



Facultad de Ingeniería
Ingeniaritza Fakultatea

Grado en Ingeniería en Tecnologías Industriales

Industria Teknologien Ingeniaritzako gradua

Proyecto fin de grado

Gradu amaierako proiektua



Facultad de Ingeniería
Ingeniaritza Fakultatea

Grado en Ingeniería en Tecnologías Industriales

Industria Teknologien Ingeniaritzako gradua

Proyecto fin de grado

Gradu amaierako proiektua

Resumen

Este trabajo se sitúa en el ámbito de la ingeniería energética, centrado en el desarrollo de modelos ARX lineales que caracterizan la dinámica de la carga térmica en edificaciones.

Para la realización de este estudio, se ha utilizado información real proporcionada por la *National University of Singapore (NUS)*, que incluye datos de diversas variables correspondientes a cinco áreas diferentes de uno de los edificios en su campus universitario.

Las innovaciones en la metodología de este estudio consisten en una reorientación de los modelos de referencia que predicen la temperatura interior de la *NUS*, hacia un modelo enfocado en la predicción de la carga térmica. Esta nueva dirección está motivada por dos hallazgos clave: la escasa variabilidad de la temperatura interior y la alta significatividad de la temperatura de la hora anterior como factor explicativo crucial en los modelos.

Este modelo se construye evaluando el impacto de factores externos, como la temperatura exterior y la radiación solar, e internos, como la ocupación de los espacios. Además, se incluye la diferencia entre la consigna del controlador climático y la temperatura interior como variable innovadora.

Se consigue desarrollar un modelo que predice la carga térmica con valores de R^2 superiores a 0,6 y un MAE de aproximadamente 0,133 kWh. Por último, se detecta un claro desajuste en el modelo debido al encendido y apagado repentino del sistema de control climático, lo que sienta la base para futuras líneas de investigación.

Descriptores

Evolución térmico-energética en edificios

Predicción de carga térmica

Modelo basado en datos

Eficiencia energética

Índice

1. INTRODUCCIÓN	1
2. ANTECEDENTES.....	3
2.1 Contexto	3
2.2 Usos de modelos predictivos de la carga térmica.....	3
2.3 Métodos de predicción utilizados	4
2.4 Modelos de caja negra	5
3. JUSTIFICACIÓN	7
4. OBJETIVOS Y ALCANCE	8
4.1 Objetivos	8
4.2 Alcance	8
4.2.1 Dentro del alcance del proyecto	8
4.2.2 Fuera del alcance del proyecto y limitaciones	9
5. FUNDAMENTOS TEÓRICOS	10
5.1 Tipos de modelos de predicción.....	10
5.1.1 Modelos lineales.....	10
5.1.2 Modelos no lineales.....	11
5.1.3 Modelos de series temporales	13
5.1.4 Modelos de redes neuronales artificiales (ANN).....	14
5.1.5 Modelos ARX	15
5.2 Evaluación de modelos y métricas de error	16
5.2.1 Coeficiente de determinación (R^2)	16
5.2.2 Coeficiente de determinación ajustado ($R_{ajustado}^2$).....	16
5.2.3 Raíz del error medio cuadrático (RMSE).....	17
5.2.4 Error medio absoluto (MAE).....	17

5.3	Métodos de construcción de modelos.....	18
5.3.1	Métodos.....	18
5.3.2	Significatividad de las variables	19
6.	MEMORIA TÉCNICA.....	20
6.1	Metodología	20
6.1.1	Primera fase: mejora de los modelos de referencia	20
6.1.2	Segunda fase: creación de nuevos modelos.....	20
6.2	Herramienta software	21
6.2.1	Excel.....	21
6.2.2	IBM SPSS	22
6.2.3	MATLAB	23
6.2.4	R.....	23
6.3	Edificio y base de datos	25
6.3.1	Edificio	25
6.3.2	Base de datos.....	26
6.3.3	Habitación seleccionada	29
6.4	Primera fase: mejora de modelos de referencia	30
6.4.1	Modelo de referencia	30
6.4.2	Mejora modelos predictivos.....	31
6.4.3	Observaciones importantes sobre los modelos	44
6.4.4	Primeras conclusiones	48
6.5	Segunda fase: creación de nuevos modelos.....	50
6.5.1	Método de construcción	50
6.5.2	Variables seleccionadas.....	50
6.5.3	Otras consideraciones	52

6.5.4	Métricas de error	52
6.5.5	Modelado	53
6.5.6	Análisis de los residuos.....	62
7.	CONCLUSIONES Y TRABAJO FUTURO	66
8.	VALORACION ETICA DEL PROYECTO.....	68
9.	RIESGOS	70
10.	PLANIFICACIÓN Y PRESUPUESTO	71
10.1	Planificación.....	71
10.2	Presupuesto.....	73
11.	BIBLIOGRAFÍA	75

Índice de ilustraciones

Ilustración 1. Estructura de una red neuronal [34]	15
Ilustración 2. Ubicación de Singapur en el mapa del mundo	25
Ilustración 3. Plano de la habitación 4	29
Ilustración 4. Evolución de la temperatura interior y la exterior del 20 al 24 de septiembre de 2021	46
Ilustración 5. Evolución de la temperatura interior con respecto a la temperatura exterior (08/09/2021)	47
Ilustración 6. Evolución de la temperatura interior con respecto a la temperatura exterior (14/09/2021)	47
Ilustración 7. Relación entre la temperatura interior y su consigna	48
Ilustración 8. Tabla de correlaciones entre las variables seleccionadas	51
Ilustración 9. Representación gráfica de la energía consumida para la climatización medida y la energía consumida predicha para el modelo inicial.....	54
Ilustración 10. Representación gráfica de la energía consumida para la climatización medida y la energía consumida predicha para el modelo inicial.....	60
Ilustración 11. Representación gráfica de la energía consumida para la climatización medida y la energía consumida predicha para el modelo ARX optimizado.	60
Ilustración 12. Representación gráfica de la energía consumida para la climatización medida y la energía consumida predicha para el modelo ARX optimizado, con intervalos de confianza	61
Ilustración 13. Gráfico de dispersión de los residuos frente a la diferencia entre la consigna y la temperatura interior	62
Ilustración 14. Gráfico de dispersión de los residuos frente a la ocupación por conteo	63
Ilustración 15. Gráfico de dispersión de los residuos frente a la temperatura exterior	63
Ilustración 16. Gráfico de dispersión de los residuos frente a la radiación solar	64
Ilustración 17. Evolución de la radiación para la semana del 20 septiembre de 2021. Las líneas rojas punteadas se colocan sobre las 7:30 y 20:30.....	65
Ilustración 18. Evolución de la carga térmica para la semana del 20 septiembre de 2021. Las líneas rojas punteadas se colocan sobre las 7:30 y 20:30.....	65
Ilustración 19. Diagrama de Gantt para la planificación del proyecto.....	72

Ilustración 20. Diagrama de Gantt para la planificación del proyecto.....	72
Ilustración 21. Información general sobre las horas empleadas por estudiante y director	73

Índice de tablas

Tabla 1. Comparativa entre software para análisis estadístico y predicción de la carga térmica	24
Tabla 2. Información general sobre las habitaciones medidas [29].....	26
Tabla 3. Variables disponibles para cada habitación [29]	27
Tabla 4. Significatividad de los modelos de partida 1 y 3	34
Tabla 5. Comparativa de significatividad entre los modelos 1 y 3 y el modelo mejorado para 1 y 3...	35
Tabla 6. Comparativa de variables, R^2 y MAE entre los modelos 1 y 3, y el mejorado	35
Tabla 7. Coeficientes para las variables del modelo 1 y 3 mejorado	36
Tabla 8. Significatividad de las variables del modelo de partida 2.....	37
Tabla 9. Comparativa de n.º de variables, R^2 y MAE entre el modelo 2 y el mejorado	38
Tabla 10. Coeficientes para las variables del modelo 2 mejorado.....	38
Tabla 11. Análisis de significatividad de las variables del modelo 4 de partida.....	39
Tabla 12. Comparativa de n.º de variables, R^2 y MAE entre el modelo 4 y el mejorado	40
Tabla 13. Coeficientes para las variables del modelo 4 mejorado.....	40
Tabla 14. Análisis de significatividad de las variables del modelo 5, 6, 7 y 8 de partida	43
Tabla 15. Modelos 5, 6, 7 y 8 mejorados	44
Tabla 16. Coeficientes estandarizados para el modelo completo	45
Tabla 17. Tabla de correlaciones para la temperatura interior y la energía consumida para refrigerar agua	49
Tabla 18. Resumen de los aspectos más relevantes del modelo inicial.....	54
Tabla 19. Resumen de los aspectos más relevantes del modelo inicial y radiación solar	55
Tabla 20. Resumen de los aspectos más relevantes del modelo inicial y radiación solar + ocupación por conteo.....	56
Tabla 21. Resumen de los aspectos más importantes del modelo óptimo utilizando solo variables exógenas en el momento t.....	57
Tabla 22. Resumen de los aspectos más importantes del modelo óptimo con solo variables exógenas en el momento t y energía consumida para la climatización en t-1	58

Tabla 23. Resumen de los aspectos más importantes del modelo óptimo con solo variables exógenas en el momento t y energía consumida para la climatización en t-1 despues de stepAIC.....	59
Tabla 24. Desglose del presupuesto por empleo de recursos	74

1. INTRODUCCIÓN

La Universidad de Deusto, a través de Deustotech, ha fomentado desde 2005 la investigación aplicada en nuevas tecnologías dentro de su Facultad de Ingeniería. El grupo de investigación, Deusto Sustainable Research Group[1], concentra su misión en apoyar las actividades de TIC en empresas y en la sociedad a través de la investigación, el desarrollo tecnológico, la innovación y la transferencia de conocimientos. Focalizándose en niveles de madurez tecnológica (TRL) del 2 al 7, su actividad se articula en cuatro campos aplicados: Industria, Movilidad, Energía y Sociedad, contando además con la Cátedra de Matemáticas Aplicadas como una actividad transversal que apoya a las anteriores. El equipo trabaja con datos heterogéneos a lo largo del ciclo de vida y en conformidad con los principios éticos y humanistas que caracterizan a la Universidad de Deusto. El grupo ha liderado investigaciones en el sector ambiental, incluyendo sostenibilidad, eficiencia de recursos y economía circular, enfocándose en áreas como la economía circular, el uso sostenible de recursos, la eficiencia energética y el IoT & Industria.

Una de las líneas de investigación que ha sido prioritaria para el grupo es la mejora de modelos de predicción para aumentar la eficiencia energética en edificios. La investigación en esta área es crucial para desarrollar sistemas de control predictivo que mejoren significativamente la gestión de la energía en los edificios, optimizando el rendimiento y la eficiencia en tiempo real, adaptándose proactivamente a las variaciones y perturbaciones del sistema.

Esto se debe a que la demanda imperante de llevar a cabo la transición energética se manifiesta no solo en la búsqueda de fuentes de energía renovables y respetuosas con el medio ambiente, sino también en el impulso hacia infraestructuras energéticas más eficientes. Este imperativo encuentra respaldo en los Objetivos de Desarrollo Sostenible, específicamente en el objetivo número 7[2], titulado "Energía asequible y no contaminante". En el marco de este objetivo, destacan dos metas estrechamente vinculadas con la eficiencia energética: "7.3 duplicar la tasa mundial de mejora de la eficiencia energética" y "7. b ampliar la infraestructura y mejorar la tecnología para prestar servicios energéticos modernos y sostenibles para todos en los países en desarrollo [...]". Este contexto resalta la urgencia de abordar la eficiencia energética como un componente esencial de la transición hacia un sistema energético más sostenible.

A pesar de los avances en estándares mínimos de rendimiento y códigos de energía para edificaciones, así como del crecimiento en la adopción de tecnologías eficientes y renovables en edificaciones, el sector enfrenta el desafío de implementar cambios más rápidos para cumplir con el Escenario de Emisiones Netas Cero para 2050 [3]].

El desafío de la transición energética se hace patente al considerar que los edificios, responsables de aproximadamente el 30–40% de las emisiones globales de gases de efecto invernadero y el consumo de energía [4], representan un componente crítico en la lucha contra el cambio climático. De este porcentaje, hasta un 85% se consume durante la fase operativa de los edificios [5]. Controlar los sistemas de edificación emerge como una tarea desafiante debido a la dinámica variable del sistema y las perturbaciones inherentes al mismo. En la actualidad, los controles proporcionales, integrales y

derivativos (PID) son los más utilizados en la práctica, pero su rendimiento puede ser insatisfactorio sin una sintonización adecuada y la falta de un control supervisor multiobjetivo.

Los sistemas de control predictivo, en contraste con los sistemas reactivos convencionales, se caracterizan por anticipar y ajustar activamente las acciones de control en función de las condiciones futuras esperadas. Esta capacidad de anticipación permite a los sistemas de control predictivo optimizar el rendimiento y la eficiencia en tiempo real, adaptándose proactivamente a las variaciones y perturbaciones del sistema. Diferentes estudios en los que intervienen sistemas de control predictivo revelan ahorros energéticos notables, oscilando entre el 10% y el 46%, en comparación con los sistemas de control reactivos [6] [7].

En este contexto, la necesidad de modelos capaces de predecir la carga térmica futura es fundamental para el funcionamiento efectivo de los sistemas de control predictivo en edificaciones. Al prever la carga térmica, los sistemas de control predictivo pueden ajustar de manera más eficiente el uso de energía, regulando, por ejemplo, el funcionamiento de los sistemas de calefacción, ventilación y aire acondicionado (HVAC) para maximizar la eficiencia energética y el confort, al mismo tiempo que minimizan el costo y el impacto ambiental.

2. ANTECEDENTES

En este apartado se revisa la bibliografía de predicción de cargas térmicas de edificios y de estrategias de control predictivo para entender mejor el contexto. Esta búsqueda de información ha sido realizada de manera sencilla y ordenada gracias al artículo escrito por Huiheng Liu, Jinrui Liang, Yanchen Liu y Huijun Wu de acceso abierto que consiste en un repaso sobre sistemas de predicción del consumo energético de edificios basado en datos[8].

2.1 CONTEXTO

Los edificios representaron el 36% de la demanda global de energía y el 37% de las emisiones de CO₂ relacionadas con la energía en 2020[4]. Durante la fase operativa de los edificios, se consume más del 80% de la energía total de estos. En los países en desarrollo, las emisiones de carbono de los edificios pueden llegar al 52% [5]. Por ejemplo, China, el país en desarrollo más grande del mundo, produjo 9467 millones de toneladas de CO₂ relacionado con la energía en 2018, lo que representó aproximadamente el 29% de las emisiones globales de CO₂. En este contexto, el sector de la construcción contribuyó con el 20% del consumo total de energía y el 30% de las emisiones totales de CO₂ [9]. Por lo tanto, la reducción del consumo de energía y las emisiones de carbono en los edificios es una necesidad urgente en el contexto de la neutralidad de carbono.

