



Revista Electrónica de  
Tecnología, Educación y Ciencia  
ISSN: 2953-5654  
<http://retec.unsa.edu.ar>  
Universidad Nacional de Salta

**Aprendizaje Automático Aplicado: Predicciones para un  
problema de convección natural transitoria**

**Sonia Esteban<sup>1</sup>, Loraine Gimson<sup>2</sup>**

<sup>1</sup>Facultad de Ciencias Exactas – Universidad Nacional de Salta

<sup>2</sup>C.I.D.I.A. – Centro de Investigación y Desarrollo en Informática Aplicada  
[s.esteban593@gmail.com](mailto:s.esteban593@gmail.com) , [loraine@cidia.unsa.edu.ar](mailto:loraine@cidia.unsa.edu.ar)

**Revista Electrónica de Tecnología, Educación y Ciencia,  
Volumen 1, Número 3, pág. 32-45 jun, 2026. ISSN: 2953-5654**

Disponible en <http://retec.unsa.edu.ar/>

## Aprendizaje Automático Aplicado: Predicciones para un problema de convección natural transitoria

Sonia Esteban<sup>1</sup>, Loraine Gimson<sup>2</sup>

<sup>1</sup>Facultad de Ciencias Exactas – Universidad Nacional de Salta

<sup>2</sup>C.I.D.I.A. – Centro de Investigación y Desarrollo en Informática Aplicada  
[s.esteban593@gmail.com](mailto:s.esteban593@gmail.com) , [loraine@cidia.unsa.edu.ar](mailto:loraine@cidia.unsa.edu.ar)

**Resumen:** En este trabajo se aplica aprendizaje automático supervisado para predecir el flujo de calor promedio  $Q$  en cavidades bidimensionales con geometría tipo ático, sometidas a convección natural transitoria. A partir de datos obtenidos por simulación numérica con el software OpenFOAM®, se analiza el comportamiento de  $Q$  en función del salto de temperatura impuesto ( $\Delta T$ ) y el ángulo de inclinación de la pared activa ( $\gamma$ ). La base teórica se fundamenta en la resolución de ecuaciones diferenciales que modelan la transferencia de calor y dinámica de fluidos, con condiciones de borde variables. Para reducir el costo computacional de las simulaciones, se entrenaron modelos predictivos utilizando regresión lineal múltiple (RLM) y Random Forest (RF), evaluando su desempeño mediante el coeficiente de determinación  $R^2$ . El modelo RF mostró un ajuste superior ( $R^2 > 0,99$ ) y fue validado con nuevos casos simulados no incluidos en el entrenamiento, demostrando su capacidad de generalización. Este enfoque permite obtener estimaciones confiables del flujo de calor sin necesidad de resolver nuevamente el problema completo, lo cual resulta útil en aplicaciones como diseño térmico de recintos o análisis preliminar de sistemas pasivos.

**Palabras clave:** Aprendizaje Automático Supervisado, Modelo predictivo, Problema de convección natural transitoria.

### 1. Contexto

Un grupo de docentes-investigadores pertenecientes al instituto del INENCO (Instituto de Energía No Convencional, Universidad Nacional de Salta) junto con docente del departamento de Matemática realizaron la simulación numérica del problema de convección natural transitoria en una cavidad bidimensional trapezoidal, utilizando OpenFoam® [1]. La cavidad consiste de dos paredes verticales, una inferior horizontal y otra superior inclinada. El ángulo de inclinación ( $\gamma$ ) varía de  $0^\circ$  a  $30^\circ$ . La temperatura impuesta en la pared inferior es mayor que la superior, mientras que las paredes laterales se mantienen aisladas térmicamente. La diferencia de temperatura impuesta ( $\Delta T$ ) varía entre  $0^\circ\text{C}$  y  $30^\circ\text{C}$ . La figura 1 muestra las configuraciones estudiadas en diferentes trabajos. Los resultados obtenidos se pueden utilizar en el estudio del acondicionamiento térmico en áticos, performance de colectores solares, entre otras aplicaciones.

### 2. Problema a analizar

El tiempo de las simulaciones puede ser de varias horas dependiendo de la red de discretización utilizada, para cada ángulo de inclinación y salto de temperatura; por lo que contar con un modelo que permita obtener un valor estimado de  $Q$  sería muy provechoso para el equipo de investigación.

