos a partir de la composición de la mezcla de alimentación:

- Grupo 1: mezcla CO₂/H₂/CH₄/Ar. (27 experimentos)
- Grupo 2: mezcla CO₂/H₂/CH₄/CO/Ar (32 experimentos)
- Grupo 3: mezcla CO₂/H₂/CH₄/H₂O/Ar (31 experimentos)

Siguiendo la Fig. 3 y estableciendo como modelo 1 el modelo LH-HW, se ajustaron los coeficientes a través del optimizador PCHO. Siendo $n=90$ el número de experimentos involucrados, el ajuste se realiza minimizando la función indicada en la Ec.6, donde R_{an} es el vector de los resultados analíticos arrojado por el modelo, R_{exp} son los resultados obtenidos experimentalmente y p es el vector de coeficientes que actúa como variable de decisión.

$$resi(p) = \sqrt{\frac{\sum_{i=1}^n (R_{an}(p) - R_{exp})^2}{n-1}} \quad (6)$$

En sucesivas ejecuciones se obtuvieron diferentes conjuntos de coeficientes, pero todos mostrando un mismo orden en el valor residual, cómo se puede observar en la Tabla 1.

En primer lugar, el filtro establecido se basa en discriminar aquellos experimentos que aporten más valor residual a la función $resi(p)$. Entonces, por cada experimento i tenemos $R_{exp}^{(i)}$ y $R_{an}^{(i)}$ y su diferencia dada por la Ec. 7.

$$Dif^{(i)} = (R_{exp}^{(i)} - R_{an}^{(i)})^2 \quad (7)$$

Para establecer un umbral a partir de la distancia entre los valores experimentales y los que arroja el modelo utilizamos la función tangente hiperbólica de la diferencia

entre R_{an} y R_{exp} (Ec. 8). Así se logra convertir esa diferencia en un valor interpretable como una probabilidad.

$$P^{(i)} = \text{Tanh}(\text{Dif}^{(i)}) = n \quad ; \quad n \in [0,1] \quad (8)$$

Tabla 1. Conjuntos de coeficientes ajustados con el PCHO para M_1 y sus correspondientes valores residuales.

K_f	K_l	K_{CO_2}	K_o'	Residuo
5.0527	0.4392	0.3180	701.2493	0.1603
18.5463	0.3549	0.6924	37.3680	0.1606
11.7663	0.4712	0.5251	862.7935	0.1605
5.0527	0.4392	0.3180	701.2493	0.1608
19.0336	0.8597	0.2024	737.4774	0.1605
17.5896	0.1397	0.1327	540.7241	0.1605
25.7510	0.4104	0.4736	242.8354	0.1607
20.1578	0.2645	0.2677	758.4986	0.1607
9.4407	6.5225	3.2050	9.0067	0.1611

Si este valor $P^{(i)}$ es menor que una cierta tolerancia tol , $P^{(i)} < tol$ entonces se considera que el modelo ajusta bien al experimento i , es decir, el experimento i es aceptado pues queda dentro del conjunto correspondiente al modelo.

El conjunto de coeficientes con los que realizamos la experiencia numérica es {5.0527, 0.4392, 0.3180, 701.2493}. La Tabla 2 muestra el valor del residuo y su porcentaje de reducción al evaluar el modelo exclusivamente en los puntos aceptados por el filtro, a partir de diferentes valores del umbral de tolerancia.

Tabla 2. Clasificación de los experimentos para distintos umbrales de tolerancia.

