

# Nuevo modelo matemático basado en inteligencia artificial y análisis de sensibilidad para el dimensionamiento de colectores solares de placa plana

A.J. Cetina-Quiñones, Diana Romero, A. J. Hernández Bautista, E. Chan-Dzib, A. Bassam

*Laboratorio de Modelado y Optimización de Procesos Energéticos y Ambientales, Facultad de Ingeniería, Universidad Autónoma de Yucatán, Av. Industrias no contaminantes, Mérida, Yucatán, México.*

*Fecha de recepción: 24 de febrero de 2024 - Fecha de aceptación: 4 de abril de 2025*

## Resumen

Este estudio presenta el desarrollo de un modelo surrogado basado en algoritmos de aprendizaje automático y un análisis de sensibilidad para un sistema de colectores solares de placa plana. El modelo fue entrenado utilizando una base de datos generada mediante el método f-chart, considerando siete variables de entrada: mes, número de colectores, irradiación solar, temperatura ambiente, velocidad del viento, temperatura y consumo de agua caliente sanitaria. La variable de salida objetivo fue la cobertura solar mensual. Los resultados revelaron que las redes neuronales artificiales demostraron mejor rendimiento reportando valores del coeficiente de determinación de 0.9956, un error cuadrático medio de 0.0188 y un error medio absoluto de 0.0159. El análisis de sensibilidad basado en el algoritmo de Garson reveló que el número de colectores fue la variable más influyente con 44 % de influencia sobre la cobertura solar. Estos resultados destacan la importancia de optimizar los parámetros de entrada para mejorar el rendimiento del sistema, así como el potencial de los algoritmos de aprendizaje automático como herramientas útiles y alternativas a los modelos convencionales en sistemas energéticos para la mejor toma de decisiones e identificación de áreas de oportunidad en la mejora de estos sistemas.

*Palabras claves: Inteligencia artificial, desarrollo sustentable, modelo inteligente, dimensionamiento de sistemas energéticos, cobertura solar.*

## New mathematical model utilizing artificial intelligence and sensitivity analysis for sizing flat-plate solar collectors

### Abstract

This study presents the development of a surrogate model based on machine learning algorithms and a sensitivity analysis for a flat-plate solar collector system. The model was trained with a database generated using the f-chart method, considering seven input variables: month, number of collectors, solar radiation, ambient temperature, wind velocity, temperature, and consumption of sanitary hot water. The target output variable was the monthly solar coverage. The results revealed that artificial neural networks demonstrated better performance, reporting a coefficient of determination value of 0.9956, a mean squared error of 0.0188, and a mean absolute error of 0.0159. The sensitivity analysis based on the Garson algorithm revealed that the number of

---

\*armando.cetina7@gmail.com

collectors was the most influential variable with 44% influence on solar coverage. These results highlight the importance of optimizing input parameters to improve system performance, as well as the potential of machine learning algorithms as useful and alternative tools to conventional models in energy systems for better decision-making and identifying areas of opportunity in the improvement of these systems.

**Keywords:** *Artificial intelligence, sustainable development, intelligent model, energy systems sizing, solar coverage.*

## 1. Introducción

La energía solar representa la fuente de energía renovable más abundante y tiene el potencial de satisfacer las necesidades energéticas del mundo (Hachicha et al., 2019). Dentro de este contexto, las tecnologías de energía solar térmica son esenciales para un futuro sostenible ya que proporcionan una fuente de energía limpia, renovable y doméstica (Buker y Ayompe, 2015). La energía solar térmica se produce aprovechando la energía del Sol, y el colector solar de placa plana (CSPP) es el componente principal que convierte la energía solar en calor (Cetina-Quiñones et al., 2021). El uso de CSPPs para la generación de energía térmica es la aplicación más común, típicamente utilizada para la generación de agua caliente comercial y doméstica. La principal ventaja de esta tecnología es que aprovecha la energía solar, por lo que se puede minimizar el uso de energía externa para el calentamiento de agua (Çapik et al., 2012). Como resultado, el diseño eficiente de los sistemas de calefacción de agua solar es fundamental para garantizar el beneficio óptimo para el cliente, especialmente para los sistemas grandes. Un sistema de agua caliente solar requiere el tamaño adecuado de diversos componentes dependiendo del aislamiento solar esperado y la demanda de agua calentada (Deepika et al., 2016). Un método ampliamente utilizado para dimensionar estos sistemas es el conocido método f-chart desarrollado por Klein y Beckman, que se utiliza para calcular el rendimiento térmico anual de los sistemas de calefacción activa para edificios (utilizando líquido o aire como fluido de trabajo) donde la temperatura mínima de suministro de energía es cercana a 20 °C (Deepika et al., 2016; Formacion et al., 2009). Un trabajo de

investigación en esta área fue desarrollado por (Rosli et al., 2019), donde se utilizó el método f-chart en el diseño de sistemas de calefacción solar líquido para estimar la fracción de la carga térmica total suministrada por el sistema solar de calentamiento hacia el sistema de aplicación de agua caliente doméstica en Melaka, Malasia. Los resultados revelaron que la demanda anual total de calefacción de agua caliente doméstica en Melaka era de 9.55 GJ, y la proporción anual de la carga suministrada por la energía solar fue del 78.42 %, lo que es aceptable para la implementación y la instalación. Rocha et al. (2022) presentaron el análisis de un calentador solar utilizando el método f-chart. Mediante el método f-chart, se realizaron simulaciones considerando diversos parámetros. Los resultados revelaron que, a pesar de que la superficie de la unidad experimental estaba más allá del rango sugerido, los resultados experimentales de la temperatura de salida alrededor de 55 °C coincidían con los modelos producidos. Sujith (2013) realizó un estudio de investigación sobre el diseño de un sistema de calefacción de agua solar para una familia de seis miembros en Calicut, India, utilizando el método f-chart. También se estudió la variación de la fracción de carga total suministrada por la energía solar y se realizó un análisis económico. Los resultados indicaron que el número óptimo de colectores solares era de seis metros cuadrados, considerando una capacidad de almacenamiento de 300 litros y una tasa de flujo de líquido de 0.06 L/s, cuya cobertura solar era del 81.12 % con un período de retorno esperado y ahorro de energía de 4.94 años y 1.000216 Rs/kWh, respectivamente.

Con base en los trabajos previamente reportados y de acuerdo con el conocimiento de

los autores, hay una carencia del estudio de la influencia de los parámetros de entrada que intervienen en el método f-chart. Esto indica que existe una brecha de oportunidad para el estudio de estos parámetros. Específicamente en el ámbito de los procesos térmicos, el análisis de sensibilidad desempeña un papel crucial al evaluar y optimizar la eficiencia de sistemas térmicos. Este enfoque permite comprender cómo diversos parámetros, tanto ambientales, de diseño y operacionales afectan el rendimiento de estos sistemas, lo que permitirá identificar factores determinantes que impactan directamente en las variables de interés (Borgonovo, 2023). Por otra parte, uno de los principales inconvenientes de los colectores solares durante su análisis es la necesidad de emplear múltiples variables de entrada en tiempo real para mejorar su rendimiento, lo que lo convierte en un sistema complejo (Pandey y Chaurasiya, 2017). Por esta razón, en los últimos años, diferentes trabajos de investigación han utilizado diversas técnicas de inteligencia artificial (IA) como herramienta de modelado para analizar a profundidad estos sistemas. Las técnicas basadas en IA tienen mayor capacidad para predecir el rendimiento de los sistemas de energía renovable, ya que pueden aprender de patrones de información significativos sin necesidad de procedimientos matemáticos complejos (Gopi et al., 2023). Una de las técnicas más empleadas es la creación de modelos surrogados basados en algoritmos de aprendizaje automático, tales como las Redes Neuronales Artificiales (RNA) (Dehaj Shafiey et al., 2020; Liu et al., 2021), Árboles de Regresión (AR) (Ahmad et al., 2018; Tixeira-Pinto, 2023) y Máquinas de Vectores de Soporte (MVS) (Kaood et al., 2021; Zayed et