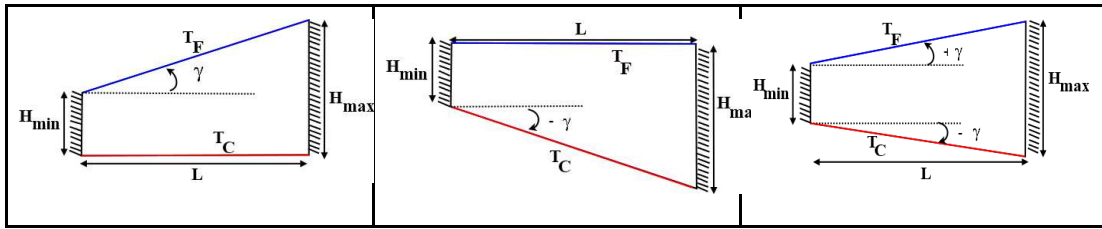


Figura 1: Tipos de geometrías

Para predecir el comportamiento del flujo de calor  $Q$  para un  $\Delta T$  y  $\gamma$  dados, se pretende aprovechar todos los datos que ya fueron simulados y recolectados en planillas de cálculo. Si bien se cuentan con datos que corresponden a simulaciones del problema de convección natural transitoria en recintos de distintas geometrías: ático, ático invertido y trapecio simétrico (figura 1), se circunscribe el presente trabajo a las geometrías tipo ático. Se tiene que para cada salto de temperatura ( $\Delta T = 1, 5, 10, \dots, 30$  °C) se varía el ángulo de inclinación ( $\gamma = 1, 2, 3, \dots, 30^\circ$ ).

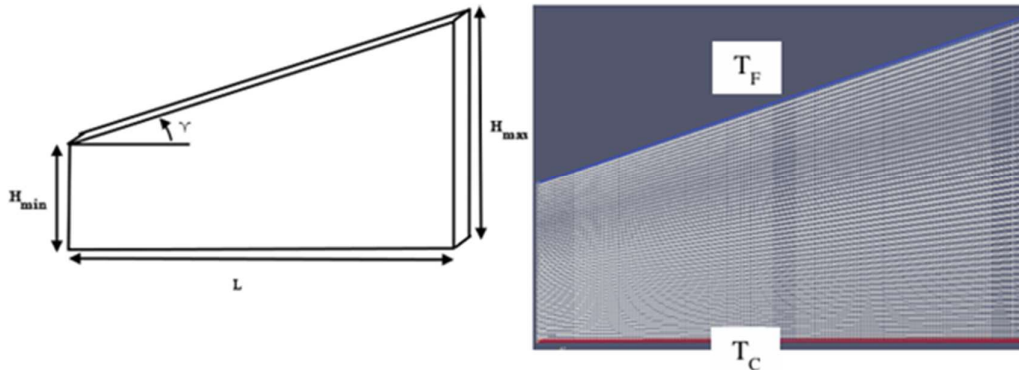


Figura 2: Tipo de geometría estudiada.

### Librerías a utilizar

Se trabaja con python y sus diferentes librerías:

- pyplot es un módulo dentro de la biblioteca Matplotlib en Python que proporciona una interfaz similar a la de MATLAB para crear gráficos de forma rápida y sencilla
- pandas utilizada para el análisis y manipulación de datos que permite trabajar con DataFrames
- Scikit-learn (sklearn) ofrece una amplia gama de herramientas para tareas de aprendizaje automático, incluyendo clasificación, regresión, clustering y reducción de dimensionalidad. Se utilizaron las siguientes clases de esta biblioteca: la división de las muestras en testeo y entrenamiento (train\_test\_split), para implementar el modelo de regresión lineal (LinearRegression), para trabajar con Random Forest (RandomForestRegressor), para poder obtener la mejor combinaciones de hiperparámetros para Random Forest (GridSearchCV) y para evaluar el rendimiento de un modelo de regresión lineal (r2\_score).

### 3. Trabajo inicial realizado

#### 3.1. Preparación de los datos

A través de simulaciones realizadas con el software OpenFOAM se obtiene el gradiente de temperatura ( $\frac{dT}{dx}$ ) sobre una de las paredes activas  $T_{flux}$  para un determinado  $\Delta T$  y  $\gamma$ . Los datos se guardan en una planilla de cálculo. Los resultados simulados para cada  $\Delta T$  se encuentran en diferentes hojas los valores de  $T_{flux}$  corresponden a un tiempo terminado (0 a 3600 segundos). Por lo tanto, lo primero es incluir una nueva columna que determine la geometría (estableciendo la convención 0 para ático, 1 para ático invertido).

Lo siguiente es tener toda la información en un único archivo .csv para poder realizar el procesamiento de datos. Para ello se genera, trabajando dentro de python con la librería pandas, un único archivo formato .csv que tenga la información del  $T_{flux}$  para todos los ángulos  $\gamma$  (0 a 30°) de cada  $\Delta T$  (0 a 30 °C, cada 5°C). Se agrega una nueva columna Q (flujo de calor) multiplicando el  $T_{flux}$  por el coeficiente 0,02 (conductividad térmica del aire). Trabajar con Q en lugar de  $T_{flux}$  directamente es útil porque Q representa la cantidad total de calor que se transfiere, mientras que  $T_{flux}$  es un gradiente. En tabla 1 se pueden ver los primeros valores del archivo que ya tiene los datos preparados para realizar el análisis.

geom	g	$\gamma$	$\Delta T$	$T_{flux}$
0	0	0	1	55.83
0	0	1	1	55.27
0	0	2	1	54.70
0	0	3	1	53.72

**Tabla 1.** Cabecera y primeros datos del archivo con el que se hizo procesamiento de datos.

#### 3.2. Análisis de valores obtenidos por simulación

Como se dijo anteriormente, se cuenta con los datos y gráficas de la evolución temporal de  $T_{flux}$ . En las figuras 3 a 6 se puede observar cómo varía Q a lo largo del tiempo para distintas inclinaciones, para un  $\Delta T = 10$  °C y  $\Delta T = 30$  °C.