Tolerancia	0.1	0.095	0.09	0.085	0.08	0.075
Residuo	0.0805	0.0732	0.0614	0.0584	0.0584	0.0539
% reducción	49.94	54.48	61.82	63.68	63.68	66.48
Experimentos excluidos	64	71	77	78	78	80

Cuanto más fina es la tolerancia, mayor es el porcentaje de reducción, pero al mismo tiempo es mayor la cantidad de puntos excluidos. Fijando la tolerancia en 0.1, quedaron 64 experimentos apartados del modelo LH-HW. La búsqueda de patrones predictivos arroja un primer resultado a partir de la interpretabilidad de los datos: este grupo de 64 experimentos rechazados incluye el total de 31 experimentos del Grupo 3. La característica relevante de este grupo respecto a los otros dos es que en los experimentos se incluye H₂O en su mezcla de alimentación, lo que permite establecer un primer criterio de discriminación. Para los restantes 33 experimentos sin clasificar se

reutilizan los datos de los 90 experimentos empleando una técnica de aprendizaje automático. En este caso se busca un perceptrón lineal que indique la separación entre los datos etiquetados con 0 (excluido del modelo) o 1 (aceptado en el modelo).

Para establecer el criterio de predicción acerca de la bondad del ajuste respecto a un nuevo dato se combinan la interpretabilidad de los datos con la pauta establecida por el perceptrón, es decir, la presencia de agua en la mezcla de alimentación y el criterio de exclusión del perceptrón. Aquellos datos rechazados se reúnen en un subconjunto con el que se realiza un ajuste de coeficientes para el modelo alternativo PL. Con este modelo alternativo se repite el proceso de filtrado en función de la diferencia entre lo experimentado y lo predicho por el nuevo modelo, y se entrena un nuevo perceptrón para anticipar la aceptación o no del dato a este nuevo modelo. La Fig. 5 esquematiza este circuito de los datos. Fig. 6 ilustra distribución resultante de los puntos experimentales.

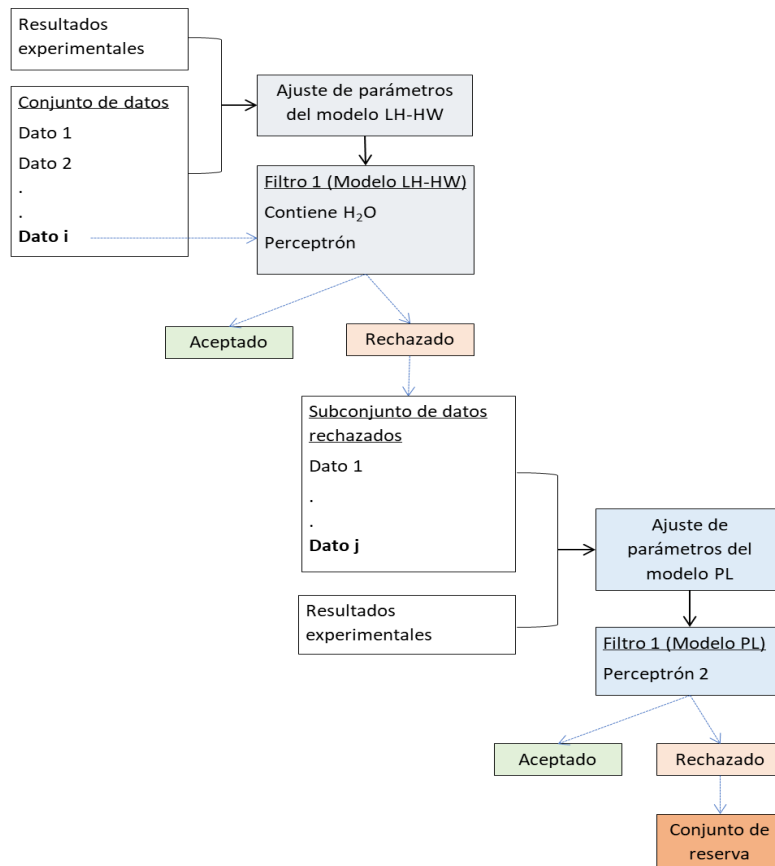


Fig. 5. Boosting con filtrado para este caso de estudio.

113 datos. Modelo LHHW. Residuo = 0.1603	
28 aceptados LHHW. Residuo = 0.0811	85 datos. Modelo PL. Residuo = 0.1154
41 aceptados PL. Residuo = 0.0964	44 rechazados

Fig. 6. Distribución de los experimentos según el criterio de clasificación.