al., 2023), los cuales han sido ampliamente empleadas en trabajos relacionados con el aprovechamiento de la energía solar térmica (Dehaj Shafiey et al., 2020; Ghritlahre et al., 2020; Shen et al., 2021; Tomy et al., 2016). Por otra parte, otra herramienta muy utilizada para el análisis del rendimiento de cualquier sistema que complementa al modelo surrogado corresponde al análisis de sensibilidad (AS), que describe el porcentaje de influencia de las variables de entrada de un sistema sobre las variables de salida (Ghritlahre et al., 2020). Este análisis permite una comprensión más profunda y una evaluación comparativa de la importancia de los parámetros del sistema, en este caso, del sistema de calentamiento de agua.

De acuerdo con el contexto anterior, en este trabajo se desarrolla un modelo surrogado a partir del entrenamiento de algoritmos de aprendizaje automático de un sistema de calentamiento solar doméstico integrado con colectores solares de placa plana. El objetivo del trabajo es crear un modelo sustituto al modelo energético convencional, así como evaluar la influencia de los parámetros de entrada del sistema sobre la variable de salida.

## 2. Metodología

### 2.1. Descripción del sistema

El sistema propuesto se enfoca en la producción de agua caliente sanitaria para uso doméstico, empleando un arreglo de colectores solares de placa plana. La Figura 1 ilustra el diagrama esquemático del sistema (Cetina-Quiñones et al., 2023).

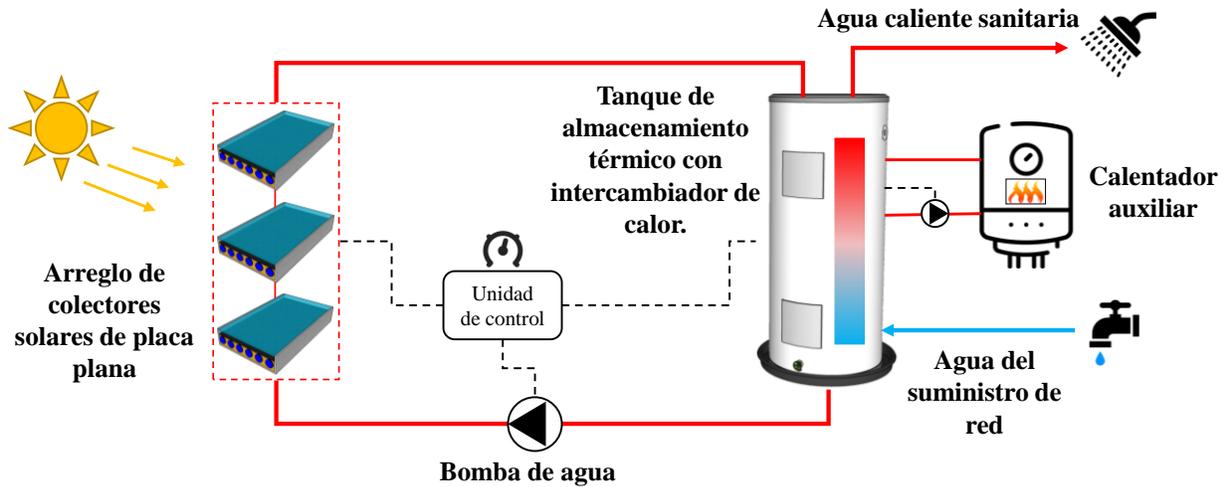


Figura 1. Diagrama esquemático del sistema solar considerado en el presente estudio.

Los componentes principales del colector consisten en una cubierta de vidrio transparente, una placa absorbidora que está conectada a seis tubos caloportadores de fluido (agua), así como un aislante térmico para minimizar las pérdidas de calor (Cetina-Quiñones et al., 2021). De igual manera, el arreglo de colectores se encuentra conectado a un tanque de almacenamiento térmico que integra un intercambiador de calor. Además, se conecta a un calentador auxiliar que funge como fuente de energía adicional para satisfacer la demanda de calor durante períodos de ausencia de la radiación solar (Company, 2008). El sistema opera nueve horas diario entre dos periodos, de seis a diez horas y de dieciocho a veintitrés horas, que son los lapsos de demanda de agua caliente sanitaria (Lévesque et al., 2004). Finalmente, el sistema incorpora una unidad de control que regula el flujo de masa, la temperatura de retroalimentación y las bombas de agua para garantizar una distribución uniforme de masa y una reducción de los requerimientos de energía.

## 2.2. Evaluación energética del sistema

Para la evaluación energética del sistema, se utilizó el método de la gráfica  $f$  ( $f$ -chart) por sus siglas en inglés, el cual permite el dimensionamiento de cualquier sistema de colectores solares considerando ciertas condiciones iniciales, y permitiendo identificar el comportamiento del sistema térmico y la cantidad de energía necesaria para cualquier aplicación térmica (Duffie et al., 2020). El método  $f$ -chart requiere el uso de datos mensuales promedio para determinar el rendimiento de la cobertura solar en las instalaciones de calentamiento que utilizan colectores solares planos. Este método se basa en la identificación de variables, a las que se asigna un valor, y la realización de una prueba a la instalación para conocer su rendimiento bajo estas condiciones. De tal manera, la Ecuación 1 representa la cobertura solar mensual obtenida con el método  $f$ -chart (Rosli et al., 2019).

$$f = 1.029Y - 0.065X - 0.245Y^2 + 0.0018X^2 + 0.0215Y^3 \quad (1)$$

Donde  $Y$  es la relación del calor absorbido por el sistema respecto a la demanda de calor y  $X$  la relación de las pérdidas del sistema respecto a

la demanda de calor, las cuales se obtienen mediante las Ecuaciones 2 y 3, respectivamente.

$$Y = \frac{\text{calor absorbido por el sistema}}{\text{demanda de calor}} = \frac{Q_{abs}}{Q_{ACS}} \quad (2)$$

$$X = \frac{\text{pérdidas de calor por el sistema}}{\text{demanda de calor}} = \frac{Q_L}{Q_{ACS}} \quad (3)$$

De las Ecuaciones 2 y 3,  $Q_{ACS}$ ,  $Q_{abs}$  y  $Q_L$ , se obtienen con las Ecuaciones 4 – 6.