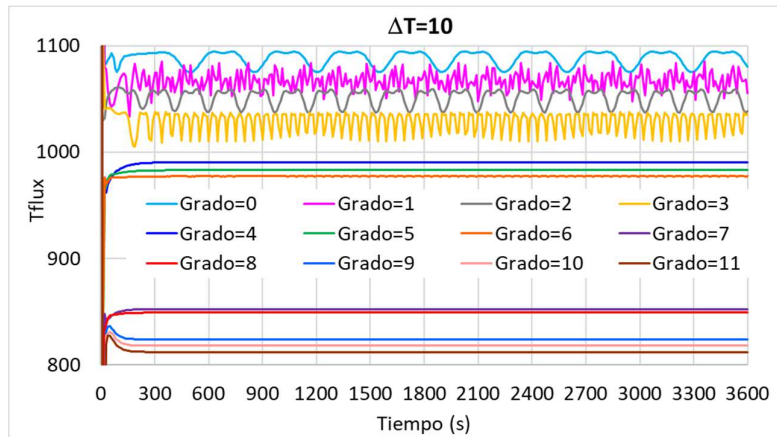


Figura 3: Evolución temporal de  $T_{flux}$  para distintas inclinaciones para  $\Delta T=10$  °C.

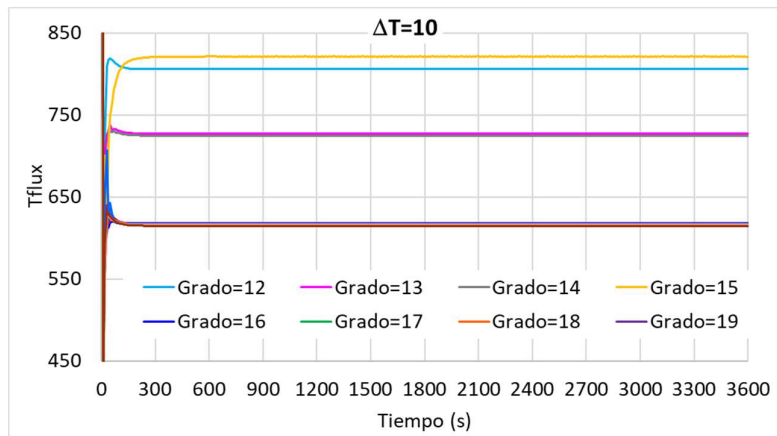


Figura 4: Evolución temporal de  $T_{flux}$  para distintas inclinaciones para  $\Delta T=10$  °C.

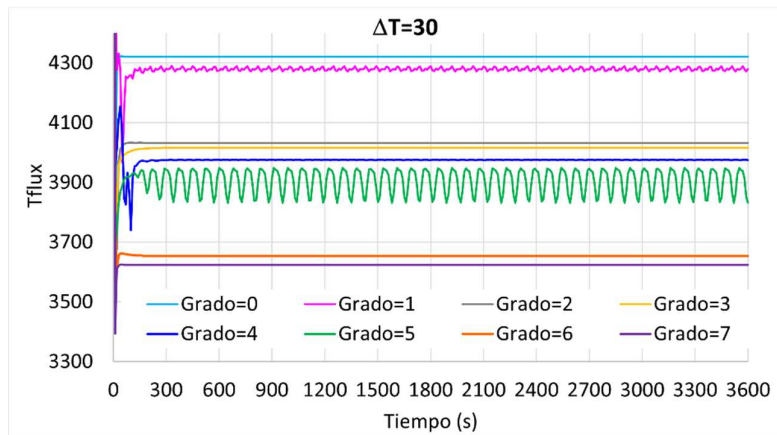
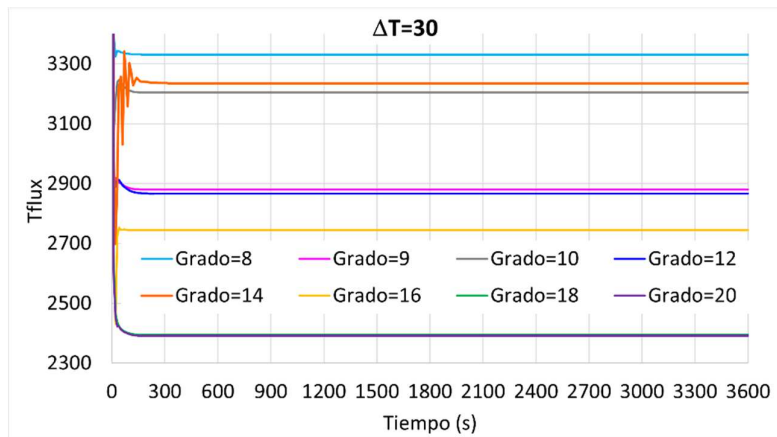


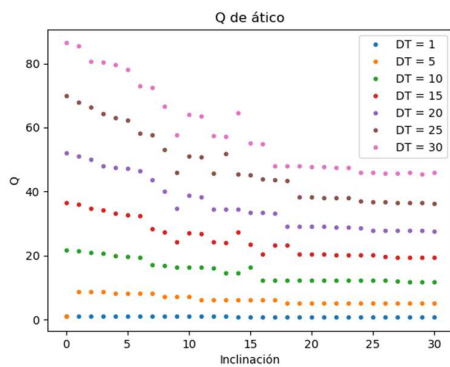
Figura 5: Evolución temporal de  $T_{flux}$  para distintas inclinaciones para  $\Delta T=30$  °C.