Teniendo establecidos criterios de aceptación de los modelos LH-HW y PL para cada experimento, se puede establecer un metamodelo que integre los dos modelos con sus filtros y un eventual conjunto de experimentos en reserva no clasificados. Hemos denominado M_1 = Modelo LH-HW; M_2 = Modelo PL y R = Conjunto de reserva. Dado un experimento, se establecen los siguientes valores para las variables lógicas:

- $h = 0$: El experimento no contiene agua; $h = 1$: El experimento contiene agua en la alimentación.
- $p_1 = 0$: El modelo M_1 no ajusta satisfactoriamente al experimento; $p_1 = 1$: El modelo M_1 es un buen ajuste para el experimento.
- $p_2 = 0$: El modelo M_2 no ajusta satisfactoriamente al experimento; $p_2 = 1$: El modelo M_2 es un buen ajuste para el experimento.

Siendo e_i el experimento i -ésimo e iniciando con todas las variables booleanas inicializadas en 0, el metamodelo que representa al experimento e_i en función de sus características se puede escribir tal como en la Ec. 9.

$$Meta(e_i) = hp_2M_2(e_i) + (1-h)(p_1M_1(e_i) + (1-p_1)p_2M_2(e_i) + (1-h)(1-p_1)(1-p_2)R(e_i)) \quad (9)$$

De esta forma, en la medida en que las variables binarias van modificando sus valores en función de los criterios de selección, cada experimento queda incluido en el conjunto correspondiente al modelo que mejor lo ajusta. En consecuencia, la calidad del metamodelo está determinada por la validación de los modelos base. En este caso de estudio, la validez se corroboró empleando los 23 datos reservados para el testeo. El resultado es satisfactorio dado que los perceptrones muestran una precisión del 86% y 76% respectivamente.

5 Conclusiones

Hemos desarrollado el diseño conceptual de una metodología basada en la física con el objetivo de encontrar un metamodelo integral para la velocidad de la reacción RWGS a 600°C, donde solo estaban disponibles unos pocos datos experimentales. Este metamodelo se apoya en la interpretabilidad de los datos y en el aprendizaje

sobre las predicciones de los modelos base. El diseño del metamodelo nos permitió unir distintas estructuras para producir resultados generales, dado que la complejidad del mecanismo de la reacción RWGS a 600°C muestra que no es representable desde una única instancia.

Finalmente, dado que el ajuste de coeficientes en estos modelos base es fundamental para obtener buenas predicciones del metamodelo, se eligió PCHO como la herramienta de optimización sólida para el ajuste de estos coeficientes dado que su comportamiento fue validado en una variedad de experimentos [10]. PCHO tiene un diseño escalable, no dependiente del cálculo de gradientes y con capacidad de manejar funciones no convexas que mostró muy buena performance en colecciones de problemas de ingeniería química [9]. Siendo PCHO el optimizador designado para resolver el subproblema de optimización, su desempeño fue altamente satisfactorio. Es un desafío ampliar su alcance mediante el diseño de procedimientos para hibridarlo con esquemas basados en datos. Además, se pueden proponer otras versiones mejoradas de PCHO con más especificidad, es decir, con capacidades para resolver problemas de optimización en línea o software especializado para problemas multiobjetivo.

Declaración de intereses. Los autores declaran que no tienen intereses contrapuestos relevantes para el contenido de este artículo.