$$Q_{ACS} = c_f \cdot C_{ACS} \cdot (T_{ACS} - T_{red}) \cdot n \quad (4)$$

$$Q_{abs} = F_R(\tau\alpha)_n \cdot \frac{F'_R}{F_R} \cdot \frac{(\tau\alpha)}{(\tau\alpha)_n} \cdot H \cdot n \cdot A_c \quad (5)$$

$$Q_L = F_R U_L \cdot \frac{F'_R}{F_R} \cdot (100 - \bar{T}_{amb}) \cdot \Delta t \cdot K_1 \cdot K_2 \cdot A_c \quad (6)$$

donde  $c_f$  es el calor específico del agua,  $C_{ACS}$  es la demanda en litros de agua,  $T_{ACS}$  es la temperatura de agua caliente sanitaria, es decir, la temperatura a la que se desea llegar el agua para la aplicación correspondiente,  $T_{red}$  es la temperatura del agua del suministro de la red,  $n$  el número de días del mes,  $F_R(\tau\alpha)_n$  es el factor de eficiencia óptica del colector el cual corresponde a la ordenada en el origen de la curva característica del colector,  $\frac{F'_R}{F_R}$  es un factor de corrección del conjunto colector-intercambiador (se recomienda tomar el valor de 0.95),  $\frac{(\tau\alpha)}{(\tau\alpha)_n}$  es el modificador del ángulo de incidencia (0.96 para colectores con cubiertas

de vidrio sencillas),  $H$  es la irradiación solar diaria mensual incidente,  $A_c$  es el área del arreglo de colectores,  $F_R U_L$  es la pendiente de la curva característica del colector,  $\bar{T}_{amb}$  es la temperatura media mensual del ambiente durante las horas de Sol,  $\Delta t$  es la cantidad de segundos que contiene un mes, y que  $K_1$  y  $K_2$  son factores de corrección del sistema de almacenamiento de calor.

De las Ecuaciones 5 y 6,  $F_R(\tau\alpha)_n$  se consideró como 0.79 y  $F_R U_L$  como 5.051, de acuerdo con la ficha técnica del colector utilizado como caso de estudio (BOSCH, 2023), mientras que  $K_1$  y  $K_2$  se calcula mediante las Ecuaciones 7 y 8, respectivamente (Formacion et al., 2009).

$$K_1 = \left( \frac{\text{kg acumulacion}}{75 \cdot A_c} \right)^{-0.25} \quad (7)$$

$$K_2 = \frac{11.6 + 1.18T_{ACS} + 3.86T_{red} - 2.32\bar{T}_{amb}}{100 - T_{amb}} \quad (8)$$

Las ecuaciones anteriores fueron utilizadas para crear una base de datos, considerando siete variables de entrada (mes, número de colectores, irradiación solar, temperatura ambiente, velocidad del viento, temperatura de ACS y consumo de ACS), y una variable de salida (cobertura solar mensual). Cabe

mencionar que la evaluación energética para la creación de la base de datos del presente trabajo se llevó a cabo bajo las condiciones ambientales de la ciudad de Mérida, Yucatán, como caso particular de estudio. El rango de las variables de entrada se describe en la Tabla 1.

Tabla 1. Rango de variables de entrada para la creación de la base de datos.

Variable	Rango	Referencia
Mes	(1 – 12)	-
Número de colectores	(1 – 25)	(Cetina-Quiñones et al., 2021)
Irradiación solar	(16,370,220 – 24,301,308)	(Meteonorm, 2024)
Temperatura ambiente	(24.02 – 28.43)	(Meteonorm, 2024)
Velocidad del viento	(0.29 – 0.52)	(Meteonorm, 2024)
Temperatura de ACS	(40 – 60)	(Buker y Ayompe, 2015)
Consumo de ACS	(100 – 4000)	(Casabianca y Aires, 2012)

Posteriormente, la base de datos generada fue usada para entrenar tres modelos de aprendizaje automático para obtener un sustituto del modelo matemático convencional: Redes Neuronales Artificiales (RNA), Árboles de Regresión (AR) y Máquinas de Vectores de Soporte (MVS), lo cual se explica con mayor detalle en la sección subsecuente.

### 2.3. Modelo surrogado basado en algoritmos de aprendizaje automático

En esta sección, se proporciona una explicación del desarrollo del modelo surrogado basado en tres algoritmos de aprendizaje automático, los cuales fueron redes neuronales artificiales, árboles de regresión y máquinas de vectores de soporte. Esta selección se debió a la robustez,

información integral, uso y fácil implementación de cada algoritmo en el entorno de programación MATLAB, gracias a las bibliotecas y paquetes necesarios para la ejecución de cada uno. Además, esos algoritmos se han utilizado frecuentemente en trabajos de regresión en sistemas energéticos. La incorporación de estos modelos se debe a su capacidad natural de abordar problemas complejos que involucran una variedad de entradas y salidas, superando las limitaciones de los modelos analíticos convencionales (He y Xu, 2010). La Figura 2 muestra un diagrama esquemático de las variables de entrada y de salida, así como de los algoritmos de aprendizaje automático empleados en este estudio.

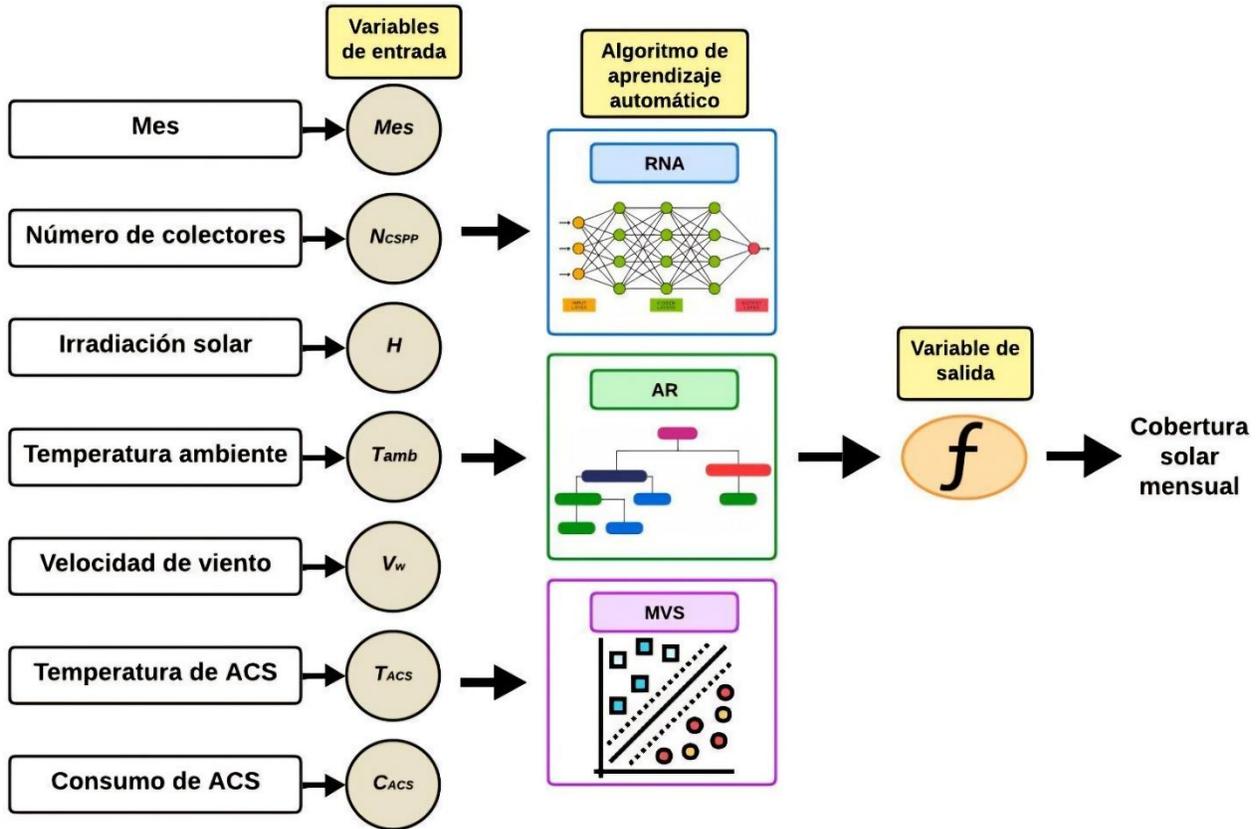


Figura 2. Diagrama esquemático del modelo surrogado empleando tres algoritmos de aprendizaje automático.