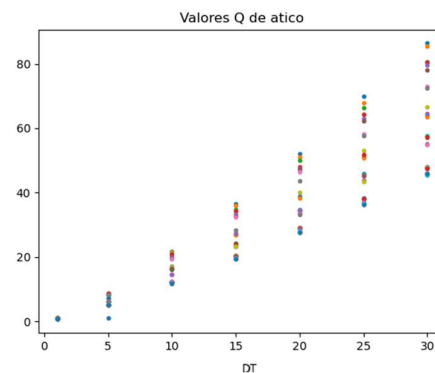
**Figura 6:** Evolución temporal de  $T_{flux}$  para distintas inclinaciones para  $\Delta T = 30\text{ }^{\circ}\text{C}$

Se puede observar en las gráficas anteriores que el  $T_{flux}$  se mantiene relativamente constante a través del tiempo, para la mayoría de las inclinaciones (figuras 4 y 6). En algunos casos se puede visualizar una variación de  $T_{flux}$ , se puede decir que la evolución temporal no alcanza el estado estacionario sino uno oscilatorio, casos grado=0 a grado=3 en figura 3 y grado=1 y grado=5 en figura 5. Este tipo de comportamientos también se observa en otros casos no mostrados en este trabajo. Se decide trabajar en este trabajo con los promedios temporales de Q, ya que no se pretende predecir el comportamiento de Q en tiempos posteriores.

El primer trabajo que se realiza con los datos es el ploteo con Python utilizando la librería pyplot, una de las gráficas obtenidas puede observarse en la Figura 4. Se grafica el Q promedio (de los 3600 segundos simulados) en función de la inclinación para cada uno de los  $\Delta T$  simulados (Figura 4.a). Se hace evidente que el  $\Delta T$  y el grado (inclinación) influyen en el valor promedio de Q obtenido por simulación. El flujo Q disminuye con el aumento de la inclinación, figura 4. a. Por otro lado, en la figura 4.b se grafica Q vs.  $\Delta T$ , para cada inclinación, se aprecia en este caso el aumento de Q con el aumento del salto de temperatura.



**Figura 4a:** Q vs.  $\gamma$  para distintos  $\Delta T$ .



**Figura 4b:** Q vs  $\Delta T$  y para distintos  $\gamma$ .

## 4. Pasos realizados para obtener un buen modelo predictivo del Q

En la teoría del *machine learning* (aprendizaje automatizado) existen muchos modelos que pueden aplicarse para poder realizar predicciones. La librería Scikit-learn (sklearn) permite configurar, entrenar y probar varios de ellos e incluso realiza el cálculo del  $R^2$  que mide cuán bien se ajusta el modelo a los datos observados.

El algoritmo de regresión más simple y uno de los más utilizados en el aprendizaje automatizado es la regresión lineal (simple - RLS- y múltiple - RLM-). Asume una relación lineal entre las variables independientes y dependientes.

Los árboles de decisión son más flexibles y pueden capturar relaciones complejas no lineales al subdividir en ramas el set de datos a entrenar.

Random forest -RF- (bosque aleatorio) es un conjunto de árboles de decisión que se combinan para mejorar la precisión y la robustez del modelo, reduciendo el riesgo de sobreajuste. Puede decirse que es una variación de la regresión lineal donde se trabaja con varios árboles tratando de ajustar mediante regresiones lineales para determinados intervalos de datos.

Se decide trabajar con Regresión Lineal (RL) y Random Forest (RF) y ver si el  $R^2$  obtenido es adecuado o si es necesario buscar otro tipo de modelo.

### 4.1. Preparación de los Set de Datos

Para aplicar cualquiera de los modelos de predicción se debe dividir el conjunto de datos en dos: entrenamiento (train) y prueba (test). En general, la proporción es 80% - 20% la selección es aleatoria. En el conjunto de entrenamiento se encuentran los datos de los que el modelo (o algoritmo) aprende. Si se pretende realizar la elección de un modelo que resulte más adecuado, es aconsejable contar con un conjunto de validación (val), cuyos datos sirven para evaluar los diferentes modelos. Los datos del conjunto de validación se toman del conjunto de entrenamiento. El conjunto de datos de validación indica el grado de aprendizaje y adaptación del modelo, lo que permite realizar ajustes y optimizaciones en sus parámetros o hiperparámetros antes de la prueba final. Una vez que se realiza la elección del modelo, se trabaja con el conjunto de prueba que proporciona una evaluación final del modelo.

### 4.2. Regresión lineal

Lo primero que se decide es aplicar regresión Lineal. Se denomina regresión lineal simple cuando depende sólo de un parámetro para predecir el target o variable objetivo (en este caso Q), y se denomina múltiple cuando son varios los parámetros que se utilizan para realizar la predicción.