Mejorar la eficiencia energética de los edificios puede reducir el consumo de energía en un 30-80% y disminuir significativamente las correspondientes emisiones de carbono de estos [4]. La predicción del consumo energético de los edificios, *BECP (building energy consumption prediction)*, es una iniciativa efectiva para mejorar la eficiencia energética, y ha desempeñado un papel importante en el control de la energía del edificio, la optimización del diseño, la evaluación de la remodelación y la orientación del costo de la energía. En la etapa de operación del edificio, el BECP se combina con los requisitos del grupo de usuarios finales para personalizar la operación de equipos consumidores de energía, optimizando su eficiencia energética sin afectar el confort térmico de los mismos [10].

2.2 USOS DE MODELOS PREDICTIVOS DE LA CARGA TÉRMICA

Mariam et al. propusieron un modelo de control predictivo basado en redes neuronales para un sistema de HVAC en el gimnasio de la Universidad de Qatar, junto con el sistema de control de gestión de planificación, logrando ahorros de energía de hasta el 46% mientras optimizaba conjuntamente el confort térmico y la calidad del aire [6]. Tomasz et al. realizaron predicciones de consumo de energía y control de eficiencia energética para una casa multifamiliar y un edificio de oficinas en Polonia, logrando ahorros de energía reales que superaron el 15% y 24% [7].

En la etapa de diseño de edificios, Kim et al. exploraron los efectos de la permeabilidad del aire, el coeficiente de ganancia de calor solar y la conductividad térmica en el consumo de energía del edificio, variando las variables de diseño basadas en el modelo de predicción desarrollado [11]. La permeabilidad del aire tiene una influencia significativa en la carga de calefacción, mientras que el coeficiente de ganancia de calor solar tiene el mayor impacto en la carga de enfriamiento. El incremento de la conductividad térmica redujo el consumo de energía de enfriamiento entre un 8-

26%. Agregar el coeficiente de transferencia de calor de la fachada conduce a un aumento en la necesidad de energía de calefacción de aproximadamente un 27-29%.

Además, el BECP puede guiar el precio de la energía suministrada a los clientes [12]. Jinseok y Ki-Il propusieron un modelo de predicción de consumo de energía de edificios (BECP) basado en memoria a largo y corto plazo, *LSTM (long short-term memory)* y un algoritmo de operación basado en el tiempo de uso, *TOU (time of use)*, que teóricamente podría reducir los costos de demanda máxima en un 22% [11]. Frente a futuras crisis energéticas, la previsión del consumo energético en los edificios puede también ofrecer información valiosa para los gestores de instalaciones, proveedores de electricidad y responsables de tomar decisiones, ayudándoles a planificar el uso de la energía, monitorizar anomalías en el consumo, regular los costos energéticos y responder a estrategias de demanda para remodelar los perfiles de carga de los edificios y reducir la demanda [13].

Otro grupo de científicos integraron la predicción del consumo energético de un sistema HVAC con el riesgo de transmisión aérea para determinar el espacio requerido y el costo energético más bajo para reducir el riesgo de [14]. Combinaron la previsión del consumo energético causado por la COVID-19 con un modelo econométrico para integrar los impactos a largo y corto plazo bajo una completa verificación [15]. Además, la predicción del consumo energético de los edificios (BECP) desempeña un papel importante en la operación de los edificios, la evaluación de la eficiencia energética, la detección y diagnóstico de errores [15], la gestión de la demanda [16] y el mantenimiento [17].

2.3 MÉTODOS DE PREDICCIÓN UTILIZADOS

Actualmente, los métodos de predicción del consumo energético de los edificios (BECP) se dividen principalmente en tres tipos: caja blanca, caja negra y caja gris [18]. El modelo de caja blanca, también conocido como modelo físico, se construye y analiza basándose en principios termodinámicos e información detallada sobre las características energéticas del edificio. Los programas de simulación de consumo energético de edificios más utilizados incluyen EnergyPlus, DesignBuilder, DeST, eQuest, TRNSYS y DOE-2. Estos programas pueden establecer modelos de predicción de consumo energético según los parámetros estructurales específicos de los edificios, datos meteorológicos externos y el rendimiento del sistema de aire acondicionado. Sin embargo, la desventaja de un modelo físico es que requiere múltiples verificaciones y ajustes complejos para asegurar una fiabilidad satisfactoria de los resultados predichos.

Por otro lado, los académicos han desarrollado el modelo de caja negra, también conocido como método basado en datos, que no se basa en los principios termodinámicos ni en las características energéticas del edificio. El método basado en datos utiliza algoritmos específicos para analizar conjuntos de datos grandes recopilados y extraer la lógica entre los datos para lograr la toma de decisiones automática. En comparación con el modelo físico, los métodos basados en datos solo se basan en grandes conjuntos de datos y algoritmos de aprendizaje automático (ML), los cuales requieren un tamaño grande de muestras de datos [19].

El modelo de caja gris es una combinación de un modelo físico y un método basado en datos para predecir el consumo energético de los edificios. Aunque este modelo ofrece una mayor precisión en las predicciones y reduce la dificultad de los cálculos en comparación con el modelo físico, el proceso

de establecimiento del modelo puede incorporar suposiciones inexactas. En general, las cajas negras proporcionan mayor conveniencia práctica que las cajas blancas o grises. El uso de análisis de datos y la minería de grandes conjuntos de datos evitan la necesidad de construir un modelo físico para prever el consumo de energía [8].

Debido a esto y a contar con una gran base de datos, se ha decidido hacer uso de un modelo de caja negra. A continuación, se presentan algunos artículos que comparan los diferentes modelos dentro de esta categoría en el ámbito de la predicción de carga térmica.

2.4 MODELOS DE CAJA NEGRA

Los modelos autorregresivos con entradas externas, identificados como ARX (AutoRegressive with eXogenous inputs), o ARMAX cuando incorporan medias móviles (AutoRegressive Moving Average with eXogenous inputs), son técnicas frecuentemente utilizadas para la predicción de la temperatura interior y la carga térmica. Estos modelos pueden clasificarse en lineales o no lineales. Los modelos no lineales son particularmente útiles para descomponer y estudiar las contribuciones no lineales dentro de un modelo, como los efectos de la radiación solar. En el apartado de fundamentos teóricos se profundiza más en estos modelos.

La implementación de un modelo ARX no lineal en diversas aulas de una escuela en Dinamarca obtuvo con un error cuadrático medio (RMSE) de entre 0,25°C y 0,6°C para predicciones a un día [20]. Este modelo incorporó como variables exógenas la temperatura exterior, la radiación solar y datos sobre los radiadores.

Para explorar la eficacia de los modelos ARX y ARMAX en la predicción de la temperatura interior, utilizando como factores la temperatura exterior, la radiación solar global, la velocidad del viento y la humedad relativa exterior, se implementaron estos modelos en diversas aulas de la Universidad Autónoma de Querétaro en México [21]. Los hallazgos indicaron que los modelos ARX proporcionaron mejores resultados que los ARMAX, con coeficientes de determinación de 0.9457 y 0.9056, respectivamente. El estudio también determinó que incorporar datos con más de 20 minutos de antigüedad no generaba mejoras significativas en las predicciones. Con un intervalo de muestreo de 5 minutos, se observó que la información de más de cuatro pasos de tiempo atrás no contribuía a mejorar la precisión de los modelos.

Otro conjunto de modelos que ha ganado relevancia recientemente son aquellos basados en aprendizaje automático e inteligencia artificial. Estos modelos requieren de grandes cantidades de datos y, por esta razón, son susceptibles a la sobre especificación [22].

Dentro de este grupo se encuentran las redes neuronales artificiales, que han demostrado ser especialmente eficaces. Utilizando redes neuronales, se han obtenido coeficientes de determinación (R^2) que superan el 0.9, indicando una alta correlación entre las temperaturas predichas y las temperaturas reales observadas [22].

También es común encontrar investigaciones que evalúan diferentes modelos en la predicción de temperaturas internas de edificios utilizando datos variados de diversas estructuras. Por ejemplo, una investigación realizada en un edificio típico de los que fueron construidos en el sector terciario en Italia durante los años 60 y 70 [23], analiza diferentes modelos: ARX y NARX (modelo autorregresivo neuronal con entradas externas). El modelo NARX tiende a ofrecer rendimientos inferiores en comparación con los otros modelos ARX a medida que se extiende el horizonte de tiempo, aunque destaca cuando las predicciones se limitan a un solo paso temporal. En general, los dos modelos son efectivos para predecir comportamientos a corto plazo, con horizontes de 15 minutos, 1 hora y 3 horas, mostrando un RMSE menor de 0.5°C y 1°C para los horizontes de 1 y 3 horas, respectivamente. Este estudio también identificó dos patrones de comportamiento de los ocupantes que causaban perturbaciones observables en el modelo, como la apertura de puertas y ventanas por las mañanas para actividades de limpieza y la constante apertura de puertas que dan a la calle al mediodía.

En el campo de la predicción del consumo energético, se han identificado patrones diarios de carga térmica en edificios utilizando una mezcla de técnicas de aprendizaje supervisado y no supervisado, alcanzando precisiones entre 0.7 y 0.85, entendiendo la precisión como la proporción de predicciones correctas sobre el total de predicciones realizadas. Los resultados indican que los principales factores que influyen en los patrones de consumo son las variaciones estacionales y la temperatura media diaria [24].

En cuanto a las estrategias de predicción basadas en "ensemble learning", que combinan árboles de decisión para generar patrones de consumo de energía seguidos de la aplicación de técnicas de ensemble para optimizar la predicción para cada patrón, se ha logrado un coeficiente de variación del error medio cuadrático de 15,6% cuando se dispone de todos los datos necesarios [25]. Este enfoque ha demostrado ser efectivo y también presenta buenos resultados incluso con un conjunto de datos reducido para el entrenamiento del modelo.

Además, se han utilizado modelos basados en datos para prever las cargas térmicas de edificios conectados a redes de calefacción urbanas, obteniendo coeficientes de determinación (R^2) de entre 0.47 y 0.95 para datos por hora, y entre 0.7 y 0.99 para datos diarios [26]. Un modelo simplificado, diseñado para la predicción a corto plazo de las cargas térmicas, logró un error absoluto menor al 4% y valores de R^2 entre 0.92 y 0.94 utilizando datos sintéticos para entrenar y evaluar el modelo, similar al modelo ARX descrito en esta tesis [27].

Por último, el proyecto de fin de grado desarrollado por Iñigo García de Eulate explora la predicción de la temperatura interior utilizando un conjunto de datos facilitado por la National University of Singapore [28]. Los modelos desarrollados fueron ARX lineales. Alcanzaron coeficientes de determinación (R^2) superiores a 0.99 y un error absoluto medio (MAE) aproximadamente de 0.137°C . Además, los resultados muestran que con un 95% de confianza, las predicciones varían, como máximo, 0.404°C por encima o 0.402°C por debajo de las mediciones reales. Los modelos creados en este estudio se toman como punto de partida para el desarrollo del presente trabajo.

3. JUSTIFICACIÓN

Este proyecto, alineado con los principios y la misión del *Deusto Sustainable Research Group* de la Universidad de Deusto, propone desarrollar y validar modelos ARX lineales para predecir la carga térmica a una hora vista en infraestructuras, abordando el desafío global de la eficiencia energética en edificaciones.

El proyecto queda justificado desde el **punto de vista técnico** ya que, la Universidad de Deusto, a través de *Deustotech*, ha enfatizado la importancia de la innovación en la aplicación de nuevas tecnologías para la mejora de la eficiencia energética. Este proyecto se fundamenta en la necesidad de desarrollar sistemas de control predictivo que ajusten de manera eficiente y anticipada los consumos de energía en los edificios, mejorando el rendimiento y adaptándose a variaciones externas e internas. Además, está alineado con una de las líneas de trabajo que había quedado abierta en estudios anteriores.

Asimismo, el modelo simplificado diseñado para la predicción a corto plazo de las cargas térmicas que logró un error absoluto menor al 4% y valores de R^2 entre 0.92 y 0.94 utilizando datos sintéticos para entrenar y evaluar el modelo utilizando un modelo ARX lineal [27] demuestra la viabilidad del presente proyecto. Dicho modelo fue desarrollado para datos tomados en Bilbao y Madrid. Para este proyecto, se emplean datos de la Universidad Nacional de Singapur, lo que justifica aún más el trabajo debido a las diferencias en clima y latitud, permitiendo así explorar la adaptabilidad y eficacia de un modelo ARX en diversas condiciones ambientales y climáticas.

Desde una **perspectiva económica**, el proyecto presenta un potencial económico significativo al permitir una gestión más eficaz de los recursos, reduciendo el costo operativo de los edificios. Los estudios han demostrado que la implementación de modelos de control predictivo en sistemas HVAC puede resultar en ahorros energéticos significativos, como los observados en la Universidad de Qatar con ahorros de hasta el 46% [6], y en edificios en Polonia con ahorros del 15% al 24% [7]. Estrategias de predicción del consumo de energía también han mostrado potencial para reducir los costos de demanda máxima en hasta un 22% [11], destacando la eficacia de los modelos predictivos en la optimización del uso y coste de la energía en distintos escenarios y emergencias, como las sanitarias, proporcionando una gestión más eficiente y estratégica.

Desde el **punto de vista social** al mejorar el control sobre los sistemas de climatización, se mejora el confort de los usuarios, aspecto que repercute directamente en la salud y bienestar de quienes habitan o utilizan los espacios edificados. Este enfoque está en consonancia con los principios éticos y humanistas que promueve la Universidad de Deusto, buscando un impacto social positivo mediante la aplicación de la tecnología. Además, desde un **prisma ambiental**, la eficiencia energética en edificios es clave para la reducción de las emisiones de CO₂, con los edificios responsables de una porción significativa del consumo energético global. Este proyecto apoya la transición hacia un modelo energético sostenible, contribuyendo al objetivo de alcanzar el Escenario de Emisiones Netas Cero para 2050 [3], un desafío imperativo ante el cambio climático.

4. OBJETIVOS Y ALCANCE

4.1 OBJETIVOS

El objetivo principal es diseñar y desarrollar un modelo de caracterización ARX lineal de la carga térmica en edificios, utilizando la base de datos proporcionada por la Universidad Nacional de Singapur, donde se encuentra información sobre variables exteriores e interiores del edificio.

Objetivos específicos:

- Alcanzar un nivel de ajuste suficientemente significativo, evaluado por el coeficiente de determinación (R^2) mayor que 0,6.
- Lograr que el modelo obtenga un error medio absoluto (MAE) inferior a 0,15 kWh.
- Reducir la cantidad de parámetros siempre y cuando el modelo mantenga su capacidad predictiva.
- Llevar a cabo un análisis de residuos que permita comprender la naturaleza de los errores más repetidos.

Contribución:

Cambio de enfoque respecto al proyecto de referencia, identificando la predicción de la carga térmica con respecto a la predicción de la temperatura interna como una mejor alternativa, debido principalmente a tratarse de un edificio climáticamente controlado y con muy poca variabilidad de la temperatura interior. Además, se sientan las bases de un nuevo modelo de partida sobre el que poder desarrollar un estudio más específico.