La Figura 2 ilustra el modelo con siete variables de entrada correspondientes al mes, número de colectores, irradiación solar, temperatura ambiente, velocidad del viento, temperatura de ACS y consumo de ACS. En contraste, la variable de salida corresponde a la cobertura solar mensual. El proceso empleado durante el entrenamiento se llevó de manera

iterativa empleando los algoritmos de aprendizaje automático previamente mencionados. En total se utilizaron 63,050 datos por cada variable. Además, para facilitar una convergencia de aprendizaje más rápida, se normalizaron todas las variables entre 0 y 1 utilizando la Ecuación 9 (Cardoso-Fernández et al., 2023).

$$x_n = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (9)$$

donde  $x_n$  representa la variable normalizada,  $\min(x)$  es el valor mínimo,  $\max(x)$  el valor máximo y  $x$  el valor actual.

### 2.3.1. Redes Neuronales Artificiales

Las RNA se basan en la estructura neural de la corteza cerebral humana, pero en tamaños mucho más pequeños (Chakraverty y Mall, 2017). Se interpretan como unidades de procesamiento están organizadas en múltiples capas (capas de entrada, ocultas y de salida), con enlaces que conectan las unidades de procesamiento en diferentes niveles. Los pesos

de la capa oculta representan las características de los patrones de entrada, y la capa de salida produce el patrón de salida de la red basado en los pesos de la capa oculta (Mohandes et al., 2019). Cada unidad de procesamiento, por ejemplo,  $l$  conlleva entradas  $x_{il} (i = 1, 2, \dots, m)$  de unidades vinculadas a ella en las otras capas. Estas entradas están asociadas con conexiones con un peso  $w_{il} (i = 1, 2, \dots, m)$ . La unidad de procesamiento luego realiza dos operaciones fundamentales en las entradas: suma y

activación, y produce una salida  $y_k$  (Wei et al., 2018).

La Ecuación 10 presenta la función tangente sigmoidea (*tansig*), la cual fue empleada como

$$y_k = \sum_{n=1}^N \left[ LW_{k,n} \left( \frac{2}{1 + \exp[-2(\sum_{m=1}^M (IW_{n,m}x_m) + b_{1,n})]} - 1 \right) \right] + b_{2,k} \quad (10)$$

A partir de la Ecuación 10,  $y_k$ , representa la  $k$  salida de la RNA, es decir, el parámetro  $f$ .  $M$  y  $N$  representan el número total de neuronas en la capa de entrada y capa oculta, respectivamente.  $IW$ ,  $LW$ ,  $b_1$  y  $b_2$  corresponden a las matrices y vectores de los pesos asociados con las conexiones en la capa de entrada y capa oculta, las conexiones entre la capa de oculta y la capa de salida, el bias de la capa oculta y el bias asociado con la neurona de salida, en el orden dado. Los subíndices  $m$ ,  $n$  y  $k$  indican las neuronas de entrada, oculta y salida, respectivamente. El entrenamiento de la RNA se realizó utilizando el algoritmo de optimización Levenberg-Marquardt, que combina las ventajas del método de gradiente descendente y el método de Gauss-Newton. Este algoritmo fue seleccionado por su eficiencia en problemas de regresión no lineales y su rápida convergencia. Como función de costo se empleó el error cuadrático medio (MSE), que mide la diferencia promedio al cuadrado entre las predicciones de la red y los valores objetivo.

### 2.3.2. Árboles de Regresión

Un Árbol de Regresión (AR) representa un instrumento de aprendizaje automático que se destaca por ofrecer predicciones precisas y, al mismo tiempo, ofrecer una interpretación sencilla. Se genera mediante un procedimiento repetitivo que segmenta cada nodo en nodos subordinados según criterios definidos. En consecuencia, se ajusta con precisión un modelo de regresión para cada nodo de hoja a fin de pronosticar los valores de las variables de salida para las nuevas muestras (Yang et al., 2017). El propósito es desarrollar un modelo para predecir con precisión el valor de una

función de activación debido a su adaptación eficiente a procesos complejos no-lineales (Singh Chauhan et al., 2015).

variable objetivo basado en reglas de decisión simples construidas a partir de las propiedades de los datos (Yu et al., 2010). Como resultado, un modelo de AR se asemeja a una red con un nodo raíz y varios nodos de ramificación. Además, se debe considerar la profundidad de un árbol de decisión, ya que un árbol más profundo tiene una gran variedad y bajo sesgo, mientras que un árbol más corto tiene un alto sesgo y baja varianza (Gupta et al., 2021).

Para la optimización del modelo AR se utilizó la herramienta de ajuste de hiperparámetros de MATLAB, considerando un tamaño de hoja variable, lo que permite al modelo encontrar todas las posibles divisiones alternativas para cada nodo. Este enfoque de optimización exhaustiva garantiza que el árbol encuentre la mejor estructura posible, lo que mejora la robustez del modelo y su capacidad para manejar datos faltantes o ruidosos. De igual manera, se realizó una evaluación cruzada de estos hiperparámetros para garantizar la generalización del modelo y prevenir el sobreajuste.

### 2.3.3. Máquinas de Vectores de Soporte

Las Máquinas de Vectores de Soporte (MVS) representan una técnica de aprendizaje supervisado versátil, capaz de abordar tanto problemas de regresión como de clasificación (Awad y Khanna, 2015). Para problemas de regresión, como el presente estudio, las MVS buscan encontrar una función que mejor se ajuste a los datos minimizando el error de predicción mientras mantiene la complejidad del modelo bajo control. Este equilibrio se logra mediante la definición de un margen  $\epsilon$  (épsilon) que establece una zona de tolerancia alrededor de la función de regresión, conocida

como el tubo  $\epsilon$ -insensible (Cardoso-Fernández et al., 2023). La técnica opera proyectando los datos de entrada en un espacio de características de mayor dimensión mediante funciones kernel, donde la relación no lineal original puede tratarse como lineal (Pisner y Schnyer, 2020). Una característica distintiva de

las MVS es que la solución depende únicamente de un subconjunto de los puntos de entrenamiento, denominados Vectores de Soporte (SV), que son los que definen el margen del tubo  $\epsilon$ -insensible (Scholkopf y Smola, 2018). Las MVS se representan mediante la Ecuación 11.

$$f(\mathbf{x}, a, a^*) = \sum_{i=1}^s (a_i - a_i^*)k(\mathbf{x}_i, \mathbf{x}) + b \quad (11)$$

donde  $a, a^*$  corresponden a la variable dual para las restricciones activas,  $k$  es la función kernel, donde se utilizan comúnmente tipos lineales, polinómicos y gaussianos, y  $b$  es el sesgo de la función de regresión (Santamaría-Bonfil et al., 2016).

En la implementación del modelo MVS para este trabajo, se utilizó un kernel lineal debido a su eficiencia computacional y buenos resultados en problemas de regresión similares (Mangasarian and Musicant, 2002). La función de pérdida empleada fue  $\epsilon$ -insensible con un valor  $\epsilon = 0.1$ , que permite una zona de tolerancia alrededor de la predicción. El parámetro de regularización  $C$  se estableció en 1.0 para mantener un equilibrio entre el ajuste a los datos y la complejidad del modelo. Estos hiperparámetros fueron evaluados con validación cruzada para asegurar la generalización del modelo y evitar el sobreajuste.