### 4.3. Regresión lineal considerando solo dependencia con DT

Se analiza primero el modelo predictivo trabajando con regresión lineal simple para  $\Delta T$  y luego para grado (pendiente de pared activa) para predecir Q. A continuación, se hizo una regresión lineal múltiple teniendo en cuenta ambos parámetros ( $\Delta T$  y  $\gamma$ ). A continuación, se muestra el gráfico obtenido con el set de validación, donde  $y_{val}$  es el valor que se obtuvo mediante la simulación de Q.

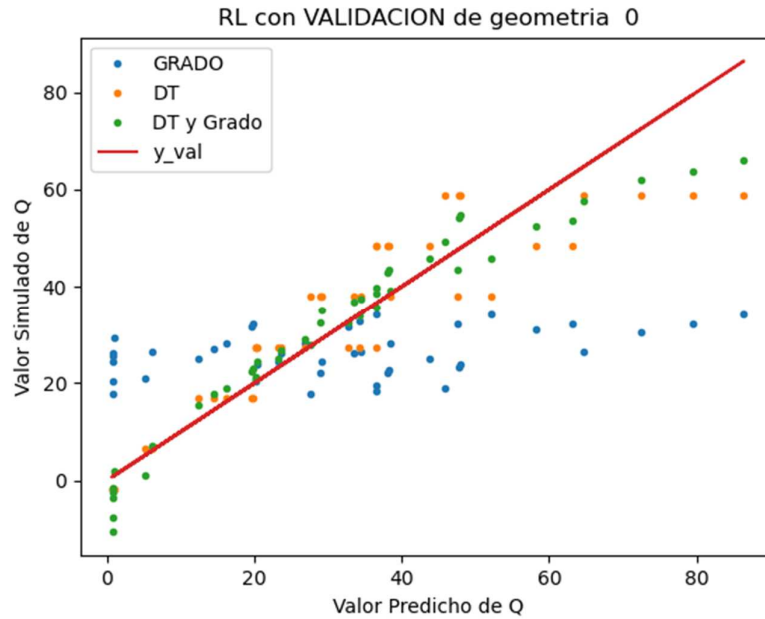


Figura 5: RL del set de pruebas, ático.

Los resultados del  $R^2$  arrojados por el programa realizado con el set de validación se muestran en la siguiente tabla.

Regresión Lineal	Parámetros	$R^2$ en entrenamiento	$R^2$ en validación
simple	$\gamma$	0.05514708818107705	0.06883859332818387
simple	$\Delta T$	0.8194570377584085	0.8165930331731346
múltiple	$\gamma$ y $\Delta T$	0.9394511647228838	0.9221485975579398

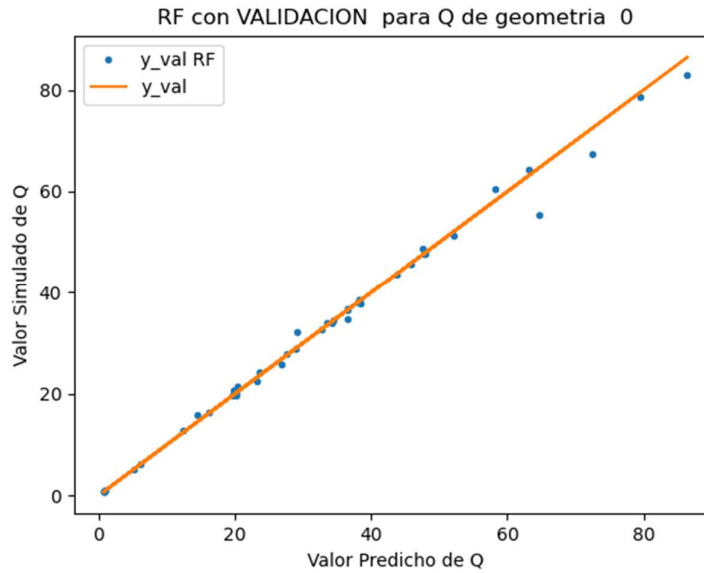
Tabla 2.  $R^2$  obtenido en los distintos casos de RL

En la tabla 2 se puede apreciar que el  $R^2$  en RLS considerando Grado es muy bajo, lo que indica que el ajuste lineal para este caso no es apropiado. Por otro lado, teniendo en cuenta la regresión lineal con parámetro  $\Delta T$  el ajuste es aceptable. Por último, el  $R^2$  obtenido para la RLM con ambos parámetros es el mayor, por lo que se puede decir que el MLR dependiendo del grado y  $\Delta T$  es el mejor de los modelos obtenidos considerando LR para poder predecir el Q,

#### 4.4. Random Forest

Para poder implementar RF se trabaja nuevamente con la biblioteca de sklearn. Se deben configurar ciertos parámetros con los que se construye el random forest. Para poder elegir la mejor combinación de hiperparámetros de todos los posibles se decide implementar gridsearch que permite entrenar y validar el modelo con cada una de las combinaciones posibles y elegir aquella que tenga el mejor desempeño.