4.2 ALCANCE

El alcance de este proyecto abarca diversas áreas esenciales para el desarrollo completo y efectivo de modelos de predicción de la carga térmica en edificaciones.

4.2.1 Dentro del alcance del proyecto

Inicialmente, se ha realizado una investigación exhaustiva del estado del arte en lo relacionado con los modelos de predicción de la carga térmica y del modelo de referencia. Esto ha incluido un resumen de los estudios previos detallando, las metodologías aplicadas y los resultados alcanzados en el campo, estableciendo así una sólida base teórica y técnica para el desarrollo del proyecto.

Además, se desarrolla un amplio conocimiento del lenguaje de programación R, esencial para la comprensión e implementación de los modelos estadísticos y predictivos. Este esfuerzo incluye tanto el aprendizaje técnico de R como la aplicación práctica en el contexto de la modelización de datos de carga térmica. La competencia en R es crucial, dado que permite la manipulación flexible y robusta de grandes conjuntos de datos y la implementación efectiva de técnicas estadísticas avanzadas.

Este estudio se centra en el análisis y uso de datos proporcionados por la Universidad Nacional de Singapur, concretamente del conjunto *ROBOD (Room-level Occupancy and Building Operation Dataset)* [29] del edificio 4 de la Escuela de Diseño y Medioambiente de la Universidad Nacional de Singapur (NUS). Este conjunto de datos abarca información detallada sobre diversas variables, incluyendo calidad medioambiental interior, Wi-Fi, energía, operativa HVAC, climatología exterior y ocupación.

El proyecto contempla un análisis profundo y detallado de las variables seleccionadas para la predicción. Este análisis involucra la exploración de cómo cada variable contribuye a la variabilidad de la carga térmica, su relevancia dentro del modelo y su influencia en los resultados de predicción. Se realiza una evaluación meticulosa para determinar las variables más significativas y su correlación con la carga térmica.

Asimismo, el proyecto abarca el desarrollo de un modelo de predicción de la carga térmica en edificaciones.

En el desarrollo del modelo también se incluye la validación de este con métricas de error como el R^2 y el MAE, así como el uso del criterio Akaike.

Finalmente, el proyecto incluye un análisis de los residuos del modelo desarrollado. El estudio de los residuos ayuda a mejorar la precisión del modelo, proporcionando insights valiosos sobre posibles mejoras y refinamientos en la metodología de predicción.

4.2.2 Fuera del alcance del proyecto y limitaciones

El alcance de este proyecto no incluye datos que no estén directamente relacionados con las variables establecidas, o derivadas de ellas, en el conjunto *ROBOD*.

No se consideran otros tipos de modelos predictivos que no estén alineados con los objetivos específicos, como los de redes neuronales artificiales *ANN (artificial neural network)*.

Aunque el proyecto no contempla el uso del modelo desarrollado para aplicaciones distintas a la optimización de la eficiencia energética de edificios, sí se reconocen posibles extensiones de su aplicación. Por ejemplo, el modelo podría ser útil para la gestión de la demanda de energía, donde predicciones precisas de la carga térmica podrían ayudar a equilibrar la carga de la red eléctrica en tiempo real.

Sin embargo, es importante aclarar que, aunque estas aplicaciones son factibles, el proyecto actual no incluye la implementación del modelo en un controlador real ni se explorarán estudios detallados sobre cómo llevar a cabo dicha implementación.

Además, el proyecto está limitado por los recursos computacionales disponibles y el tiempo asignado para su finalización, lo que ha podido afectar la profundidad de algunos análisis. Aunque el objetivo es desarrollar un modelo robusto, la aplicación de este podría estar limitada a condiciones similares a las de la base de datos utilizada, lo que podría afectar la generalización a otros contextos o tipos de edificios.

5. FUNDAMENTOS TEÓRICOS

En este apartado se explicarán los fundamentos teóricos y definiciones necesarias que se deben tener para el desarrollo del proyecto, así como permitir la comprensión del trabajo posterior.

5.1 TIPOS DE MODELOS DE PREDICCIÓN

En el ámbito de la predicción y análisis de datos, existen múltiples modelos estadísticos diseñados para abordar diferentes tipos de relaciones y dinámicas entre variables. Estos modelos varían desde los lineales, como la regresión lineal simple, hasta modelos más complejos y sofisticados que pueden manejar relaciones no lineales y patrones temporales. Algunos de los modelos más destacados incluyen los modelos no lineales, como las regresiones polinómicas y las funciones exponenciales; modelos de series temporales, que son cruciales para datos cronológicos como los ARIMA; y los modelos ARX, que combinan enfoques autorregresivos con influencias externas. Además, las Redes Neuronales Artificiales ofrecen un enfoque poderoso para detectar patrones complejos, mientras que las Máquinas de Soporte Vectorial se utilizan para problemas de clasificación y regresión en espacios de alta dimensionalidad.

En este apartado se va a realizar una breve explicación de los más comunes, profundizando finalmente en el modelo que se ha utilizado para el desarrollo del proyecto.

5.1.1 Modelos lineales

Los modelos lineales [30] son una herramienta esencial en estadísticas y ciencia de datos, usados ampliamente para predecir una variable dependiente basándose en una o más variables independientes. La premisa básica de la regresión lineal es que la relación entre la variable dependiente y las variables independientes puede ser descrita mediante una línea recta. Estos modelos son particularmente valorados por su simplicidad y claridad interpretativa.

Un modelo de regresión lineal simple se representa con la ecuación 1:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

Donde:

- Y es la variable dependiente, la variable que queremos predecir o explicar.
- X es la variable independiente, la que utilizamos para hacer la predicción.
- β_0 es la intersección o el término constante (el valor de Y cuando X es 0).
- β_1 es la pendiente de la línea, que indica el cambio en Y por un cambio unitario en X .
- ε es el término de error, que representa la variabilidad en Y no explicada por X .

Como se observa en un modelo de regresión lineal simple, únicamente se utiliza una variable explicativa para explicar la variable dependiente. Este tipo de modelos pueden hacerse más complejos añadiendo más variables predictoras, obteniendo lo que se conoce como un modelo de regresión lineal múltiple.

En un modelo de regresión lineal múltiple, cada variable predictora tiene asociado un coeficiente que mide su contribución individual, manteniendo constantes las demás variables. Esto no solo amplía la capacidad del modelo para ajustarse a la complejidad de los datos reales, sino que también facilita la evaluación de interacciones y efectos combinados entre las variables predictoras

Un modelo de regresión lineal múltiple se representa con la ecuación 2.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (2)$$

Donde:

- Y es la variable dependiente, la variable que queremos predecir o explicar.
- X_1, X_2, \dots, X_n son las variables independientes, las que utilizamos para hacer la predicción.
- β_0 es la intersección o el término constante (el valor de Y cuando X_1, X_2, \dots, X_n son 0).
- $\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes de regresión, indican el cambio en Y por un cambio unitario en X_1, X_2, \dots, X_n respectivamente.
- ε es el término de error, representa la variabilidad en Y no explicada por X_1, X_2, \dots, X_n .

5.1.2 Modelos no lineales

Los modelos no lineales [31] ofrecen una capacidad ampliada para representar las relaciones complejas entre variables en comparación con los modelos lineales. Son especialmente útiles cuando la dependencia entre la variable independiente y la variable dependiente no es proporcional o recta, lo que es común en muchos fenómenos naturales.

Un modelo de regresión no lineal se representa con la ecuación 3:

$$Y = f(X, \beta) + \varepsilon \quad (3)$$

Donde:

- f representa una función no lineal que relaciona las variables independientes con la dependiente.
- β son los parámetros del modelo que se estiman durante el ajuste.
- ε es el término de error, representa la variabilidad en Y no explicada por X .

Existen varios tipos de modelos no lineales. Entre los más comunes se encuentran.

Regresión polinómica: extiende la regresión lineal comentada anteriormente, incluyendo términos polinómicos de las variables independientes. Se representa en la ecuación 4.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \dots + \varepsilon \quad (4)$$

Modelos exponenciales: son útiles cuando el crecimiento o decaimiento de Y cambia exponencialmente con X . Se representa en la ecuación 5.

$$Y = \beta_0 e^{\beta_1 X} + \varepsilon \quad (5)$$

Modelos logarítmicos: se utilizan cuando se espera que la relación entre la variable independiente y la variable dependiente aumente o disminuya a una tasa decreciente. Se representa en la ecuación 6.

$$Y = \beta_0 + \beta_1 \log(X) + \varepsilon \quad (6)$$

En los tres modelos:

- Y es la variable dependiente.
- $\beta_0, \beta_1, \dots, \beta_n$ son los parámetros del modelo que se estiman durante el ajuste.
- ε es el término de error, representa la variabilidad en Y no explicada por X .

Cabe destacar que abordar el análisis de datos con modelos no lineales cuando el comportamiento subyacente puede ser caracterizado adecuadamente con modelos lineales puede conducir a complicaciones innecesarias. Utilizar un modelo no lineal en estas circunstancias puede no solo incrementar la complejidad del análisis sino también afectar a la interpretación y la precisión de los resultados. Los modelos no lineales, por su naturaleza, requieren métodos computacionales más complejos para su estimación y pueden ser más sensibles a las especificaciones del modelo y a la calidad de los datos. Esto puede llevar a interpretaciones erróneas o a sobreajustes, especialmente en muestras pequeñas o en datos con mucho ruido.

Para simplificar este problema se utilizan transformaciones para la linealización, convirtiendo un modelo no lineal en uno lineal a través de métodos matemáticos apropiados. Esto se hace para aprovechar la simplicidad y la robustez de los modelos lineales, facilitando la interpretación, el cálculo y la aplicación de técnicas estadísticas bien establecidas.

El ejemplo de transformación más común es la logarítmica. Se utiliza en casos donde las relaciones exponenciales pueden ser transformadas en lineales, como, por ejemplo, modelos de crecimiento exponencial o de decaimiento. La transformación logarítmica se representa en la ecuación 7.

$$\text{Si } Y = \beta_0 e^{\beta_1 X}, \quad \text{entonces } \ln(Y) = \beta_0 + \beta_1 X \quad (7)$$

5.1.3 Modelos de series temporales

Los modelos de series temporales son herramientas estadísticas fundamentales diseñadas para analizar conjuntos de datos que se recopilan en intervalos de tiempo regulares. Estos modelos son esenciales para prever tendencias futuras, basándose en observaciones pasadas. Son ampliamente utilizados en economía, meteorología, finanzas, y otros campos donde las variables se observan a lo largo del tiempo y se desea predecir futuros comportamientos o patrones.

Generalmente se basan en alguno de estos tres conceptos [32]:

- **Estacionalidad:** se refiere a variaciones periódicas en los datos que ocurren a intervalos regulares y predecibles. Esta repetición puede ser diaria, semanal, mensual, o anual, dependiendo del contexto y los datos.
- **Tendencia:** indica un aumento o disminución a largo plazo en los datos. Reconocer e incluir la tendencia en los modelos de series temporales ayuda a entender los cambios generales en los datos a lo largo del tiempo y mejora la precisión del pronóstico.
- **Autocorrelación:** es la correlación de una serie temporal con sus propios valores pasados. Es crucial para identificar la relación interna en los datos y determinar cuánto de los valores pasados puede predecir los valores futuros.

Los tipos de modelos principales dentro de este bloque son:

Modelos autorregresivos (AR): Estos modelos son una forma de análisis de series temporales en las que la variable de interés se predice a partir de sus propios valores pasados. Un modelo AR se representa según la ecuación 8.

$$Y_t = \alpha + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \dots + \delta_k Y_{t-k} + \varepsilon_t \quad (8)$$

Donde:

- Y_t es el valor de la variable explicada en el momento de la predicción.
- α es el termino independiente.
- $\delta_1, \delta_2, \dots, \delta_k$ son los coeficientes autorregresivos.
- $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}$ son las muestras pasadas de la variable explicada.
- ε_t es el residuo o término de error.

El resto de los modelos más utilizados incluyen una media móvil [33] para su construcción y son:

Modelos de media móvil (MA): Los modelos de media móvil son útiles para analizar series temporales que muestran fluctuaciones aleatorias. Este tipo de modelo asume que cada valor en la serie se puede describir como un promedio móvil de errores pasados, que son variaciones aleatorias que no se explican por el modelo. Es efectivo para series donde las observaciones presentan breves y rápidas variaciones que parecen aleatorias.

Modelos autorregresivos de media móvil (ARMA): Los modelos ARMA combinan características de los modelos autorregresivos, que usan valores pasados de la serie para predecir el futuro, con los de media móvil, que incorporan los errores pasados en las predicciones. Son adecuados para series temporales que no exhiben tendencias claras ni comportamientos estacionales, y son útiles para modelar series donde tanto los valores como los errores pasados son importantes para las predicciones.

Modelos autorregresivos integrados de media móvil (ARIMA): Los modelos ARIMA son ideales para trabajar con series temporales que no son estacionarias, es decir, series cuyas propiedades estadísticas como la media o la varianza cambian con el tiempo. Estos modelos primero transforman la serie a una forma estacionaria diferenciándola, y luego aplican un modelo ARMA para hacer predicciones.

5.1.4 Modelos de redes neuronales artificiales (ANN)

Una red neuronal artificial [34] es un sistema que intenta emular la forma en que el cerebro humano procesa la información. Consiste en varias unidades de procesamiento dispuestas en capas: una capa de entrada que recibe datos, varias capas ocultas que procesan los datos, y una capa de salida que genera la predicción o el resultado final.

Las unidades en estas redes, que se asemejan a neuronas abstractas, están interconectadas por conexiones ponderadas, conocidas como pesos. Estos pesos se ajustan durante el proceso de entrenamiento de la red, que implica múltiples iteraciones sobre un conjunto de datos. En cada iteración, la red realiza predicciones, compara sus resultados con los valores verdaderos y ajusta los pesos para mejorar las predicciones futuras.

Inicialmente, los pesos son aleatorios y las predicciones de la red pueden no tener sentido. Sin embargo, a medida que se entrena la red con ejemplos cuyos resultados son conocidos, las ponderaciones se ajustan en función de los errores en las predicciones. Este proceso se conoce como aprendizaje mediante retro propagación. Al final del entrenamiento, la red habrá afinado sus pesos a un punto en el que puede generalizar bien a nuevos datos, proporcionando predicciones precisas basadas en la entrada que recibe.

En la ilustración 1 se puede observar estructura de un modelo de red neuronal.