Referencias

1. Karniadakis, G. E. Kevrekidis, I. G. Perdikaris, L. Lu, P. Wang, S. Yang, L.: Physics-informed learning. *Nature Reviews Physics*, 3(6), 422-440 (2021)
2. Kochenderfer, M. J. Wheeler T.A.: *Algorithms for optimization*. Mit Press (2019)
3. Dietterich, T. G.: Machine-learning research: Four current directions. *AI Mag.*, 18:4, 97-136 (1997)
4. Wolpert, D.: Stacked generalization, *Neural Networks*, 5:2, 241-260 (1992)
5. Džeroski, S. Ženko, B.: Is combining classifiers with stacking better than selecting the best one? *Machine learning*, 54, 255-273 (2004)
6. Pavlyshenko, B.: Using stacking approaches for machine learning models. In: 2018 IEEE Second International Conf. on Data Stream Mining & Processing (DSMP), 255-258, IEEE (2018)
7. Solomatine, D. See, L. M. Abrahart, R. J.: Data-driven modelling. *Practical hydroinformatics: Computational intelligence and technological developments in water applications*, 2, 17-30 (2008)
8. Meng, C. Seo, S. Cao, D. Griesemer, S. Liu, Y.: When physics meets machine learning: A survey of physics-informed machine learning. *arXiv preprint arXiv:2203.16797* (2022)
9. Oteiza, P.P., Ardenghi, J.I., Brignole, N.B.: Parallel hyperheuristics for process engineering optimization. *Comp. & Chem. Eng.*, 153, 107440 (2021)

10. Debortoli, D.O., Oteiza, P.P., Ardenghi J.I., Brignole, N.B.: Evaluación de desempeño del optimizador hiperheurístico cooperativo paralelo (PCHO). MECOM 2023 (2023)
11. Zhou, Z.-H.: Ensemble learning. *Encyclopedia of Biometrics*, 1, 270–273 (2009)
12. Steinberg, D. Colla, P. CART: classification and regression trees. *The Top Ten Algorithms in Data Mining*, 9, 179 (2009)
13. Abu-Mostafa, Y. S., Magdon-Ismail, M., Lin, H. T.: *Learning from data: A short course*. AMLBook (2012).
14. Bishop, C. M.: *Pattern Recognition and Machine Learning*. Springer (2006).
15. Meckesheimer, M., Barton, R. R., Simpson, T. W., Booker, A. J.: Computationally inexpensive metamodel assessment strategies. En: *International design engineering technical conferences and computers and information in engineering conference*, 80227, 191-201, American Society of Mechanical Engineers. (2002)
16. Byron Smith., R.J., Loganathan, M., Shantha, M. S.: A review of the water gas shift reaction kinetics. *Int. Journal of Chemical Reactor Engineering*, 8(1) (2010)
17. Santos, M. F., Bresciani, A. E., Ferreira, N. L., Bassani, G. S., Alves, R. M.: Carbon dioxide conversion via reverse water-gas shift reaction: Reactor design. *Journal of Environmental Management*, 345, 118822 (2023)
18. Cañete, B. Gigola, C. E Brignole, N. B.: Enhancing the potential of methane combined reforming for methanol production via partial CO₂ hydrogenation. *Ind. Eng. Chem. Res.*, 56(22), 6480–6492 (2017)
19. Samimi, F. Hamed, N. Rahimpour, M.: Green methanol production from indirect CO₂ conversion: RWGS reactor vs RWGS membrane reactor. *J. Env. Chem. Eng.*, 7(1), 102813 (2019)
20. Whitlow, J. E. Parrish, C. F.: Operation, modelling and analysis of the reverse water gas shift process. In *AIP Conference Proceedings 654 (1)*, 1116-1123, American Inst. Physics (2003)
21. Bozzano, G. Manenti, F.: Efficient methanol synthesis: Perspectives, technologies and optimization strategies. *Progress in Energy and Combustion Science*, 56, 71-105 (2016)
22. Nelson, N. C. Szanyi, J.: Heterolytic hydrogen activation: understanding support effects in water–gas shift, hydrodeoxygenation, and CO oxidation catalysis. *ACS Catalysis*, 10(10), 5663-5671 (2020)
23. Weller, S.: Analysis of kinetic data for heterogeneous reactions. *AIChE Journal*, 2(1), 59-62 (1956)
24. Maklavany, D. M. Shariati, A. Khosravi-Nikou, M. R. Roozbehani, B.: Hydrogen production via low temperature WGS reaction: kinetic study, mathematical modelling, simulation and optimization of catalytic fixed bed reactor using gPROMS. *Chem. Prod. Process Model*, 12(3) (2017)