A continuación, se muestra el gráfico obtenido con el set de validación, donde  $y_{val}$  es el valor que se obtuvo mediante la simulación del Q.



**Figura 6:** RF del set de validación

En la tabla 3 se muestran los valores obtenidos de  $R^2$  con los Sets de entrenamiento y validación. Analizando el  $R^2$  que se obtuvo mediante este modelo, configurando sus parámetros usando gridsearch, se puede decir que RF obtenido es un muy buen modelo para poder predecir el Q.

Modelo	Parámetros	$R^2$ en entrenamiento	$R^2$ en validación
RF	$\gamma$ y $\Delta T$	0.9984397382807936	0.9924165941488159

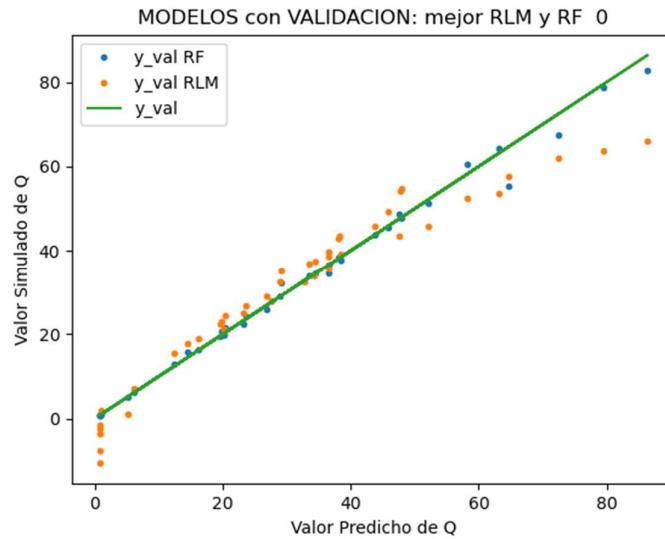
**Tabla 3.**  $R^2$  obtenido para RF

### 5. Comparación del MLR y RF

Tal como se describió previamente a través de la regresión lineal múltiple y del Random forest se obtuvieron dos modelos que presentaron valores de  $R^2$  en entrenamiento y validación muy aceptables. Para poder determinar cuál de ellos es el más recomendable para predecir el Q del caso de estudio, se compararon los valores obtenidos del  $R^2$  que pueden verse en la siguiente tabla y gráfico.

Modelo	Parámetros	$R^2$ en entrenamiento	$R^2$ en validación
RF	$\gamma$ y $\Delta T$	0.9984397382807936	0.9924165941488159
MLR	$\Delta T$ y $\gamma$	0.9394511647228838	0.9221485975579398

**Tabla 4.**  $R^2$  de entrenamiento y validación de los dos modelos que fueron considerados más apropiados



**Figura 7:** Predicción de Q para RF y RLM con set de validación

Se compararon los dos enfoques de aprendizaje supervisado para predecir el flujo de calor Q en función del salto térmico  $\Delta T$  y el ángulo de inclinación  $\gamma$ : MLR y RF. Aunque ambos modelos mostraron desempeño aceptable, tal como se puede ver en la tabla 4 y en la Figura 7, RF superó notablemente a MLR, especialmente en la capacidad de generalización a nuevos datos, con un coeficiente de determinación  $R^2$  superior al 0,99 en validación y prueba.

Esto puede explicarse por varias razones:

1. Relaciones no lineales: MLR asume una relación lineal entre las variables independientes ( $\Delta T$ ,  $\gamma$ ) y la variable dependiente (Q). Sin embargo, el comportamiento del flujo de calor en fenómenos de convección natural suele ser no lineal, especialmente cuando se modifican simultáneamente la geometría del recinto y el gradiente térmico. RF no impone esta suposición y puede capturar relaciones complejas y no lineales mediante la combinación de múltiples árboles de decisión.

2. Interacciones entre variables: MLR solo puede modelar interacciones entre variables si se introducen explícitamente como términos cruzados (por ejemplo,  $\Delta T \cdot \gamma$ ). En cambio, RF detecta automáticamente interacciones relevantes al dividir el espacio de características en múltiples ramas, ajustando mejor los patrones complejos sin intervención manual.

3. Robustez ante ruido y sobreajuste: RF es un modelo de tipo ensamble, lo que significa que promedia múltiples árboles de decisión construidos sobre subconjuntos de los datos. Esto reduce la varianza del modelo y mejora su robustez frente al sobreajuste y al ruido, en comparación con MLR que puede volverse inestable si hay correlaciones débiles o multicolinealidad.

4. Capacidad de capturar umbrales o comportamientos locales: En los datos simulados, puede haber valores críticos de  $\gamma$  o  $\Delta T$  que producen cambios abruptos en Q. RF es capaz de identificar estos umbrales y ajustar su predicción localmente, mientras que MLR solo puede ajustar un plano en el espacio de parámetros.

5. Evaluación empírica directa: Los resultados cuantitativos ( $R^2$  en validación y prueba) y los errores relativos frente a nuevos datos muestran que RF no solo ajusta mejor el conjunto de entrenamiento, sino que también generaliza adecuadamente a configuraciones no vistas, incluso en condiciones de “doble ciego”.