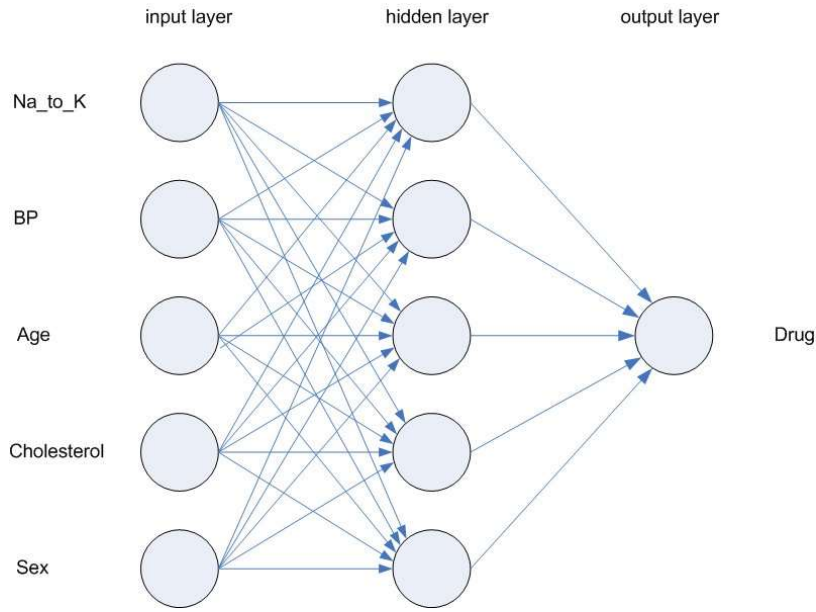


Ilustración 1. Estructura de una red neuronal [34]

5.1.5 Modelos ARX

El presente trabajo utiliza modelos ARX lineales. Los modelos ARX lineales son una clase de modelos estadísticos que se utilizan para describir sistemas dinámicos en los que la variable de interés es predicha no solo por variables externas, sino también por sus valores pasados que influyen en su comportamiento.

Los modelos ARX lineales representan una síntesis avanzada de dos conceptos fundamentales en estadística y control de sistemas: los modelos autorregresivos (AR) representados en la ecuación 8, y los modelos de regresión lineal múltiple (ecuación 2), también conocidos como modelos lineales con variables exógenas. Esta combinación permite capturar tanto la dependencia temporal interna de una serie como la influencia de variables externas que impactan en el sistema estudiado. Un modelo ARX se representa mediante la ecuación 9.

$$\begin{aligned}
 Y_t = & \alpha + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \dots + \delta_k Y_{t-k} + \\
 & \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_k X_{k,t} + \\
 & \mu_1 X_{1,t-1} + \mu_2 X_{2,t-1} + \dots + \mu_k X_{k,t-1} + \\
 & \gamma_1 X_{1,t-k} + \gamma_2 X_{2,t-k} + \dots + \gamma_k X_{k,t-k} + \varepsilon_t
 \end{aligned}
 \tag{9}$$

Donde:

- Y_t es el valor de la variable explicada en el momento de la predicción.
- α es el termino independiente.
- $\delta_1, \delta_2, \dots, \delta_k$ son los coeficientes autorregresivos.
- $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}$ son las muestras pasadas de la variable explicada.
- $X_{1,t}, X_{2,t}, \dots, X_{k,t}$ son las variables independientes y $X_{1,t-1}, X_{1,t-2}, \dots, X_{k,t-k}$ sus muestras pasadas
- β, μ y γ son los coeficientes de las variables independientes y sus muestras pasadas

5.2 EVALUACIÓN DE MODELOS Y MÉTRICAS DE ERROR

La evaluación de modelos consiste en verificar cuán bien un modelo realiza predicciones, y es importante para garantizar que el modelo sea generalizable más allá del conjunto de datos con el que fue entrenado. Para llevar a cabo esta evaluación, se utilizan diversas métricas de error que cuantifican de diferentes maneras la precisión de las predicciones del modelo.

5.2.1 Coeficiente de determinación (R^2)

R^2 es un indicador de cuánta variabilidad de la variable dependiente puede ser explicada por el modelo. Es la métrica de error más reconocida en la comunidad científica para validar la bondad de un modelo. Sigue la ecuación 10.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (10)$$

Donde:

- \bar{y} es la media de la muestra.
- \hat{y}_i son cada una de las predicciones de la variable explicada o que se quiere predecir.
- y_i son los valores reales de la variable explicada o que se quiere predecir.
- SSR es la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo.
- SST es la suma de los cuadrados de las diferencias entre los valores observados y el promedio de los valores observados.

5.2.2 Coeficiente de determinación ajustado (R_{ajustado}^2)

El R_{ajustado}^2 ajusta el R^2 para el número de predictores en el modelo, ofreciendo una medida más precisa de la calidad del ajuste cuando se añaden variables al modelo, ya que se penaliza la introducción de variables cuando no son significativas. Se calcula ajustando los grados de libertad para la variabilidad, como se representa en la ecuación 11:

$$R_{\text{ajustado}}^2 = 1 - \frac{SCR/(n - k)}{SCT/(n - 1)} \quad (11)$$

Aquí, n es el número total de observaciones y k es el número de variables predictoras. Este ajuste es fundamental cuando se comparan modelos con diferentes números de predictores.

Tanto el R^2 como el ajustado oscila entre 0 (nulo ajuste) y 1 (ajuste perfecto).

5.2.3 Raíz del error medio cuadrático (RMSE)

El ECM es una medida de la calidad del ajuste que toma en cuenta la magnitud de los errores del modelo, penalizando más los errores grandes debido al cuadrado de las diferencias.

Es especialmente útil porque magnifica el impacto de errores más grandes. Esto es crucial en contextos donde los errores grandes son más costosos o tienen consecuencias más graves. Por ejemplo, en la gestión financiera o en la salud, un modelo que ocasionalmente falla de manera significativa puede ser más peligroso que uno cuyos errores son más frecuentes, pero menos severos.

Además, es una métrica que permite comparar modelos, siempre y cuando compartan la misma variable dependiente.

Se calcula con la ecuación 12.

$$RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{(n - k)}} \quad (12)$$

Donde:

- \bar{y} es la media de la muestra.
- y_i son los valores reales de la variable explicada o que se quiere predecir.
- \hat{y}_i son cada una de las predicciones de la variable explicada o que se quiere predecir.

5.2.4 Error medio absoluto (MAE)

El MAE es otra métrica complementaria útil. A diferencia del ECM, el MAE mide el promedio de los errores absolutos sin elevarlos al cuadrado, ofreciendo una perspectiva más tolerante a los errores atípicos.

Es muy útil ya que es una métrica fácil e intuitiva de interpretar, puesto que representa el error medio directamente en las unidades de la variable que se quiere predecir. Además, El MAE es menos sensible a los outliers que el RMSE, lo que lo hace ideal para modelos en aplicaciones donde los outliers no deben influir de manera desproporcionada en la evaluación del modelo.

Al igual que con el RMSE, esta métrica nos permite comparar modelos siempre y cuando tengan la misma variable explicada.

Se calcula con la ecuación 13.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (13)$$

Donde:

- y_i son los valores reales de la variable explicada o que se quiere predecir.
- \hat{y}_i son cada una de las predicciones de la variable explicada o que se quiere predecir.

5.3 MÉTODOS DE CONSTRUCCIÓN DE MODELOS

5.3.1 Métodos

En la construcción de un modelo de regresión, dos enfoques predominan: la extensión y la reducción por pasos. Cada método ofrece una estrategia sistemática para determinar qué variables predictoras incluir en el modelo final.

Método de extensión por pasos: este enfoque comienza con un modelo simple, que incluye solo una variable predictora. Gradualmente, se añaden variables adicionales una por una. Después de añadir cada nueva variable, se evalúa su contribución a la precisión y la capacidad explicativa del modelo. Si la inclusión de la nueva variable mejora el modelo significativamente, se conserva y se procede a añadir otra variable. Este proceso se repite hasta que no se encuentran más mejoras significativas al añadir nuevas variables.

Método de reducción por pasos: este método adopta una aproximación opuesta. Se inicia con un modelo que incluye todas las variables predictoras potenciales. Luego, se eliminan las variables de forma secuencial, evaluando la importancia de cada una basándose en su significatividad estadística. Después de cada eliminación, se analiza cómo afecta la capacidad predictiva del modelo. Si la eliminación de una variable no deteriora el modelo de manera significativa, se considera que la variable no es relevante y se procede con la eliminación de la siguiente. Este proceso continúa hasta que todas las variables restantes son consideradas esenciales para el modelo.

Para llevar a cabo estos dos métodos uno de los criterios más utilizado es el criterio de Akaike.

Criterio Akaike: el criterio de información de Akaike (*AIC, Akaike Information Criterion*) [35] es una herramienta estadística utilizada para la selección de modelos en estadística. Su principal función es evaluar la calidad de cada modelo estadístico para un conjunto dado de datos y, al mismo tiempo, considerar la complejidad del modelo.

El AIC se basa en el concepto de entropía de la información, una medida de la cantidad de información perdida cuando se usa un modelo específico para representar el proceso que genera los datos. El criterio ofrece un equilibrio entre la bondad del ajuste del modelo y la simplicidad de este, penalizando la complejidad para evitar el sobreajuste.

La fórmula para calcular el AIC es la representada en la ecuación 14:

$$AIC = 2k - 2\ln(L) \quad (14)$$

Donde:

- k es el número de parámetros del modelo
- L es la verosimilitud máxima del modelo, es decir, la probabilidad de que el modelo haya predicho los datos observados, dados los valores estimados de los parámetros.

5.3.2 Significatividad de las variables

La significatividad de las variables en un modelo estadístico o de regresión es esencial para entender qué factores tienen un impacto real sobre la variable dependiente. Analizar la significatividad de cada variable permite juzgar si una variable aporta información útil y es estadísticamente significativa para hacer predicciones o para explicar variabilidad en los datos.

En el contexto de un modelo de regresión, cada variable independiente tiene asociado un coeficiente que cuantifica su relación con la variable dependiente. La significatividad de estos coeficientes se evalúa comúnmente mediante pruebas estadísticas, como la prueba t de Student, que determina si un coeficiente es significativamente diferente de cero.

La prueba t para un coeficiente en un modelo de regresión evalúa la siguiente hipótesis nula:

$$\begin{aligned} H_0: \beta_i &= 0 \\ H_a: \beta_i &\neq 0 \end{aligned} \quad (15)$$

β_i es el coeficiente de la variable independiente. Si la hipótesis nula es rechazada (generalmente, si el p -valor asociado con la prueba t es menor que el nivel de significancia, como 0,05), entonces se concluye que hay suficiente evidencia estadística para afirmar que la variable correspondiente tiene un efecto significativo sobre la variable dependiente.

Las variables que no son estadísticamente significativas sugieren que no están contribuyendo a explicar la variabilidad observada en la variable dependiente, que están mal especificadas, o que no hay suficientes datos para demostrar su impacto. Estas variables pueden ser candidatas para excluir del modelo para simplificarlo y mejorar la precisión de este.

6.1 METODOLOGÍA

El presente trabajo se ha desarrollado en dos fases distintas debido a los descubrimientos realizados durante el análisis inicial. La primera fase se ha centrado en la mejora de los modelos existentes proporcionados por Iñigo García de Eulate [28]. Sin embargo, a medida que se han realizado pruebas y se han analizado los resultados, han surgido conclusiones que han llevado a replantear ciertos aspectos dados por correctos en el trabajo previo. Esto ha dado lugar a una segunda fase metodológica, en la que se ha creado un nuevo enfoque para la predicción, cambiando la variable explicada de temperatura interior a carga térmica, tal y como se explicará más adelante.

6.1.1 Primera fase: mejora de los modelos de referencia

El objetivo principal de esta primera fase es analizar y mejorar los modelos ARX desarrollados previamente por el proyecto de referencia. El equipo de investigación de la Universidad de Deusto estaba interesado en continuar con algunas líneas de investigación abiertas en el proyecto, entre ellas, la ampliación del horizonte temporal de predicción de sus modelos y la reducción de parámetros de estos. Por tanto, la primera idea del presente proyecto ha sido realizar ese estudio e implantación, optimizando los modelos existentes al evaluar la significancia de las variables y mejorar la precisión predictiva mediante ajustes en los parámetros del modelo y reducción de variables.

Paso 1.1. Análisis de los modelos existentes: se han analizado los modelos ARX proporcionados por Iñigo García de Eulate. Esto ha incluido la evaluación de las variables utilizadas en los modelos para identificar aquellas que no sean significativas y estudiar su comportamiento con respecto a la predicción de la temperatura interior. Se llevarán a cabo pruebas de rendimiento utilizando métricas como el error medio absoluto (MAE).

Paso 1.2. Reducción de parámetros de los modelos: se ha procedido a ajustar los modelos existentes. Esto ha implicado optimizar los modelos reduciendo el número de variables utilizadas, ajustando los parámetros del modelo y explorando diferentes configuraciones para mejorar su capacidad predictiva.

Paso 1.3. Validación de los modelos mejorados: se han validado los modelos ajustados para determinar si se puede mejorar la precisión utilizando menos variables o ajustando los parámetros del modelo. Se han comparado las predicciones del modelo con los datos reales para evaluar su precisión y fiabilidad utilizando métricas como el R^2_{ajustado} y el MAE.

6.1.2 Segunda fase: creación de nuevos modelos

Debido a las conclusiones obtenidas en la primera fase, se ha replanteado el problema original. En lugar de predecir la temperatura interior, el nuevo enfoque se ha centrado en la predicción de la carga térmica de la habitación. Este cambio ha permitido una mejor optimización y control de los sistemas de climatización, abordando directamente la demanda energética del edificio.

Paso 2.1. Redefinición del problema: a partir de las conclusiones obtenidas en la primera fase, se ha decidido replantear el problema. En lugar de predecir la temperatura interior, el nuevo enfoque ha sido predecir la carga térmica de la habitación.

Paso 2.2. Preprocesado de datos: se ha realizado un nuevo preprocesado de los datos para adecuarlos a la predicción de la carga térmica. Esto ha incluido la transformación de variables y la agrupación de datos en intervalos de tiempo adecuados.

Paso 2.3. Ajuste de nuevos modelos: se han seleccionado y ajustado nuevos modelos ARX utilizando los datos preprocesados. Se han explorado diferentes configuraciones y se han ajustado los parámetros del modelo para optimizar su capacidad predictiva de la carga térmica.

Paso 2.4. Evaluación de los nuevos modelos: una vez ajustados los nuevos modelos, se ha evaluado su rendimiento utilizando métricas adecuadas como el MAE y el RMSE. Se han comparado las predicciones del modelo con los datos reales para determinar su precisión y fiabilidad.

Paso 2.5. Análisis de los residuos: se ha llevado a cabo un análisis de los residuos arrojados por el modelo desarrollado para detectar y comprender patrones de fallo y posibles soluciones para mejorar el ajuste del modelo.

6.2 HERRAMIENTA SOFTWARE

En este apartado, se presentan y comparan las cuatro principales alternativas de software que se han valorado. Todas ellas son ampliamente utilizadas en análisis de datos y modelado estadístico: Excel, IBM SPSS, R y MATLAB. Estas herramientas ofrecen capacidades variadas que pueden adaptarse a diferentes necesidades de investigación y análisis en el ámbito de la predicción energética y otros estudios similares.

6.2.1 Excel



Excel es una de las herramientas más populares y ampliamente utilizadas en el mundo para análisis de datos debido a su interfaz de usuario amigable y su integración accesible con otros productos de Microsoft Office. Ofrece varias características que lo hacen apto para una variedad de tareas analíticas y de modelado en campos menos técnicos o para análisis exploratorios rápidos. A continuación, se detallan algunos de las principales ventajas y limitaciones de Excel para su uso en proyectos estadísticos y de predicción.