Como se mencionó previamente, el entorno de programación de MATLAB fue utilizado para entrenar los diferentes algoritmos de aprendizaje automática. Se utilizaron los módulos “*neural net fitting*” y “*regression learner*” (MathWorks, 2023). La base de datos fue particionada en tres secciones, de las cuales el 70 % de los datos fue empleado para el entrenamiento, 15 % para la comprobación y 15 % para la validación. Se utilizaron métodos iterativos para encontrar los parámetros que optimizan cada algoritmo, mediante un proceso de ajuste automático (*hyperparameter tuning*) que maximiza el rendimiento de los modelos en términos de exactitud y garantiza el mejor desempeño predictivo de cada modelo. Finalmente, se utilizaron diversas métricas estadísticas, incluido el Coeficiente de Determinación ( $R^2$ ), el Error Cuadrático Medio (ECM), y Error Medio Absoluto (EMA), para evaluar la exactitud y fiabilidad del modelo de los algoritmos. Estas métricas fueron estimadas usando las Ecuaciones 12 – 14.

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \text{prom}(x_i))^2} \quad (12)$$

$$ECM = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (13)$$

$$EMA = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (14)$$

donde  $x_i$  denota el valor real, el término “*prom*” se refiere al valor promedio,  $y_i$  es el valor predicho y  $n$  indica el total de datos.

La selección de estas métricas de desempeño se realizó considerando su complementariedad

para evaluar el modelo. El coeficiente de determinación ( $R^2$ ) indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes, donde valores más cercanos a

1 indican mejor ajuste del modelo. Por otra parte, el error cuadrático medio (ECM) penaliza errores grandes y proporciona una medida de la magnitud promedio de los errores de predicción en las mismas unidades que la variable objetivo y finalmente el error medio absoluto (EMA) proporciona una medida de la precisión del modelo que es menos sensible a valores atípicos que el ECM. La combinación de estas tres métricas permite una evaluación integral del desempeño del modelo, considerando tanto su capacidad de ajuste ( $R^2$ ) como la magnitud y distribución de sus errores (ECM y EMA).

#### 2.4. Análisis de sensibilidad basado en el algoritmo de Garson

El análisis de sensibilidad permite evaluar la importancia relativa de los parámetros de entrada en el modelo surrogado con respecto a la variable de salida del sistema. Como parte de esta investigación, se seleccionó el algoritmo de Garson para el análisis de sensibilidad debido a su especial adaptación para redes neuronales artificiales, las cuales demostraron el mejor desempeño entre los algoritmos evaluados. El algoritmo de Garson está diseñado para cuantificar el porcentaje de influencia relativa de cada variable de entrada considerando la división de los pesos de

conexión de la red neuronal artificial (Ghanizadeh et al., 2020).

Estas conexiones de peso se pueden estimar empleando la Ecuación 14, donde  $IR_j$  representa la importancia relativa de cada variable de entrada  $j$  en el modelo predictivo. La ecuación considera los valores absolutos de los pesos de conexión en diferentes niveles de la red: entre las neuronas de entrada y las ocultas ( $IW_{j,n}$ ), así como entre las neuronas ocultas y las de salida ( $LW_{n,k}$ ). El denominador  $\sum |IW_{m,n}|$  actúa como factor de normalización, representando la suma de todos los pesos de conexión entrantes a una neurona oculta específica.

La ecuación realiza un proceso de normalización y ponderación donde  $M$  representa el número total de variables de entrada y  $N$  el número de neuronas en la capa oculta. Este proceso asegura que la suma total de las importancias relativas de todas las variables de entrada sume 100%, facilitando la interpretación de los resultados como porcentajes de influencia. El algoritmo considera tanto las conexiones directas como las interacciones entre variables a través de las capas de la red, proporcionando una medida robusta de la importancia de cada variable en el modelo.

$$IR_j = \frac{\sum_1^N \left[ \left( \frac{|IW_{j,n}|}{\sum_1^M |IW_{m,n}|} \right) \times |LW_{n,k}| \right]}{\sum_1^N \left[ \sum_1^N \left( \frac{|IW_{j,n}|}{\sum_1^M |IW_{m,n}|} \right) \times |LW_{n,k}| \right]} \quad (14)$$

### 3. Resultados y discusión

#### 3.1. Validación del modelo

El modelo desarrollado con base en el método f-chart para este trabajo fue validado con los

resultados experimentales reportados por (Okafor y Akubue, 2012). La Figura 3 reporta la comparación de los resultados obtenidos con el presente estudio y la comparación con los resultados reportados por los autores previos.

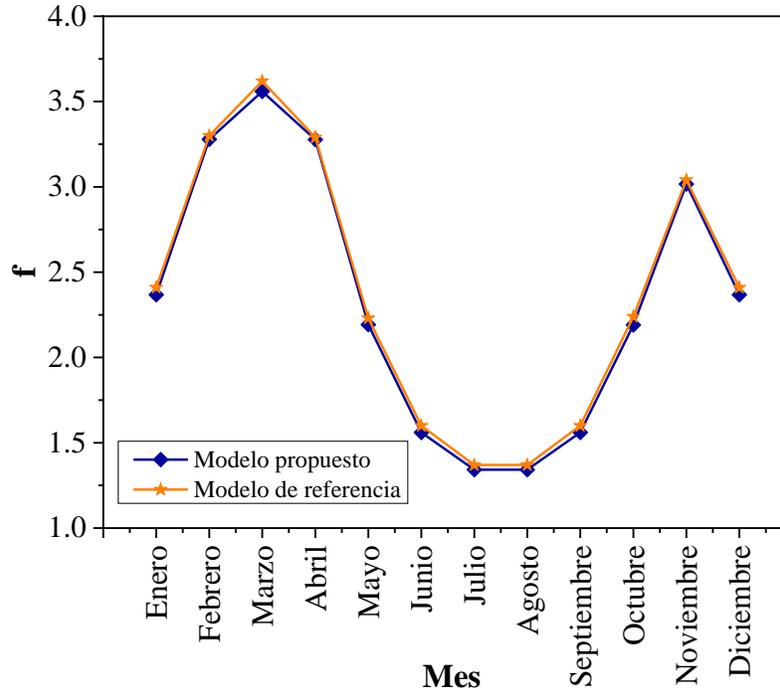


Figura 3. Validación del modelo propuesto considerando la cobertura solar mensual respecto al modelo de referencia.

Se puede observar de la Figura 3 que los resultados obtenidos utilizando el modelo actual se encuentran estrechamente en concordancia con los resultados de referencia reportados por (Okafor y Akubue, 2012). El valor más alto de  $f$  fue de 3.62 para el valor de referencia y de 3.56 para este estudio, correspondiente al mes de marzo. La máxima diferencia obtenida entre ambos resultados fue de 0.02112 alcanzada durante el mes de febrero. Por otra parte, los valores de ECM y EMA correspondientes a 0.0375 y 0.0353 subrayan la exactitud de las predicciones, indicando desviaciones mínimas entre la referencia y los valores obtenidos con el modelo presente. Finalmente, el alto valor del coeficiente de determinación (0.9976) demuestra que el modelo propuesto explica una proporción significativa de la variabilidad en los datos de temperatura interior, con un valor cercano a 1, lo que indica una ligera

inconsistencia aceptable con los datos de referencia que para fines prácticos se pueden considerar despreciables.