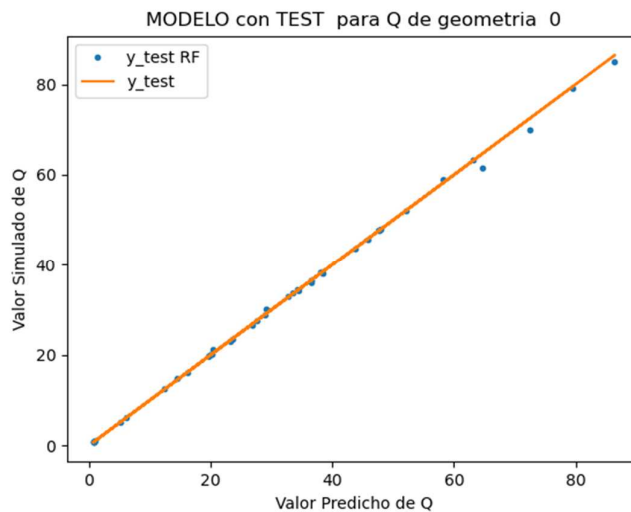
### 6. Puesta a punto del modelo elegido

Una vez elegido el modelo con el cuál se trabajará (RF), se incorporan los datos del conjunto de validación al de entrenamiento para así realizar un último entrenamiento con una mayor cantidad de datos que le permitirá al modelo realizar un mejor ajuste. Posteriormente con el modelo entrenado se realiza la prueba con el set de datos de testeo ( $y_{test}$ ).

En la siguiente tabla (Tabla 5) se puede observar el  $R^2$  obtenido tanto en el set de entrenamiento como en el de prueba. Y en la gráfica (Figura 8) se puede ver cuanto ajusta el Q obtenido por el modelo ( $y_{test}$  RF) respecto del valor obtenido por simulación ( $y_{test}$ ).

Modelo Elegido	$R^2$ en entrenamiento	$R^2$ en prueba
RF	0.9990912484762635	0.9989551091032514

**Tabla 5.**  $R^2$  de entrenamiento y prueba del modelo RF re-entrenado.



**Figura 8:** Predicción de Q para RF re-entrenado con Set de Prueba

### 7. Aplicación del modelo a nuevos datos

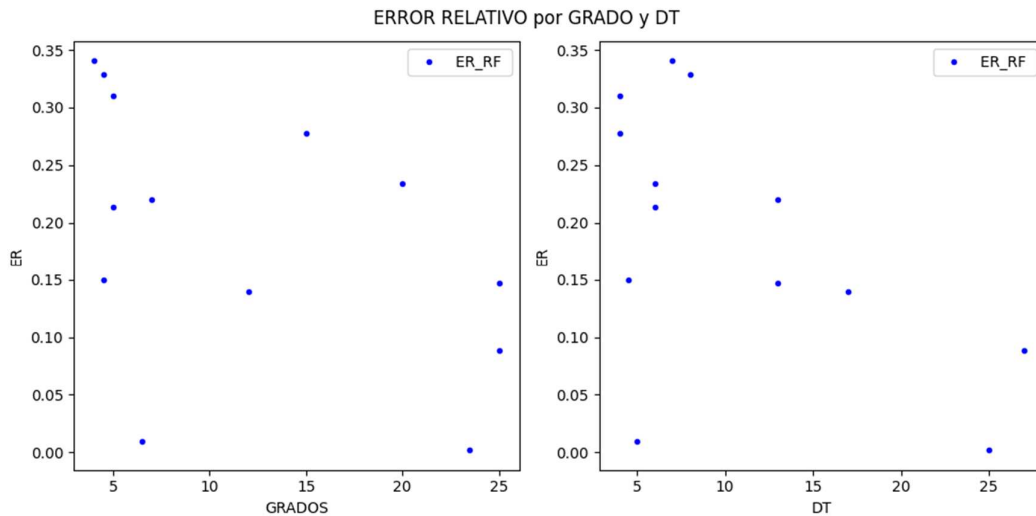
Para realizar una última prueba, se decide contrastar casos extras simulados, como pruebas de doble ciego para ver si el modelo funciona para  $\Delta T$  o grados con los que no fue entrenado. En la siguiente tabla se muestran los valores de  $y$  y  $\Delta T$  utilizados junto con los resultados obtenidos por la simulación en OF, predichos por el modelo RF y los errores relativos que sirven de comparación entre ambos.

$$ER_{RF} = \frac{\text{simulado} - RF}{RF}$$

Casos	grado	DT	Simulado	RF	ER_RF
0	23.5	25.0	378.610	37,927848	0.001766
1	25.0	27.0	403.990	36,828928	0.088370
2	4.0	7.0	12,616	8.319.792	0.340536
3	6.5	5.0	81.040	8,025932	0.009633
4	4.5	4.5	72.000	8,277692	0.149679
5	4.5	8.0	148.972	19,790036	0.328440
6	5.0	4.0	62.860	8.238.120	0.310550
7	15.0	4.0	4,771	6,094368	0.277377
8	5.0	6.0	104.810	8,23812	0.213995
9	20.0	6.0	68.430	5,23964	0.234307
10	7.0	13.0	24	29,2738	0.219742
11	25.0	13.0	175.430	20,114864	0.146603
12	12.0	17.0	288.490	24,825.628	0.139463