Ventajas:

- **Accesibilidad y facilidad de uso:** Excel es conocido por su interfaz intuitiva. La mayoría de los usuarios están familiarizados con su entorno, lo que reduce la curva de aprendizaje para realizar análisis básicos y manipulaciones de datos.
- **Funcionalidades básicas de análisis de datos:** ofrece una gama de funciones estadísticas y analíticas integradas, como análisis de tendencias y proyecciones, que son suficientes para muchas necesidades empresariales y educativas básicas.

- **Integración con otros softwares de Microsoft:** la integración perfecta con otras aplicaciones de Microsoft como Word y PowerPoint facilita la presentación y el intercambio de datos.

Limitaciones:

- **Capacidad de manejo de datos:** Excel tiene limitaciones en cuanto a la cantidad de datos que puede manejar eficientemente. Los grandes conjuntos de datos pueden ralentizar o incluso colapsar el programa, lo que lo hace menos adecuado para análisis de datos a gran escala.
- **Análisis avanzado:** las capacidades estadísticas de Excel son básicas en comparación con las de software especializado como R, SPSS, o MATLAB. Carece de la capacidad para realizar análisis más complejos requeridos en investigaciones avanzadas sin recurrir a plugins o extensiones externas.
- **Reproducibilidad:** repetir los pasos o replicar los análisis puede ser complicado, ya que la documentación de las operaciones y manipulaciones de datos no es tan estructurada o clara como en otros entornos de programación.

6.2.2 IBM SPSS



IBM SPSS es un software avanzado de análisis estadístico ampliamente utilizado en investigación académica, salud, empresas, gobierno y casi cualquier industria que requiera análisis complejos de datos. Su facilidad de uso, junto con profundas capacidades analíticas, lo hacen ideal para investigadores que necesitan realizar análisis sofisticados sin una profunda experiencia en programación. A continuación, se examinan las características principales, ventajas y limitaciones de IBM SPSS

Ventajas:

- **Interfaz gráfica de usuario:** SPSS es famoso por su interfaz amigable con el usuario, que facilita la realización de análisis complejos a través de menús y cuadros de diálogo, lo que permite a los usuarios menos técnicos utilizar el software de manera efectiva.
- **Amplia gama de funcionalidades estadísticas:** SPSS ofrece una extensa variedad de técnicas estadísticas incluyendo análisis de regresión, correlaciones, pruebas de significancia, y más. Además, soporta análisis multivariante, análisis de supervivencia y modelado complejo.
- **Gestión de datos:** SPSS permite manejar grandes conjuntos de datos y ofrece herramientas robustas para la manipulación y transformación de datos, incluyendo la capacidad de manejar datos faltantes de manera avanzada.

Limitaciones:

- **Costo:** uno de los mayores inconvenientes de SPSS es su precio, que puede ser prohibitivo especialmente para individuos que no tienen el presupuesto para licencias de software costosas.
- **Flexibilidad en análisis avanzados:** aunque SPSS es muy potente, se pueden encontrar limitaciones en su flexibilidad comparado con software de código abierto como R, especialmente cuando se trata de análisis muy específicos o personalizados.

6.2.3 MATLAB



MATLAB es un entorno de software y lenguaje de programación desarrollado por MathWorks, ampliamente utilizado en ingeniería, matemáticas, ciencias, y educación. Es especialmente valorado por su capacidad para realizar cálculos numéricos y análisis de datos complejos, simulación, y el modelado de sistemas dinámicos. A continuación, se describen las características principales, ventajas y limitaciones de MATLAB:

Ventajas:

- **Capacidad de cálculo y simulación:** MATLAB es altamente eficaz en el manejo de matrices y operaciones matemáticas complejas, lo que lo hace ideal para simulaciones numéricas y cualquier tipo de análisis matemático.
- **Toolboxes especializadas:** ofrece una amplia variedad de toolboxes especializadas que se pueden adquirir para expandir las capacidades del programa. Estas incluyen herramientas para señales y sistemas, control automático, aprendizaje de máquinas, y muchas otras áreas.
- **Visualización de datos avanzada:** proporciona extensas opciones para la visualización de datos que facilitan la interpretación de información compleja y el análisis de resultados.
- **Integración con dispositivos y hardware:** capaz de conectarse y controlar instrumentación externa, lo que es útil en entornos de prueba y medida, así como en la investigación aplicada.

Limitaciones:

- **Costo:** similar a SPSS, MATLAB puede ser bastante caro, especialmente cuando se agregan múltiples toolboxes especializadas, lo que puede ser prohibitivo para algunos usuarios con presupuestos limitados.
- **Curva de aprendizaje:** aunque MATLAB es menos técnico que algunos lenguajes de programación, aprender a utilizarlo eficientemente aún presenta una curva de aprendizaje, particularmente para quienes no tienen experiencia en programación.
- **Aplicaciones especializadas:** a pesar de su potencia y versatilidad en el manejo de cálculos numéricos, MATLAB es predominantemente utilizado en contextos de simulación y diseño de controladores más que en estadística pura. Su entorno y sus herramientas están optimizados para aplicaciones de ingeniería donde la modelización dinámica y las simulaciones controladas son cruciales. Esta orientación lo hace ideal para prototipos y desarrollo de algoritmos en sistemas controlados, pero menos conveniente para tareas de estadística descriptiva o inferencia estadística.

6.2.4 R



R es un entorno de programación y lenguaje de código abierto específicamente diseñado para análisis estadístico y visualización gráfica. Es ampliamente reconocido en el ámbito académico y profesional por su versatilidad y potencia en el manejo de datos estadísticos, investigación ingeniería, finanzas y muchas otras áreas de investigación científica y análisis de datos. A continuación, se detallan las principales ventajas y limitaciones de R.

Ventajas:

- **Código abierto y gratuito:** uno de los mayores beneficios de R es que es completamente gratuito, lo que facilita su acceso a cualquier usuario sin restricciones de licencia.
- **Extensa comunidad de soporte:** R tiene una comunidad muy activa y colaborativa que contribuye con una extensa gama de paquetes que extienden sus funcionalidades, cubriendo desde análisis estadístico avanzado hasta aprendizaje automático y visualización de datos.
- **Flexibilidad y potencia:** R es excepcionalmente flexible y potente para análisis estadísticos y modelado de datos. Permite la implementación de los últimos métodos estadísticos y algoritmos personalizados, haciendo que sea muy apreciado en el ámbito de la investigación.
- **Integración de datos y gráficos avanzados:** ofrece capacidades robustas para la manipulación y visualización de datos. Los usuarios pueden crear gráficos avanzados y personalizados que son difíciles de generar en otras plataformas.

Limitaciones:

- **Curva de aprendizaje:** aunque R es muy poderoso, su sintaxis puede ser inicialmente intimidante para los nuevos usuarios, especialmente para aquellos sin experiencia previa en programación.
- **Documentación variable:** Aunque hay mucha documentación disponible, la calidad puede ser inconsistente debido a la naturaleza de código abierto y la gran cantidad de contribuyentes.

En la tabla 1 se puede observar una comparativa general entre los cuatro softwares.





VENTAJAS		LIMITACIONES
EXCEL 	Facilidad de uso Funcionalidades básicas de análisis Integración con softwares Microsoft	Capacidad de manejo de datos Análisis avanzado Reproducibilidad
IBM SPSS 	Interfaz gráfica de usuario Funcionalidades estadísticas Gestión de datos	Costo Flexibilidad en análisis avanzados
MATLAB 	Capacidad de cálculo y simulación Toolboxes especializadas Visualización de datos avanzada Integración con dispositivos	Costo Curva de aprendizaje Aplicaciones especializadas
R 	Código abierto y gratuito Extensa comunidad de soporte Flexibilidad y potencia Integración datos y gráficos avanzados	Curva de aprendizaje Documentación variable

Tabla 1. Comparativa entre software para análisis estadístico y predicción de la carga térmica

Tras una evaluación detallada de las herramientas disponibles, R ha sido elegido por su combinación única de accesibilidad y flexibilidad. Su naturaleza gratuita y de código abierto reduce las barreras económicas, mientras que su extenso repertorio de paquetes permite realizar análisis complejos y personalizados de manera eficiente. Además, la capacidad de R para realizar transformaciones complejas de datos de forma sencilla y efectiva permite adaptar los datasets a las necesidades específicas del análisis de carga térmica, facilitando la limpieza, normalización y preparación de los datos para el modelado predictivo. Por último, la amplia variedad de paquetes para la creación de gráficos detallados y personalizados permite la interpretación visual de los resultados para entender el comportamiento de los modelos creados.

6.3 EDIFICIO Y BASE DE DATOS

6.3.1 Edificio

El estudio se apoya en el conjunto de datos ROBOD [29] (Room-level Occupancy and Building Operation Dataset) del edificio SDE4 de la Escuela de Diseño y Medioambiente de la Universidad Nacional de Singapur. Este conjunto de datos es fundamental, ya que proporciona mediciones detalladas de variables críticas recogidas en cinco diferentes espacios del edificio, las cuales incluyen temperatura interior y exterior, humedad, niveles de CO₂, energía consumida por diferentes sistemas y datos de ocupación.

A continuación, se adjunta una imagen donde se puede ver la localización geográfica de Singapur en el mapa del mundo con el fin de ilustrar al lector.



Ilustración 2. Ubicación de Singapur en el mapa del mundo

Es fundamental considerar la ubicación geográfica del edificio en Singapur, situado cerca del ecuador donde las condiciones climáticas son predominantemente ecuatoriales. Esto implica que las temperaturas suelen ser altas y estables a lo largo del año, con poca variación estacional, lo que resulta en una demanda constante y elevada de energía para aire acondicionado. Esto es interesante ya que cuando observemos el listado de variables, la variable relacionada con la climatización corresponderá a “energía consumida por agua refrigerada” pero no encontraremos ninguna relacionada con la calefacción.

6.3.2 Base de datos

En la tabla 2, puede verse la información acerca de cada una de las habitaciones que forman parte del conjunto de datos: su uso, el tipo de ocupante, la planta del edificio en la que se encuentra, la superficie, la capacidad máxima, y la densidad máxima de ocupantes medida en metros cuadrados por ocupante.

Room	Space function	Occupant type	Level	Floor area [m ²]	Floor to ceiling height [m]	Room volume [m ³]	Seating capacity [person]	Maximum occupant density [m ² /person]
Room 1	Lecture room	Students	4	118.6	4.1	486.2	40	3.0
Room 2	Lecture room	Students	4	53.7	4.1	220.2	40	1.3
Room 3	Office space	Administrative staff	5	98.4	4.2	413.2	15	6.6
Room 4	Office space	Researchers	3	141.9	4.1	581.7	25	5.6
Room 5	Library space	Students	2	182.8	7.5	1363.3	36	5.0

Tabla 2. Información general sobre las habitaciones medidas [29]

- Las **habitaciones 1 y 2** funcionan como aulas, y su tipo de ocupante más común son los estudiantes. Ambas aulas se encuentran en el cuarto piso del edificio y cuentan con capacidad para albergar a 40 personas sentadas. Sin embargo, en cuanto a superficie, la habitación 1 es más del doble de grande que la habitación 2.
- La **habitación 3** se utiliza como oficina y está destinada al personal administrativo. Ubicada en el quinto piso del edificio, tiene capacidad para 15 personas sentadas. En cuanto a superficie, es prácticamente el doble de grande que la habitación 2.
- La **habitación 4**, al igual que la habitación 3, está destinada a uso de oficina, pero específicamente para investigadores. Tiene capacidad para albergar hasta 25 personas sentadas y es la sala más grande del edificio, a excepción de la biblioteca, con una superficie de 141,9 m² y un volumen total de 581,7 m³. Está ubicada en el tercer piso del edificio.
- La **habitación 5**, que se utiliza como biblioteca, es la más grande de las cinco habitaciones disponibles en la base de datos. Está ubicada en el segundo piso del edificio y está destinada principalmente al uso de los estudiantes.

Las variables disponibles en el conjunto se clasifican en 6 grandes grupos: calidad medioambiental interior, Wi-Fi, energía, operativa HVAC, climatología exterior y ocupación.

Data category	Measured variable	Data unit	Sensor type (Brand)	Sensor range	Sensor accuracy
Indoor environmental quality	VOC	ppb	IAQ unit (Awair Omni)	0–60000	±10%
	Sound pressure	dB(A)		Unspecified	Unspecified
	Relative humidity	%RH		0–100	±2%RH
	Air temperature	°C		–40 to 125	±0.2 °C
	Illuminance	lux		0–64000	Unspecified
	PM _{2.5}	µg/m ³		0–1000	±15 µg/m ³
	CO ₂	ppm		400–5000	±75 ppm
Wi-Fi	Wi-Fi connected devices	Number	Wi-Fi router (Cisco)	NA ^a	NA ^a
Energy	Ceiling fan energy	kWh	Energy meter (Schneider Electric Acti9 iEM3000)	0–999999	±1%
	Lighting energy				
	Plug load energy				
	Chilled water energy		BTU meter (Integra Metering CALEC ST II)		± 2%
	AHU/FCU fan energy		Energy meter (Schneider Electric PM5300)		± 0.5%
HVAC operations	Supply airflow ^b	CMH	VAV box (Johnson Controls)	0–3375	±15%
	Damper position ^b	%		0–100	NA ^a
	Temperature setpoint	°C	NA ^a	NA ^a	NA ^a
	Cooling coil valve position ^b	%	Valve (Johnson Controls)	0–100	NA ^a
	Cooling coil valve command ^b				
	AHU/FCU fan speed	Hz	Variable speed drive (ABB)	0–50	±0.2%
	Offcoil air temperature ^b	°C	NTC thermistor (Greystone TSDC)	–40 to 60	±0.2°C
	Offcoil temperature setpoint ^b	°C	NA ^a	NA ^a	NA ^a
	Supply air humidity ^b	%RH	Capacitive (Greystone HSDT)	0–100	±2%RH
	Pressure across filter ^b	Pa	Capacitive (Setra 264)	Unspecified	±1%
	Supply air pressure				
Outdoor weather ^c	Supply air temperature	°C	NTC thermistor (Greystone TSAP)	–40 to 60	±0.2°C
	Barometric pressure	hPa	Piezoresistive	600–1100	±0.5hPa
	Dry bulb temperature	°C	Pt100	–40 to 60	±15 °C
	Global horizontal solar radiation	W/m ²	Thermophile	0–2000	2 nd class pyranometer
	Wind direction	Degree	Ultrasonic	0–360	±15 RMSE
	Wind speed	m/s		0–60	±0.2 m/s
	CO ₂	ppm	Non-dispersive infrared	0–2000	±5 ppm + 2%
	Relative humidity	%RH	Capacitive	0–100	±1.5%RH
Occupancy	Occupant presence	Binary (1/0)	Surveillance camera (Xeron Vision)	NA ^a	NA ^a
	Occupant count	Number			

^aNA refers to “Not Applicable”.

^b Indicated measurements are not applicable for Room 1 and Room 2.

^c All the outdoor weather data were collected by a weather station (Delta OHM HD52.3D).