### 3.2. Modelo surrogado basado en algoritmos de aprendizaje automático

La Figura 4 presenta los resultados comparativos del desempeño de los tres algoritmos de aprendizaje automático implementados: Redes Neuronales Artificiales (RNA), Árboles de Regresión (AR) y Máquinas de Vectores de Soporte (MVS). Para cada algoritmo se muestran las gráficas de correlación (a), respuesta (b) y error absoluto (c) entre los valores predichos y actuales. La correlación se midió utilizando la Ecuación 12 descrita previamente, mientras que la respuesta representa la comparación directa entre los valores predichos y actuales en función del número de datos tal como  $f = f(x)$  donde  $f$  es el valor de la cobertura solar (tanto predicho como actual) y  $x$  representa el número de dato. Finalmente, el error absoluto se calculó mediante la Ecuación 15.

$$Error\ absoluto = |x_i - y_i| \tag{15}$$

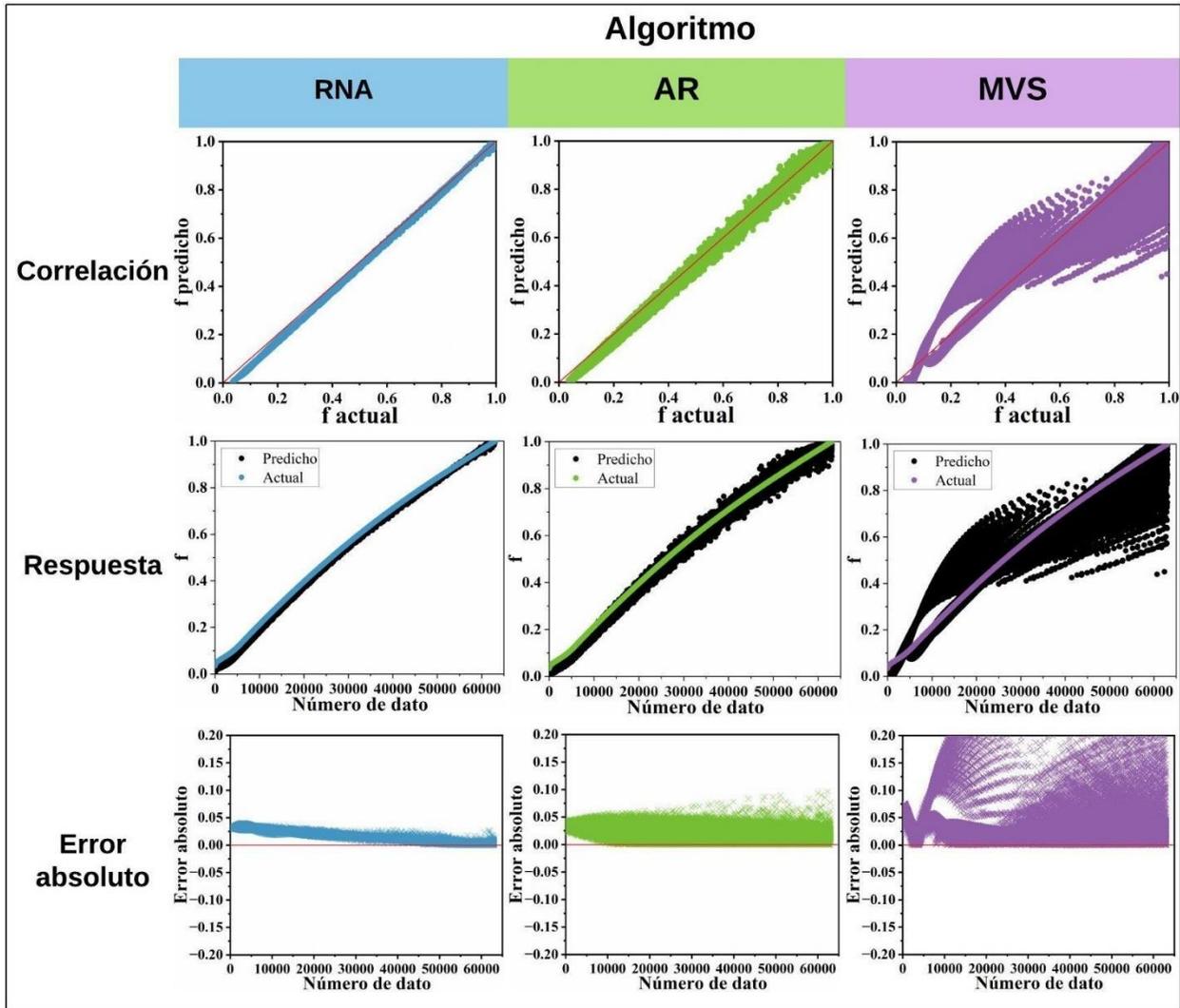


Figura 4. Gráfica de los resultados de entrenamiento de los modelos RNA, AR, MVS correspondiente a: (a) correlación entre valores actuales y predichos, (b) gráfica de respuesta y (c) verificación del error absoluto.

Como se puede observar en la Figura 4, los tres algoritmos mostraron diferentes niveles de desempeño en la predicción de la cobertura solar mensual. En el caso de las Redes Neuronales Artificiales, durante el proceso iterativo, el algoritmo de Levenberg-Marquardt con 10 neuronas en la capa oculta y función de activación *tansig* en la capa de salida fue la configuración óptima, obteniendo valores de  $R^2 = 0.9956$ ,  $ECM = 0.0188$  y  $EMA = 0.0159$ , lo que indica una correlación altamente fiable. Para el algoritmo de Árboles de Regresión, el proceso de optimización determinó que la mejor estructura fue un árbol con profundidad de dos niveles y dos hojas terminales, obteniendo valores de  $R^2 = 0.9931$ ,  $ECM = 0.0545$  y  $EMA = 1.9827$ . Finalmente, para el

algoritmo de Máquinas de Vectores de Soporte, la optimización reveló que la mejor configuración se logró utilizando un kernel lineal con función de pérdida cuadrática, obteniendo valores de  $R^2 = 0.9599$ ,  $ECM = 3.1936$  y  $EMA = 3.6712$ .

Comparando los resultados de los tres algoritmos, se observa que las Redes Neuronales Artificiales proporcionaron el mejor desempeño general, con los valores más altos de  $R^2$  y los errores más bajos ( $ECM$  y  $EMA$ ). Aunque los Árboles de Regresión también mostraron un buen desempeño con un  $R^2$  cercano al de las RNA, sus errores fueron ligeramente superiores. Por su parte, las Máquinas de Vectores de Soporte, a pesar de

utilizar un kernel lineal que generalmente es eficiente para problemas de regresión, mostraron el desempeño más ineficiente los tres algoritmos, con errores significativamente mayores.

Por otra parte, analizando las gráficas de correlación (a), se puede apreciar que el modelo RNA exhibe la relación más lineal y consistente entre valores predichos y actuales, seguido por el modelo AR, mientras que MVS muestra una mayor dispersión en sus predicciones. Las gráficas de respuesta (b) confirman este comportamiento, donde la RNA mantiene la mayor coincidencia entre valores predichos y actuales a lo largo de toda la serie de datos, mientras que AR y MVS presentan mayores desviaciones, particularmente MVS en los valores más altos del rango.