**Tabla 6.** Datos de valores nuevos y resultado de RF reentrenado

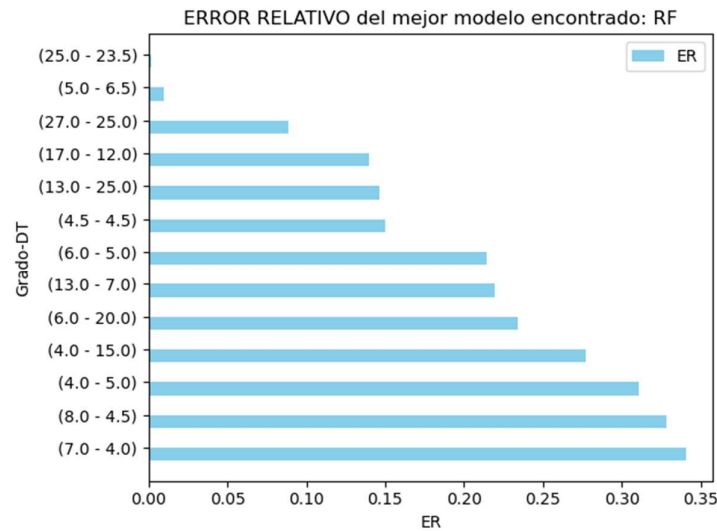
La figura 9, muestra el error relativo para los casos mostrados en la tabla 2, ordenados de menor a mayor. En la gráfica puede apreciarse que el 75% de los nuevos valores tienen un ER menor al 25%. Siendo para el 25% de los casos restantes un ER inferior al 35%. Calculando el  $R^2$



**Figura 9:** Error relativo obtenido contrastando el modelo con valores simulados analizando un parámetro

## 8. Conclusiones

La obtención de un valor promedio de Q proviene de la resolución de un sistema de ecuaciones diferenciales para un dominio de cálculo discretizado, con diferentes condiciones de borde ( $\Delta T$ ) y geometrías del recinto ( $\gamma$ ), tiene tiempos de simulación de varias horas. Contar con una herramienta que pueda predecir el valor de Q para determinados valores de  $\gamma$  y  $\Delta T$  resulta bastante conveniente, a la hora de contar con valores de Q cuando se pretende avanzar con el estudio de la transferencia de calor en recintos cerrados.



**Figura 10:** Error relativo obtenido contratando el modelo con valores simulados

En este trabajo se desarrolló un modelo predictivo para estimar el flujo de calor promedio ( $Q$ ) en una cavidad tipo ático, a partir de variables de entrada como el salto térmico ( $\Delta T$ ) y la inclinación de la pared activa ( $\gamma$ ). Utilizando datos obtenidos por simulación numérica con OpenFOAM, se entrenaron y evaluaron modelos de regresión lineal y Random Forest para determinar su capacidad de generalización.

Los resultados muestran que el modelo RF, optimizado mediante búsqueda en rejilla, presenta un desempeño significativamente superior al de la regresión lineal múltiple, alcanzando valores de  $R^2$  mayores al 0.99 tanto en entrenamiento como en validación. Asimismo, en la prueba con nuevos datos no utilizados durante el entrenamiento, el modelo mantuvo una buena capacidad predictiva, con errores relativos aceptables en la mayoría de los casos.

Se concluye que el modelo obtenido con Random Forest es una herramienta confiable y eficiente para reemplazar, en ciertas etapas del análisis, las simulaciones completas, que requieren varias horas de cómputo. Esta aproximación permite explorar más rápidamente el espacio de parámetros ( $\gamma$ ,  $\Delta T$ ), facilitando el diseño preliminar de sistemas de transferencia de calor en geometrías similares.

Como trabajos futuros, se sugiere:

- Evaluar el desempeño del modelo para otras configuraciones geométricas como ático invertido o trapecio simétrico.
- Considerar modelos de aprendizaje profundo que capten patrones temporales si se desea predecir la evolución de  $Q$  en el tiempo.
- Investigar la interpretación de los modelos (feature importance) para identificar con mayor claridad la sensibilidad del flujo de calor respecto a cada variable.

## Referencias

- [1] Esteban, E.S. et al, CONVECCIÓN NATURAL TRANSITORIA EN CAVIDADES TRAPEZOIDALES TIPO ATICO. *Mecánica Computacional Vol XXXVII*, págs. 1939-1950 (artículo completo). Editores: A. Cardona, L. Garelli, J.M. Gimenez, P.A. Kler, S. Márquez Damián, M.A. Storti. Santa Fe - Argentina, 5-7 Noviembre 2019. Sitio web: <http://www.amcaonline.org.ar>
- [2] Ganesh, N et al. Random Forest Regression-Based Machine Learning Model for Accurate Estimation of Fluid Flow in Curved Pipes. *Processes* 2021, 9, 2095. Sitio web: <https://doi.org/10.3390/pr9112095>