Tabla 3. Variables disponibles para cada habitación [29]

Para facilitar la comprensión, las variables se explican brevemente a continuación. Es importante destacar que todas las variables son cuantitativas y continuas, excepto el número de dispositivos Wi-Fi conectados y las dos variables de consigna. Además, se resaltan aquellas variables que, por su significancia, serán explicadas en detalle más adelante.

Grupo 1, calidad medioambiental interior:

- Concentración de compuestos orgánicos volátiles (VOC, *volatile organic compounds*) medida en partes por billón (ppb).
- Nivel de presión sonora medido en decibelios de tipo A (dB(A)).
- Humedad relativa medida en porcentaje (%).
- **Temperatura del aire medida en grados Celsius (°C).**
- Iluminancia medida en lux.
- Materia particulada 2,5 (PM_{2.5}) medida en microgramos por metro cúbico (µg/m³).
- Concentración de dióxido de carbono (CO₂) medido en partes por millón (ppm).

Grupo 2, Wi-Fi, cuya única variable es dispositivos conectados al Wi-Fi.

Grupo 3, energía:

- Energía de los ventiladores del techo (kWh).
- Energía de iluminación (kWh).
- Energía de enchufes (kWh).
- **Energía de agua refrigerada (kWh).**
- Energía del ventilador AHU/FCU (kWh).

Grupo 4, operativa HVAC:

- Suministro de flujo de aire medido en metro cúbicos por hora (CMH, *cubic meters per hour*).
- Posición de la compuerta medida en porcentaje (%).
- **Consigna de temperatura medida en grados Celsius (°C).**
- Posición de la válvula de la bobina de refrigeración medida en porcentaje (%).
- Comando de la válvula de la bobina de refrigeración medido en porcentaje (%).
- Velocidad del ventilador AHU/FCU medida en hercios (Hz).
- Temperatura de salida del aire después de la bobina de refrigeración medida en grados Celsius (°C).
- Consigna de temperatura de salida del aire después de la bobina de refrigeración medida en grados Celsius (°C).
- Humedad del suministro de aire medido en porcentaje (%).
- Presión a través del filtro medida en pascales (Pa).
- Presión del suministro de aire medida en pascales (Pa).
- Temperatura del suministro de aire medida en grados Celsius (°C).

Grupo 5, climatología exterior:

- Presión barométrica medida en hectopascales (hPa).
- Temperatura de bulbo seco medida en grados Celsius (°C).
- **Radiación solar global horizontal medida en vatios por metro cuadrado (W/m²).**
- Dirección del viento medida en grados.
- Velocidad del viento medida en metros por segundo (m/s).
- Concentración de dióxido de carbono (CO₂) medido en partes por millón (ppm).
- Humedad relativa medida en porcentaje (%).

Grupo 6, ocupación:

- Presencia de ocupantes medida en binario: 1 significa presencia de ocupantes y 0 ausencia de ocupantes.
- **Conteo de ocupantes.**

6.3.3 Habitación seleccionada

En el desarrollo de este proyecto, se ha utilizado específicamente información de una de las habitaciones del conjunto de datos total, con el objetivo de analizar la evolución térmica de una zona particular donde se instalará el controlador. La selección de la habitación se basó en criterios clave para simplificar el modelo y asegurar la replicabilidad del estudio. Se optó por una habitación con una sola exposición al exterior, lo cual facilita el cálculo de la radiación solar. Además, se buscó un espacio de tamaño considerable para representar adecuadamente un ambiente de oficina típico.

La habitación seleccionada, la número 4, está situada en la tercera planta del edificio y es utilizada por investigadores. Con una superficie de 141.9 m² y una altura de 4.1 metros, es la segunda más grande de las cinco habitaciones del edificio de las cuales se dispone de información. Esta habitación tiene un aforo máximo de 25 personas sentadas, lo cual es relevante para modelar la carga térmica interna generada por los ocupantes. Interesantemente, aunque tiene una exposición única, la habitación posee dos zonas orientadas hacia el exterior, con la zona suroeste siendo la más amplia, como se ilustra en las visualizaciones proporcionadas. Esta configuración la convierte en un caso de estudio ideal para desarrollar y validar el modelo de predicción térmica que se implementará en el sistema de control del edificio.

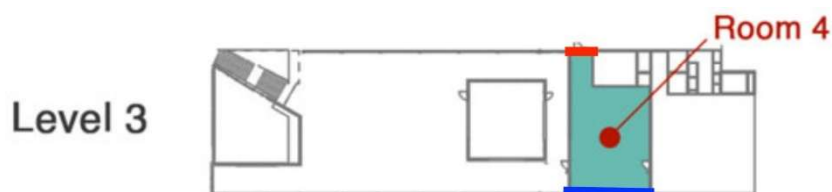


Ilustración 3. Plano de la habitación 4

6.4 PRIMERA FASE: MEJORA DE MODELOS DE REFERENCIA

6.4.1 Modelo de referencia

Se va a tomar como trabajo de referencia el proyecto de fin de grado realizado por Iñigo García de Eulate. En este trabajo se realizaron una serie de importantes avances a partir de los cuales se pueden desarrollar mejoras e innovaciones relevantes. En concreto, hay tres aspectos de ese trabajo que, en un primer momento, no se han cuestionado y sobre los que se van a empezar a implantar mejoras. Estos aspectos son:

6.4.1.1 Elección de las variables y preprocesado de datos:

El proceso de preprocesado de datos empleado en los modelos de referencia ha sido revisado y validado por el equipo de investigación de Deusto y por el autor del presente trabajo. Este procedimiento incluye un análisis de correlaciones y una selección de variables que han sido fundamentales para identificar los factores clave que influyen en la temperatura interna. Además, este proceso abarca la identificación y el tratamiento adecuado de valores atípicos, la normalización de los datos y la eliminación de redundancias.

En concreto, se consideran como variables relevantes para explicar la evolución de la temperatura interior en una habitación las siguientes, justificadas desde el punto de vista físico:

- **Temperatura exterior:** el intercambio de calor entre el interior y el exterior de un edificio, a través de elementos como paredes, ventanas y techos, afecta significativamente a la temperatura interna.
- **Radiación solar:** la energía solar que incide sobre el edificio, especialmente en las ventanas, eleva la temperatura de estas y, en consecuencia, del interior del edificio.
- **Energía consumida por la climatización del edificio:** la cantidad de energía utilizada para la climatización del edificio es un factor determinante en la predicción de la temperatura interna, particularmente cuando se cuenta con información adicional de otras variables. En este contexto, se considera la energía de agua refrigerada, ya que la localización climática del edificio solo requiere refrigeración, no calefacción.
- **Ocupación por conteo:** esta variable proporciona al modelo una indicación del aumento de la temperatura interior debido al calor generado por los ocupantes de la habitación y su transferencia al aire circundante.

6.4.1.2 Modificación y transformación del *dataframe*

Se trabaja con el *dataframe* modificado en el proyecto de referencia, donde se realizaron procesos de limpieza y transformación de algunas variables. En particular, se estableció una media móvil en la variable de ocupantes por conteo para obtener una variable más limpia y representativa de la ocupación real. Por otro lado, se transformó la radiación global horizontal a la radiación real en la habitación, ajustándola al ángulo real del sol con respecto al edificio.

La primera modificación se realizó porque la variable de ocupación por conteo se presentaba poco robusta. Para mejorar la robustez y estabilidad de la señal se incorporó una media móvil de 3 elementos que derivó en una señal más limpia.

La segunda modificación se hizo al observar que la radiación solar máxima no coincidía con el mediodía. Esto se debe a que Singapur, al ser un microestado, utiliza el huso horario de los países con los que tiene fuertes relaciones políticas, económicas, etc. no coincidiendo con la hora solar que le correspondería. Este desfase se corrigió utilizando las ecuaciones propuestas en el artículo *Solar Engineering of Thermal Processes* de Duffie y Beckman [36].

Por último, se realizó una transformación para ajustar la radiación solar medida en plano horizontal y reflejar con precisión la radiación que afecta específicamente a la habitación 4 del edificio. Inicialmente, se calculó la geometría solar (ángulos como el cenit, altura y acimut solar) usando ecuaciones del artículo previamente mencionado [36], permitiendo así entender cómo incide la luz del sol sobre el edificio en diferentes momentos. Luego, se transformó esta radiación de "horizontal global" a "global sobre fachada", utilizando los datos de geometría solar calculados. Además, se distinguió entre radiación solar directa y difusa, y se estableció un mecanismo para medir la radiación que realmente incide sobre la fachada considerando la orientación y características del edificio, como la presencia de ventanas y la ubicación geográfica. Sin embargo, esta última transformación no fue incorporada, por mostrar una menor correlación con la temperatura interior que la radiación global sobre fachada.

6.4.2 Mejora modelos predictivos

Se toma como base el desarrollo integral de los 8 modelos de referencia, que se construyeron siguiendo diferentes criterios. Estos modelos se han analizado más a fondo con el objetivo de identificar posibles mejoras. En particular, se ha buscado reducir el número de variables utilizadas en los modelos y mejorar su precisión predictiva. Para ello, se han explorado técnicas de optimización de modelos y selección de características.

Para el desarrollo se ha establecido que el modelo puede considerar, como máximo, información de las últimas seis horas de cada variable. Esto permite introducir el carácter autorregresivo del modelo al incluir las seis observaciones pasadas de la temperatura interior y de las demás variables exógenas, dando lugar a un modelo inicial con las características de la ecuación 16:

$$\begin{aligned}
 T_{int} = & \beta_0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \cdots + \beta_6 T_{int,t-6} + \\
 & + \beta_7 T_{ext,t} + \beta_8 T_{ext,t-1} + \beta_9 T_{ext,t-2} + \cdots + \beta_{13} T_{ext,t-6} + \\
 & + \beta_{14} R_t + \beta_{15} R_{t-1} + \beta_{16} R_{t-2} + \cdots + \beta_{20} R_{t-6} + \\
 & + \beta_{21} O_t + \beta_{22} O_{t-1} + \beta_{23} O_{t-2} + \cdots + \beta_{27} O_{t-6} + \\
 & + \beta_{28} Q_t + \beta_{29} Q_{t-1} + \beta_{30} Q_{t-2} + \cdots + \beta_{34} Q_{t-6}
 \end{aligned} \tag{16}$$

Donde:

- β_i son los coeficientes de la regresión.
- T_{int} es la variable temperatura interior.
- T_{ext} es la variable temperatura exterior.
- R es la variable radiación solar global sobre fachada.
- O es la variable ocupación por conteo.
- Q es la variable energía de agua refrigerada.
- El subíndice t hace referencia al valor de las variables en el momento que se está prediciendo.
- Los subíndices $t - 1, \dots, t - 6$ hacen referencia al valor de las observaciones pasadas siendo $t - 1$ el instante actual, es decir, una hora antes del momento de predicción y $t - 6$, el momento temporal 6 horas antes del momento de predicción.

En el caso del trabajo tomado como referencia se utilizaron los siguientes criterios para realizar los distintos modelos. Esta decisión fue tomada por el autor arbitrariamente, con el fin de ver con que proceso obtenía el modelo más óptimo:

Término independiente (intersección):

- Forzado a ser nulo.
- Libre.

Extensión/reducción:

- Extender el modelo desde el instante de predicción de las variables exógenas e ir añadiendo pasos de tiempo anteriores.
- Reducir el modelo comenzando con todos los instantes de tiempo anteriores y eliminando pasos de tiempo.

Método de extensión/reducción:

- Por pasos de tiempo: todas las variables tienen el mismo número de pasos de tiempo anteriores.
- Por significatividad de las variables

De este procedimiento se extraen 8 modelos bastante parecidos entre sí, donde se descartan variables que no son significativas según el método que se haya utilizado para llegar hasta ellos. Los 8 modelos se presentan a continuación.

6.4.2.1 Mejora de los modelos 1 y 3

El primer modelo fuerza el término independiente a ser 0 y utiliza un método de extensión, donde cada paso añade un paso de tiempo para todas las variables. El modelo comienza únicamente de variables exógenas en el momento de la predicción hasta llegar a las seis horas anteriores. El modelo 3 se construye de la manera opuesta. Comienza desde el modelo con más variables (ecuación 17) y se va reduciendo hasta llegar al modelo con únicamente las variables exógenas en el momento de la predicción. Cabe destacar que en ambos casos el modelo obtenido es exactamente igual.

El modelo obtenido se representa en la ecuación 16:

$$\begin{aligned} T_{int} = & 0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \\ & + \beta_3 T_{ext,t} + \beta_4 T_{ext,t-1} + \beta_5 T_{ext,t-2} + \\ & + \beta_6 R_t + \beta_7 R_{t-1} + \beta_8 R_{t-2} + \\ & + \beta_9 O_t + \beta_{10} O_{t-1} + \beta_{11} O_{t-2} + \\ & + \beta_{12} Q_t + \beta_{13} Q_{t-1} + \beta_{14} Q_{t-2} \end{aligned} \quad (17)$$

El problema con este modelo viene cuando se observa la significatividad de cada una de sus variables y se ve cómo muchas de ellas no son significativas. La manera para comprobar la significatividad de la variable viene determinada por su p-valor. La hipótesis nula dice que si el p-valor de una variable es inferior a 0,05 se considera significativa. En el caso contrario, en el que el p-valor sea superior a 0,05, se aceptará la hipótesis alternativa que dice que la variable en cuestión no es significativa y que, por lo tanto, no aporta información al modelo.

En la tabla 4 se observan todas las variables del modelo junto con su p-valor:

	p-valor modelo 1
Temperatura exterior t	0,43
Temperatura exterior t-1	0,915594
Temperatura exterior t-2	0,624229
Radiación solar global t	0,56215
Radiación solar global t-1	0,623834
Radiación solar global t-2	0,569887
Ocupantes t	0,505097
Ocupantes t-1	0,134358
Ocupantes t-2	0,000132
Energía consumida en climatización t	0,213634
Energía consumida en climatización t-1	0,859393
Energía consumida en climatización t-2	0,853574
Temperatura interior t-1	2,00E-16
Temperatura interior t-2	6,23E-05

Tabla 4. Significatividad de los modelos de partida 1 y 3

Con esta prueba se puede observar cómo, a excepción de ocupantes por conteo 2 horas antes de la predicción y la temperatura interior 1 y 2 horas antes de la predicción, el resto de las variables parecen no significativas. Sin embargo, puede que, al ir reduciendo las variables del modelo por pasos, variables que ahora se muestran cómo no significativas, terminen siéndolo en un modelo más reducido.

Para obtener un modelo con únicamente variables significativas se va a utilizar la función stepAIC de R. Este método de eliminación hacia atrás comienza con un modelo completo que incluye todas las variables predictoras disponibles (modelo 1). En cada paso, evalúa el efecto de eliminar cada variable del modelo, recalculando el criterio de información de Akaike ([35]) para cada variante del modelo sin esa variable. La variable que, al ser eliminada, resulta en la menor reducción (o incluso una mejora) del AIC, es retirada del modelo. Este proceso se repite hasta eliminar variables adicionales que no mejora el AIC, buscando así el modelo más sencillo que aún explica efectivamente la variabilidad de los datos. En resumen, esta técnica devuelve el modelo con un mejor equilibrio entre complejidad y capacidad explicativa.