En cuanto al error absoluto (c), la RNA mantiene la menor variación, con errores consistentemente alrededor de 0.05, mientras que AR presenta una dispersión ligeramente mayor, y MVS muestra la mayor variabilidad en el error. Estos resultados demuestran que, entre los tres algoritmos evaluados, la Red Neuronal Artificial desarrollada proporciona el modelo más confiable y preciso para predecir la cobertura solar mensual del sistema de producción de agua caliente sanitaria bajo las condiciones ambientales y operativas establecidas.

### *3.3. Análisis de sensibilidad*

En la presente sección se describe el porcentaje relativo de influencia de cada variable de entrada sobre el parámetro de salida del modelo, como producto del análisis de sensibilidad empleando el algoritmo de Garson y definido por la función matemática resultante del mejor modelo obtenido correspondiente a RNA previamente entrenado y validado.

Los resultados obtenidos revelaron que el número de colectores representó la variable más influyente, con un impacto significativo del 44 %. Este resultado se debe a la naturaleza

directa y proporcional entre la cantidad de colectores instalados y la capacidad global del sistema para capturar y aprovechar la radiación solar. Es decir, un mayor número de colectores implica una mayor área de captura, resultando en una mejora sustancial en la eficiencia global del sistema y por ende en mayor cobertura solar. De igual manera, el consumo de ACS reveló el segundo valor más alto de influencia, después del número de colectores, lo que representa una influencia significativa en el rendimiento de los sistemas solares térmicos debido a su impacto en la demanda de energía térmica y su relación intrínseca con la operación del sistema. Por otra parte, otras variables como la temperatura del ACS, irradiación solar y temperatura ambiente, aunque también influyentes, mostraron una variación relativa en su impacto, con valores de 17.9%, 3.8% y 2.8%, respectivamente. Finalmente, la velocidad del viento (0.16%) y el mes (0.033%) revelaron las influencias más bajas, destacando una importancia no significativa de estos parámetros en comparación con los demás.

En conclusión, los resultados del análisis de sensibilidad destacan el predominio del número de colectores como la variable determinante en la cobertura solar mensual de los sistemas solares térmicos en entornos residenciales, lo cual resalta la necesidad de una cuidadosa consideración y optimización de la cantidad de colectores instalados para lograr un diseño eficiente y sostenible de estos sistemas en aplicaciones domésticas para producción de agua caliente sanitaria.

## **Conclusiones**

Este trabajo presentó el desarrollo de un modelo surrogado basado tres algoritmos de aprendizaje automático tales como arboles de regresión, máquinas de vectores de soporte y redes neuronales artificiales y un análisis de sensibilidad de un sistema de colectores solares de placa plana. El modelo surrogado fue entrenado a partir de una base de datos creada mediante la evaluación energética del sistema

de colectores empleando el método f-chart. Además, siete variables de entrada fueron consideradas en el modelo surrogado correspondientes al mes, número de colectores, irradiación solar, temperatura ambiente, velocidad del viento, temperatura de ACS y consumo de ACS, así como la variable de salida objetivo correspondiente a la cobertura solar mensual. El análisis de sensibilidad fue desarrollado con el algoritmo de Garson.

Los resultados obtenidos revelaron que el modelo surrogado basado en RNA predijo eficazmente los valores de la cobertura solar mensual, obteniendo un coeficiente de determinación de 0.9956, un error cuadrático medio de 0.0188 y un error medio absoluto de 0.0159, indicando una correlación fiable para predecir la variable de salida del sistema. Por otra parte, el análisis de sensibilidad reveló que el número de colectores representó la variable más influyente sobre la cobertura solar mensual, con un porcentaje de influencia de 44 %, destacando la relación directa entre la cantidad de colectores instalados y el área de aprovechamiento de la radiación solar para el aumento de la cobertura solar.

Finalmente, los hallazgos de este estudio permiten destacar la importancia de los modelos de inteligencia artificial basados en algoritmos de aprendizaje automático como una herramienta útil y alternativa de los modelos convencionales de estos sistemas energéticos. Además, la integración con el análisis de sensibilidad permite el mejor

control de las variables de entrada del modelo sobre las variables de salida, enfatizando la necesidad de una cuidadosa consideración y optimización de los parámetros de entrada para el mejor aprovechamiento de los sistemas energéticos.

Como trabajo a futuro, el modelo desarrollado en este estudio puede emplearse en el análisis de sistemas de colectores solares de placa plana evaluándolos en diferentes tipos de climas y bajo diferentes condiciones de operación, lo que permitiría optimizar su rendimiento y eficiencia e identificar posibles mejoras en el diseño de estos sistemas para contribuir a la transición hacia fuentes de energía más sostenibles y limpias.

### Agradecimientos

Los autores A.J. Cetina-Quiñones, Diana Romero, y A.J. Hernández Bautista, agradecen al CONAHCYT por el apoyo financiero concedido durante sus estudios de posgrado en la opción de Energías Renovables en la Facultad de Ingeniería, Universidad Autónoma de Yucatán con CVU 861995 (A.J. Cetina-Quiñones), 1171844 (Diana Romero), y 1101214 (A.J. Hernández Bautista).

El autor E. Chan-Dzib agradece al CONAHCYT por el apoyo financiero concedido durante sus estudios de posgrado en la opción de Ciencias en Energía Renovable en el Centro de Investigación Científica de Yucatán con CVU 405462.

### Referencias

- Ahmad, M. W., Reynolds, J., and Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production*, 203, 810–821.
- Awad, M., and Khanna, R. (2015). Support vector regression. In *Efficient learning machines* (pp. 67–80). Springer.
- Borgonovo, E. (2023). Sensitivity analysis. *Tutorials in Operations Research: Advancing the Frontiers of OR/MS: From Methodologies to Applications*, 52–81.
- BOSCH. (2023). *Colector solar de placa plana SO1500TFV*.

- Buker, M. S., and Ayompe, L. (2015). A Comparison of Flat Plate and Evacuated Tube Solar Collectors with F-Chart Method. *Isi Bilimi Ve Teknigi Dergisi-Journal of Thermal Science and Technology*, 35(1), 77–86.
- Çapik, M., Yılmaz, A. O., and Çavuşoğlu, İ. (2012). Present situation and potential role of renewable energy in Turkey. *Renewable Energy*, 46, 1–13.
- Cardoso-Fernández, V., Bassam, A., May Tzuc, O., Barrera Ch., M. A., Chan-González, J. de J., Escalante Soberanis, M. A., Velázquez-Limón, N., and Ricalde, L. J. (2023). Global sensitivity analysis of a generator-absorber heat exchange (GAX) system's thermal performance with a hybrid energy source: An approach using artificial intelligence models. *Applied Thermal Engineering*, 218, 119363. <https://doi.org/10.1016/j.applthermaleng.2022.119363>
- Casabianca, G., and Aires, B. (2012). *Predimensionado simplificado de instalaciones solares de agua caliente sanitaria según los criterios y normativas básicas de arquitectura escolar*. 13–18.
- Cetina-Quiñones, A. J., Sanchez-Dominguez, I., Casillas-Reyes, A., and Bassam, A. (2023). 9E analysis of a flat plate solar collector system implementation: A new approach based on digital twin model coupled with global sensitivity analysis and multi-objective optimization. *Journal of Renewable and Sustainable Energy*, 15. <https://doi.org/10.1063/5.0151617>
- Cetina-Quiñones, A. J., Xamán, J., Bassam, A., Escalante Soberanis, M. A., and Perez-Quintana, I. (2021). Thermo-economic analysis of a flat solar collector with a phase changing material under tropical climate conditions: Residential and industrial case. *Applied Thermal Engineering*, 182(March 2020). <https://doi.org/10.1016/j.applthermaleng.2020.116082>
- Chakraverty, S., and Mall, S. (2017). *Artificial neural networks for engineers and scientists: solving ordinary differential equations*. CRC Press.
- Company, E. M. (2008). Measurement of Domestic Hot Water Consumption in Dwellings. *Energy Saving Trust*, 62.
- Deepika, D., Baig, M. A. A., Reddy, A. R., and Maneaih, D. (2016). Utilization of f-Chart Method for Designing Solar Thermal Heating System. *IOSR Journal of Mechanical and Civil Engineering*, 16(053), 23–28. <https://doi.org/10.9790/1684-16053042328>
- Dehaj Shafiey, M., Zamani Mohiabadi, M., and Sadegh Hosseini, S. M. (2020). Prediction of the outlet flow temperature in a flat plate solar collector using artificial neural network. *Environ Monit Assess*.
- Duffie, J. A., Beckman, W. A., and Blair, N. (2020). *Solar engineering of thermal processes, photovoltaics and wind*. John Wiley & Sons.
- Formacion, B. V., Muñiz, J. M. M., and Garcia, R. C. (2009). *Energia Solar Termica. 2da. Edicion*.
- Ghanizadeh, A. R., Heidarabadzadeh, N., and Jalali, F. (2020). Artificial neural network back-calculation of flexible pavements with sensitivity analysis using Garson's and connection weights algorithms. *Innovative Infrastructure Solutions*, 5, 1–19.
- Ghritlahre, H. K., Chandrakar, P., and Ahmad, A. (2020). Application of ANN model to predict the performance of solar air heater using relevant input parameters. *Sustainable Energy Technologies and Assessments*, 40(June), 100764. <https://doi.org/10.1016/j.seta.2020.100764>