El modelo de regresión que se obtiene después de aplicar el método se expresa en la ecuación 18:

$$\begin{aligned}
 T_{int} = & 0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \\
 & + \beta_3 T_{ext,t} + \\
 & + \beta_4 O_{t-2} + \\
 & + \beta_5 Q_t
 \end{aligned}
 \tag{18}$$

En el modelo mejorado se observa que las variables de temperatura interior se han mantenido, que, de las variables de temperatura exterior, ocupación y carga térmica solo se mantiene un parámetro y que la variable radiación ha quedado eliminada por completo.

En la tabla 5 se observan las diferencias en los p-valores de las variables seleccionadas. Se observa como en el nuevo modelo todas las variables son significativas.

	p-valor modelo 1	p-valor modelo 1 stepAIC
Temperatura exterior t	0,43	0,0032
Temperatura exterior t-1	0,915594	
Temperatura exterior t-2	0,624229	
Radiación solar global t	0,56215	
Radiación solar global t-1	0,623834	
Radiación solar global t-2	0,569887	
Ocupantes t	0,505097	
Ocupantes t-1	0,134358	
Ocupantes t-2	0,000132	3,76E-10
Energía consumida en climatización t	0,213634	0,050
Energía consumida en climatización t-1	0,859393	
Energía consumida en climatización t-2	0,853574	
Temperatura interior t-1	2,00E-16	2,00E-16
Temperatura interior t-2	6,23E-05	6,33E-05

Tabla 5. Comparativa de significatividad entre los modelos 1 y 3 y el modelo mejorado para 1 y 3

Por último, se muestra una comparativa de cómo el nuevo modelo, con menos variables, obtiene un R^2 similar al modelo de partida y que además consigue una ligera mejora del MAE. Esto demuestra la sobre parametrización del modelo de partida, así como la correcta eliminación de variables no significativas del mismo.

	Modelo 1 y 3	Modelo 1 y 3 mejorado
N.º de variables	14	5
Reducción de n.º variables		9
R²	0,99994	0,99994
MAE [°C]	0,13725	0,13662
Variación del MAE		0,464%

Tabla 6. Comparativa de variables, R^2 y MAE entre los modelos 1 y 3, y el mejorado

En la tabla 7 se presentan los coeficientes utilizados para cada variable en el nuevo modelo:

	Coeficientes (β)
Temperatura exterior t	0,010859
Ocupantes t-2	-0,012607
Energía consumida en climatización t	-0,0382
Temperatura interior t-1	1,109099
Temperatura interior t-2	-0,118249

Tabla 7. Coeficientes para las variables del modelo 1 y 3 mejorado

6.4.2.2 Mejora del modelo 2:

Los modelos 2 y 4 desarrollados en el trabajo de referencia siguen con la peculiaridad principal de tener el termino independiente forzado a 0. Estos modelos si presentan algunas diferencias significativas con una característica destacable en cada uno.

El modelo numero 2 deja descarta por completo la variable de energía por agua refrigerada, es decir, la energía consumida para la climatización. Esto difiere con el primer modelo mejorado presentado, el cual descartaba por completo la variable de radiación global, pero no la de energía por agua refrigerada.

La ecuación del modelo 2 de referencia se expresa con la ecuación 19:

$$\begin{aligned}
 T_{int} = & 0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \beta_3 T_{int,t-3} \\
 & + \beta_4 T_{ext,t} + \\
 & + \beta_5 R_t + \\
 & + \beta_6 O_t + \beta_7 O_{t-1} + \beta_8 O_{t-2} + \beta_9 O_{t-3}
 \end{aligned}
 \tag{19}$$

Haciendo un análisis de la significatividad de las variables se observa lo mostrado en la tabla 8:

	p-valor modelo 2
Temperatura exterior t	0,046623
Radiación solar global t	0,138266
Ocupantes t	0,855359
Ocupantes t-1	0,858934
Ocupantes t-2	0,882286
Ocupantes t-3	0,000425
Temperatura interior t-1	2,00E-16
Temperatura interior t-2	4,91E-13
Temperatura interior t-3	2,26E-10

Tabla 8. Significatividad de las variables del modelo de partida 2

Como ya se ha explicado antes cuando el p-valor de una variable supera el 0,05 se puede considerar que esa variable no es significativa y, por lo tanto, potencialmente descartable.

Se aplica el stepAIC para descartar las variables no relevantes y se obtiene modelo de la ecuación 20:

$$\begin{aligned}
 T_{int} = & 0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \beta_3 T_{int,t-3} \\
 & + \beta_4 T_{ext,t} + \\
 & + \beta_5 R_t + \\
 & O_{t-3}
 \end{aligned}
 \tag{20}$$

En el modelo mejorado se observa que las variables de temperatura interior se han mantenido y que, de las variables de temperatura exterior, radiación y ocupación solo se mantiene un parámetro.

En la tabla 9, se muestra una comparativa de como el nuevo modelo, con menos variables, obtiene un R^2 similar al modelo de partida y que además consigue una ligera mejora del MAE. Esto demuestra la sobre parametrización del modelo de partida, así como la correcta eliminación de variables no significativas del mismo.

	Modelo 2	Modelo 2 mejorado
Nº de variables	9	6
Reducción de Nº variables		3
R2	0,99995	0,99995
MAE [°C]	0,13529	0,13520
Variación del MAE		0,067%

Tabla 9. Comparativa de n.º de variables, R² y MAE entre el modelo 2 y el mejorado

En la tabla 10, se presentan los coeficientes del modelo mejorado:

	Coeficientes (β)
Temperatura exterior t	0,00859
Radiación solar global t	-0,00014
Ocupantes t-3	-0,01569
Temperatura interior t-1	1,12300
Temperatura interior t-2	-0,31910
Temperatura interior t-3	0,18930

Tabla 10. Coeficientes para las variables del modelo 2 mejorado

6.4.2.3 Mejora del modelo 4:

Con el modelo 4 pasa algo similar a lo del modelo 2. En este caso la variable que queda totalmente descartada es la variable de temperatura exterior. Como se ha visto en el modelo mejorado de los modelos 1 y 3, que contaban con todas las variables seleccionada en un primer momento, la temperatura exterior si parecía tener significatividad. Sin embargo, en este modelo queda descartada por completo. Además, este modelo 4, es sin duda el que más parámetros utiliza por lo que es más probable que se le puedan eliminar variables no significativas.

La ecuación que sigue es la 21:

$$\begin{aligned}
 T_{int} = & 0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \dots + \beta_6 T_{int,t-6} + \\
 & + \beta_7 R_t + \beta_8 R_{t-1} + \beta_9 R_{t-2} + \beta_{10} R_{t-3} + \beta_{11} R_{t-4} + \\
 & + \beta_{12} O_t + \beta_{13} O_{t-1} + \beta_{14} O_{t-2} + \dots + \beta_{18} O_{t-6} + \\
 & + \beta_{19} Q_t + \beta_{20} Q_{t-1} + \beta_{21} Q_{t-2} + \dots + \beta_{25} Q_{t-6}
 \end{aligned}
 \tag{21}$$

Haciendo un análisis de la significatividad de las variables se observa lo mostrado en la tabla 11:

	p-valor modelo 4
Energía consumida en climatización t	0,68
Energía consumida en climatización t-1	0,77550
Energía consumida en climatización t-2	0,03489
Energía consumida en climatización t-3	0,00057
Energía consumida en climatización t-4	0,14950
Energía consumida en climatización t-5	0,03645
Energía consumida en climatización t-6	1,38E-11
Radiación solar global t	0,69943
Radiación solar global t-1	0,58669
Radiación solar global t-2	0,86580
Radiación solar global t-3	0,00971
Radiación solar global t-4	1,87E-05
Ocupantes t	0,95847
Ocupantes t-1	0,57801
Ocupantes t-2	0,63068
Ocupantes t-3	0,04397
Ocupantes t-4	0,79677
Ocupantes t-5	0,01584
Ocupantes t-6	0,00436
Temperatura interior t-1	2,00E-16
Temperatura interior t-2	1,77E-11
Temperatura interior t-3	0,00083
Temperatura interior t-4	0,39324
Temperatura interior t-5	0,86928
Temperatura interior t-6	0,00010

Tabla 11. Análisis de significatividad de las variables del modelo 4 de partida

Se aplica el stepAIC para descartar las variables no relevantes y se obtiene el modelo de la ecuación 22:

$$\begin{aligned}
 T_{int} = & 0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \beta_3 T_{int,t-3} + \beta_4 T_{int,t-6} + \\
 & + \beta_5 R_{t-1} + \beta_6 R_{t-3} + \beta_7 R_{t-4} + \\
 & + \beta_8 O_{t-3} + \beta_9 O_{t-5} + \beta_{10} O_{t-6} + \\
 & + \beta_{11} Q_{t-2} + \beta_{12} Q_{t-3} + \beta_{13} Q_{t-4} + \beta_{14} Q_{t-5} + \beta_{15} Q_{t-6}
 \end{aligned}
 \tag{22}$$

En el modelo mejorado se observa que se han mantenido 4 de las 6 variables de temperatura interior, 3 de las 5 de radiación, 3 de las 7 de ocupación y 5 de las 7 de carga térmica.

En la tabla 12 se muestra una comparativa de como el nuevo modelo. Con menos variables, obtiene un R^2 similar al modelo de partida, pero en este caso se ve una ligera perdida de precisión del MAE. Esta ligera reducción del MAE a costa de eliminar 10 variables del modelo se considera suficientemente pequeña como para poder asumirla.

	Modelo 4	Modelo 4 mejorado
Nº de variables	25	15
Reducción de Nº variables		10
R2	0,99995	0,99995
MAE [°C]	0,12975	0,13011
Variación del MAE		-0,273%

Tabla 12. Comparativa de n.º de variables, R^2 y MAE entre el modelo 4 y el mejorado

En la tabla 13 se presenta la tabla de coeficientes del modelo mejorado:

	Coeficientes (β)
Energía consumida en climatización t-2	0,00531
Energía consumida en climatización t-3	0,00029
Energía consumida en climatización t-4	0,15182
Energía consumida en climatización t-5	0,03833
Energía consumida en climatización t-6	5,48E-12
Radiación solar global t-1	0,08416
Radiación solar global t-3	0,00032
Radiación solar global t-4	6,98E-06
Ocupantes t-3	4,71E-06
Ocupantes t-5	0,00240
Ocupantes t-6	0,00394
Temperatura interior t-1	2,00E-16
Temperatura interior t-2	1,28E-11
Temperatura interior t-3	0,00015
Temperatura interior t-6	5,60E-11

Tabla 13. Coeficientes para las variables del modelo 4 mejorado

6.4.2.4 Mejora de los modelos 5, 6, 7 y 8

Los modelos 5, 6, 7 y 8 son modelos donde el método de desarrollo es igual a los 4 primeros, con la diferencia de que el término independiente queda libre. En el trabajo de referencia estos modelos, aunque son desarrollados en su totalidad, quedan descartados en la conclusión posterior. La justificación del descarte se debe a dos razones principales. El primero es que, los modelos donde el término independiente es forzado a cero presentan un R^2 superior a los que se deja libre. En concreto, los primeros cuatro modelos, como ya se ha podido ver en los apartados anteriores, presentan un R^2 de prácticamente 1. No obstante, los modelos 5, 6, 7 y 8 presentan un R^2 de entorno al 0,8. Esto, aunque parece un razonamiento lógico, no es necesariamente un argumento de peso.

Para comprender este asunto de manera correcta hace falta saber cómo el software R calcula el coeficiente de determinación R^2 . A fin de cuentas, el R^2 de un modelo mide la proporción de la variabilidad total de los datos que es explicada por el modelo, con un valor de 1 indicando un ajuste perfecto y un valor de 0 indicando que el modelo no explica nada de la variabilidad. Se calcula con la ecuación 11.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (11)$$

El numerador es la suma de los cuadrados de los residuos, es decir, las diferencias cuadradas entre los valores observados y los predichos, indicando el error de predicción del modelo. El denominador es la suma total de los cuadrados de las diferencias entre los valores observados y su media, representando la variabilidad total de los datos.

$$R_0^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i y_i^2} \quad (23)$$

En cambio, R_0^2 , que se calcula cuando el intercepto es forzado a cero, ignora el ajuste alrededor de la media y , en su lugar, considera que la línea de regresión pasa por el origen. Esto puede exagerar la eficacia del modelo en casos donde la relación entre las variables no necesariamente se origina en el punto cero, pero puede parecer satisfactoria en este ajuste simplificado. Así, R_0^2 , tiende a ser más alto simplemente porque el denominador que representa la variabilidad total sin centrar es substancialmente más grande.

Este enfoque afecta directamente la comparabilidad entre R^2 y R_0^2 . Un modelo puede mostrar un R_0^2 alto, no porque las variables independientes explican eficazmente la variación en la variable

dependiente de una manera que sería útil más allá de la muestra de datos, sino porque la métrica usada no penaliza adecuadamente la falta de un ajuste más generalizable. En consecuencia, mientras que R_0^2 puede parecer indicar un ajuste mejor o comparable, podría estar simplemente ocultando las deficiencias del modelo debido a la no inclusión del intercepto.

En el trabajo de referencia, se da un segundo argumento de por qué un buen modelo de regresión debería presentar un término independiente igual a cero, postulando que las variables predictoras deberían explicar completamente el comportamiento de la variable predicha. Si bien se reconoce la validez de este enfoque bajo ciertas circunstancias teóricas, se considera importante resaltar que, en la práctica, y específicamente en la modelación de sistemas complejos como la temperatura interior de un ambiente, puede haber factores adicionales no capturados directamente por las variables incluidas.

El término independiente en un modelo estadístico actúa como un ajuste que puede capturar la influencia de todas aquellas variables latentes o no observadas que no están explícitamente incluidas en el modelo. En el caso de predecir la temperatura interior, algunas de estas variables pueden incluir la inercia térmica del edificio, diferencias en el aislamiento, o incluso variaciones en la capacidad de los sistemas de climatización que no están directamente medidas por las variables de energía consumida.

Forzar el término independiente a cero podría simplificar inapropiadamente el modelo, asumiendo que la relación entre las variables predictoras (como la temperatura exterior, la radiación solar, los ocupantes y la energía consumida en climatización) y la temperatura interior comienza exactamente en el origen, es decir, que no hay variabilidad base en la temperatura interior que no sea explicada por estas variables. Sin embargo, esto rara vez es el caso en escenarios del mundo real donde la dinámica del sistema puede ser afectada por factores no capturados directamente por las variables medibles.

Dejar el término independiente libre permite que el modelo ajuste automáticamente un nivel base de temperatura que podría ser atribuido a condiciones constantes o promedios a largo plazo no explicados por las variables directas. Por ejemplo, la temperatura base de una habitación sin calefacción ni refrigeración activa podría diferir de cero debido a factores como el calor residual de aparatos eléctricos o el efecto térmico de las paredes y el suelo.