- Gopi, A., Sharma, P., Sudhakar, K., Ngui, W. K., Kirpichnikova, I., and Cuce, E. (2023). Weather Impact on Solar Farm Performance: A Comparative Analysis of Machine Learning Techniques. *Sustainability (Switzerland)*, 15(1). <https://doi.org/10.3390/su15010439>
- Gupta, A., Bansal, A., and Roy, K. (2021). Solar energy prediction using decision tree regressor. *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 489–495.
- Hachicha, A. A., Yousef, B. A. A., Said, Z., and Rodríguez, I. (2019). A review study on the modeling of high-temperature solar thermal collector systems. *Renewable and Sustainable Energy Reviews*, 112(May), 280–298. <https://doi.org/10.1016/j.rser.2019.05.056>
- He, X., and Xu, S. (2010). Artificial neural networks. *Process Neural Networks: Theory and Applications*, 20–42.
- Kaood, A., Abubakr, M., Al-Oran, O., and Hassan, M. A. (2021). Performance analysis and particle swarm optimization of molten salt-based nanofluids in parabolic trough concentrators. *Renewable Energy*, 177, 1045–1062. <https://doi.org/10.1016/j.renene.2021.06.049>
- Lévesque, B., Lavoie, M., and Joly, J. (2004). Residential water heater temperature. *The Canadian Journal of Infectious Diseases = Journal Canadien Des Maladies Infectieuses*, 15(1), 11–12. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2094925&tool=pmcentrez&rendertype=abstract>
- Liu, P., Wu, J., Chen, L., Liu, Z., and Liu, W. (2021). Numerical analysis and multi-objective optimization design of parabolic trough receiver with ribbed absorber tube. *Energy Reports*, 7, 7488–7503. <https://doi.org/10.1016/j.egy.2021.10.084>
- Mangasarian, O. L., and Musicant, D. R. (2002). Large scale kernel regression via linear programming. *Machine Learning*, 46, 255–269.
- MathWorks. (2023). *The Language of Technical Computing*. <https://www-eio.upc.es/lceio/manuals/matlab/techdoc/pdfdocs/getstart.pdf>
- Meteonorm. (2024). *Meteonorm: World irradiation data*. <https://meteonorm.com/en/>
- Mohandes, S. R., Zhang, X., and Mahdiyar, A. (2019). A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing*, 340, 55–75. <https://doi.org/10.1016/j.neucom.2019.02.040>
- Okafor, I. F., and Akubue, G. (2012). *F-Chart Method for Designing Solar Thermal Water Heating Systems*. 3(9).
- Pandey, K. M., and Chaurasiya, R. (2017). A review on analysis and development of solar flat plate collector. *Renewable and Sustainable Energy Reviews*, 67, 641–650. <https://doi.org/10.1016/j.rser.2016.09.078>
- Pisner, D. A., and Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101–121). Elsevier.
- Rocha, J. H. de O., Gonçalves, R. S., Marrero, A. I. P., and Martins, K. C. R. (2022). Análise de um aquecedor solar através do método F-Chart e modelagem estatística. *Brazilian Journal of Development*, 8(11), 73544–73566. <https://doi.org/10.34117/bjdv8n11-183>
- Rosli, M. A. M., Zaki, D. S. M., Rahman, F. A., Sepeai, S., Hamid, N. A., and Nawam, M. Z.

- (2019). F-chart method for design domestic hot water heating system in Ayer Keroh Melaka. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, 56(1), 59–67.
- Santamaría-Bonfil, G., Reyes-Ballesteros, A., and Gershenson, C. (2016). Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy*, 85, 790–809. <https://doi.org/10.1016/j.renene.2015.07.004>
- Scholkopf, B., and Smola, A. J. (2018). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Shen, C., Liu, F., Qiu, S., Liu, X., Yao, F., and Zhang, Y. (2021). Numerical study on the thermal performance of photovoltaic thermal (PV/T) collector with different parallel cooling channels. *Sustainable Energy Technologies and Assessments*, 45, 101101.
- Singh Chauhan, P., Kumar, A., and Tekasakul, P. (2015). Applications of software in solar drying systems: A review. *Renewable and Sustainable Energy Reviews*, 51, 1326–1337. <https://doi.org/10.1016/j.rser.2015.07.025>
- Sujith, K. (2013). *Design Of Active Solar Water Heating System Using F-Chart Method*. 2(7), 296–299.
- Teixeira-Pinto, A. (2023). *Regression and Classification Trees*. Bio Statistics Collaboration of Australia. [https://bookdown.org/tpinto\\_home/Beyond-Additivity/regression-and-classification-trees.html](https://bookdown.org/tpinto_home/Beyond-Additivity/regression-and-classification-trees.html)
- Tomy, A. M., Ahammed, N., Subathra, M. S. P., and Asirvatham, L. G. (2016). Analysing the performance of a flat plate solar collector with silver/water nanofluid using artificial neural network. *Procedia Computer Science*, 93, 33–40.
- Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M., and Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82(May 2017), 1027–1047. <https://doi.org/10.1016/j.rser.2017.09.108>
- Yang, L., Liu, S., Tsoka, S., and Papageorgiou, L. G. (2017). A regression tree approach using mathematical programming. *Expert Systems with Applications*, 78, 347–357.
- Yu, Z., Haghghat, F., Fung, B. C. M., and Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, 42(10), 1637–1646.
- Zayed, M. E., Kabeel, A. E., Shboul, B., Ashraf, W. M., Ghazy, M., Irshad, K., Rehman, S., and Zayed, A. A. A. (2023). Performance augmentation and machine learning-based modeling of wavy corrugated solar air collector embedded with thermal energy storage: Support vector machine combined with Monte Carlo simulation. *Journal of Energy Storage*, 74, 1–40. <https://doi.org/10.1016/j.est.2023.109533>