Dicho esto, para la mejora de los modelos 5, 6, 7 y 8 se va a tratar de obtener un único modelo que mejore los 4 únicamente considerando las variables significativas. Para desarrollar este modelo se va a realizar una reconsideración de los pasos de tiempo máximos a incluir. Después de ver los 4 modelos previos se puede observar que la inclusión de tres pasos de tiempo para las variables explicativas es suficiente para explicar con suficiente precisión la variabilidad de la temperatura interior. Esto implica que el modelo de partida esta vez será el representado en la ecuación 24.

$$\begin{aligned}
T_{int} = & \beta_0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \beta_3 T_{int,t-3} + \\
& + \beta_4 T_{ext,t} + \beta_5 T_{ext,t-1} + \beta_6 T_{ext,t-2} + \beta_7 T_{ext,t-3} + \\
& + \beta_8 R_t + \beta_9 R_{t-1} + \beta_{10} R_{t-2} + \beta_{11} R_{t-3} + \\
& + \beta_{12} O_t + \beta_{13} O_{t-1} + \beta_{14} O_{t-2} + \beta_{15} O_{t-3} + \\
& + \beta_{16} Q_t + \beta_{17} Q_{t-1} + \beta_{18} Q_{t-2} + \beta_{19} Q_{t-3}
\end{aligned}
\tag{24}$$

Como ya se ha expuesto anteriormente, lo primero es hacer un análisis de significatividad de las variables para comprobar que hay variables que potencialmente pueden ser eliminadas.

	p-valor modelo con intercepto
Intercepto	2,00E-16
Temperatura exterior t	0,5706
Temperatura exterior t-1	0,9832
Temperatura exterior t-2	0,1223
Temperatura exterior t-3	0,0124
Radiación solar global t	0,1337
Radiación solar global t-1	0,9441
Radiación solar global t-2	0,3385
Radiación solar global t-3	0,0498
Ocupantes t	0,4204
Ocupantes t-1	0,5665
Ocupantes t-2	0,4502
Ocupantes t-3	7,57E-06
Energía consumida en climatización t	0,1323
Energía consumida en climatización t-1	0,6082
Energía consumida en climatización t-2	0,0854
Energía consumida en climatización t-3	0,0003
Temperatura interior t-1	2,00E-16
Temperatura interior t-2	6,04E-13
Temperatura interior t-3	2,08E-05

Tabla 14. Análisis de significatividad de las variables del modelo 5, 6, 7 y 8 de partida

Lo primero que debe de llamar la atención de este análisis es la clara significatividad que muestra el término independiente, siendo su p-valor de prácticamente 0, lo que confirma los dos argumentos previos. Por otro lado, se observa la no significatividad de varias variables lo que hace interesante llevar a cabo un proceso de eliminación de estas.

Se aplica el método stepAIC para descartar las variables no relevantes y se obtiene el siguiente modelo:

$$\begin{aligned}
 T_{int} = & \beta_0 + \beta_1 T_{int,t-1} + \beta_2 T_{int,t-2} + \beta_3 T_{int,t-3} + \\
 & + \beta_4 T_{ext,t-2} + \beta_5 T_{ext,t-3} + \\
 & + \beta_6 R_t + \beta_7 R_{t-3} + \\
 & + \beta_8 O_{t-3} + \\
 & + \beta_9 Q_t + \beta_{10} Q_{t-2}
 \end{aligned}
 \tag{25}$$

Se muestra una comparativa de como el nuevo modelo, con menos variables, obtiene un R² similar al modelo de partida, y como consigue ser el modelo más equilibrado entre el número de variables utilizadas y el MAE.

	Modelo 5	Modelo 6	Modelo 7	Modelo 8	Modelo 5, 6, 7 y 8 mejorados
Nº de variables	20	7	15	26	12
Intersección	4,2205	3,76926	4,38046	2,90179	4,18151
R2	0,81331	0,80314	0,80698	0,82872	0,812686413
MAE [°C]	0,13286	0,13692	0,13482	0,12806	0,13255

Tabla 15. Modelos 5, 6, 7 y 8 mejorados

6.4.3 Observaciones importantes sobre los modelos

Aunque desde el punto de vista teórico los modelos presentan mejoras respecto a los modelos previos, surgen algunas cuestiones relevantes que merecen la pena cuestionarse.

La primera de ellas es la clara significatividad predominante que presenta la temperatura interior una hora antes del momento de la predicción. Para reflejar esta clara dominancia en la explicación del modelo de la temperatura interior es interesante llevar a cabo un análisis de los coeficientes estandarizados de todas las variables. Esta estandarización de los coeficientes permite comparar de manera equitativa la influencia de cada variable predictora sobre la variable dependiente, ajustando todas las variables a una escala uniforme. Al estandarizar los coeficientes, se transforman las unidades originales de las variables en desviaciones estándar, lo que facilita la interpretación directa y comparativa de su impacto en el modelo. Esto resulta especialmente útil en modelos donde las variables operan en distintas escalas o unidades, permitiendo que los coeficientes reflejen los cambios en la variable de respuesta en términos de desviaciones estándar de cada variable predictora.

En la tabla 16 se muestran los coeficientes estandarizados del modelo con todas la variables y 6 pasos de tiempo anteriores.

	Coeficientes estandarizados
Intercepto	5,21459E-15
Energía consumida en climatización t	-0,016
Energía consumida en climatización t-1	-0,009
Energía consumida en climatización t-2	-0,047
Energía consumida en climatización t-3	0,081
Energía consumida en climatización t-4	0,027
Energía consumida en climatización t-5	0,056
Energía consumida en climatización t-6	-0,122
Radiación solar global t	0,017
Radiación solar global t-1	-0,007
Radiación solar global t-2	-0,029
Radiación solar global t-3	0,113
Radiación solar global t-4	-0,070
Radiación solar global t-5	-0,001
Radiación solar global t-6	-0,047
Ocupantes t	-0,005
Ocupantes t-1	0,031
Ocupantes t-2	-0,029
Ocupantes t-3	-0,104
Ocupantes t-4	-0,018
Ocupantes t-5	0,120
Ocupantes t-6	-0,084
Temperatura interior t-1	1,012
Temperatura interior t-2	-0,288
Temperatura interior t-3	0,137
Temperatura interior t-4	-0,042
Temperatura interior t-5	0,010
Temperatura interior t-6	0,081
Temperatura exterior t	-0,040
Temperatura exterior t-1	-0,007
Temperatura exterior t-2	0,093
Temperatura exterior t-3	-0,038
Temperatura exterior t-4	-0,009
Temperatura exterior t-5	0,027
Temperatura exterior t-6	-0,045

Tabla 16. Coeficientes estandarizados para el modelo completo

Desde una perspectiva lógica, resulta poco intuitivo que variables como la ocupación, la radiación solar o el consumo energético del agua refrigerada presenten una mayor relevancia predictiva tres o seis horas antes del evento, en comparación con una hora antes o incluso en el momento mismo de

la predicción. Esta observación sugiere que podría existir una dinámica subyacente en los datos que no es inmediatamente evidente.

Por esta razón, se ha optado por realizar una serie de visualizaciones gráficas de las variables explicativas en relación con la temperatura interior, para explorar más detalladamente el comportamiento observado. La primera de estas visualizaciones se centra en la relación entre la temperatura interior y la exterior a lo largo de una semana.

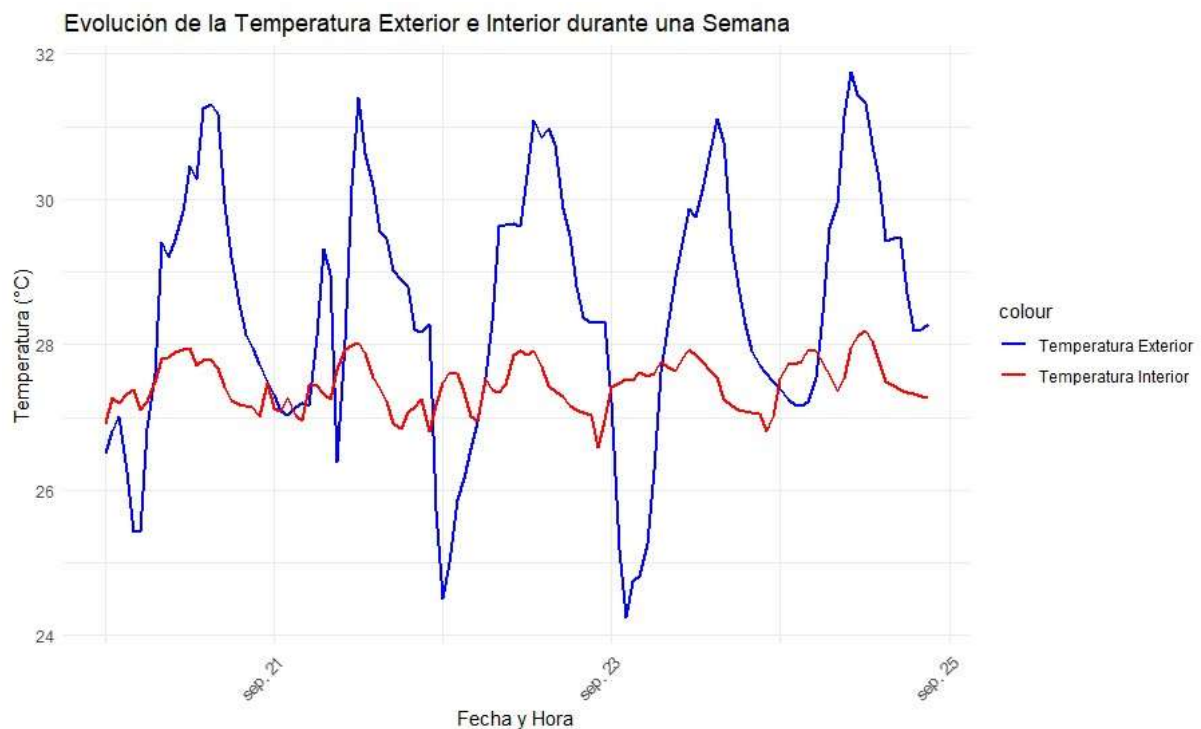


Ilustración 4. Evolución de la temperatura interior y la exterior del 20 al 24 de septiembre de 2021

En esta gráfica, dos aspectos son particularmente destacables. Primero, se observa que la temperatura interior no siempre muestra una correlación directa con la temperatura exterior. Un claro ejemplo de esto ocurre al inicio del día 22 de septiembre, donde, a pesar de que la temperatura exterior disminuye, la interior aumenta. Este fenómeno puede explicarse por la acción del sistema de control de temperatura del edificio, representado en este análisis por la variable que mide la energía consumida para refrigerar el agua. Este sistema podría estar mitigando activamente los efectos de los cambios externos en la temperatura.

Según la información proporcionada por la base de datos de la Universidad Nacional de Singapur, que recopila datos las 24 horas del día, se ha realizado un análisis limitado a las horas nocturnas, entre las 21:00 y las 7:00. Este intervalo se ha seleccionado específicamente porque minimiza las posibles interferencias de otros factores: durante estas horas, la radiación solar es inexistente, el sistema de climatización generalmente está apagado y hay una notable reducción en la presencia de personas en el área. Estas condiciones permiten una comparación más clara y directa entre las temperaturas

interior y exterior, facilitando una mejor comprensión de su interacción en ausencia de influencias externas significativas.

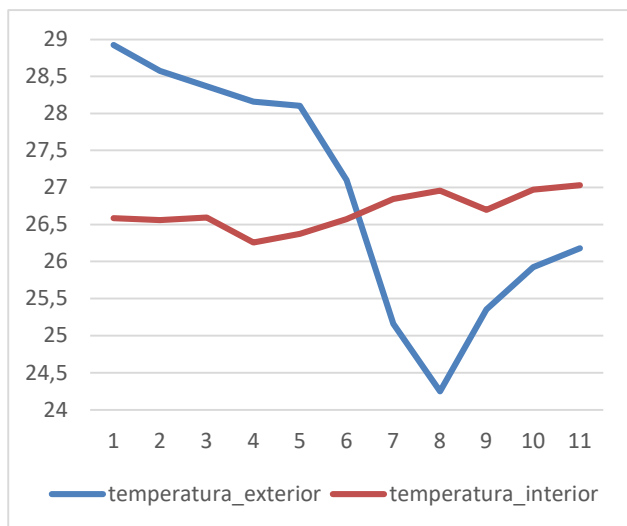


Ilustración 5. Evolución de la temperatura interior con respecto a la temperatura exterior (08/09/2021)

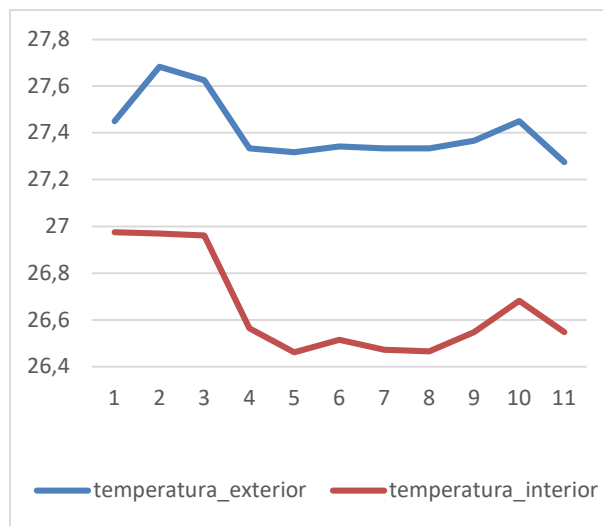


Ilustración 6. Evolución de la temperatura interior con respecto a la temperatura exterior (14/09/2021)

La Ilustración 6 del estudio evidencia una relación consistente entre la temperatura interior y la exterior, lo cual es esperado especialmente en ausencia de factores disruptivos como la radiación solar, el funcionamiento del sistema de climatización y la presencia variable de ocupantes. Este vínculo directo sugiere que, bajo condiciones controladas, la temperatura exterior tiene un impacto predecible y directo sobre la temperatura interior.

Sin embargo, en la Ilustración 5 se observa un comportamiento aparentemente contraintuitivo de la temperatura interior. A pesar de una disminución significativa de cinco grados en la temperatura exterior, la temperatura interior muestra una notable estabilidad e incluso experimenta un incremento de medio grado. Por este motivo, se ha descartado la opción de mejorar los modelos actuales separando los análisis de noche y día. Las observaciones indican que la temperatura exterior no es el único factor que influye en la temperatura interior. Aunque el sistema de climatización esté apagado y no haya radiación solar, otros factores como el aislamiento de las habitaciones y el calor acumulado en paredes y suelos pueden seguir influyendo en la temperatura interior. Estos elementos pueden actuar como reservorios térmicos, liberando o absorbiendo calor, lo que resulta en variaciones de temperatura que no son directamente atribuibles a las condiciones externas inmediatas.

El otro punto destacable del análisis es que la variación de la temperatura interior es mucho más controlada que la de la temperatura exterior, manteniéndose la primera entre 27 y 28 °C a lo largo de toda la semana, mientras que la temperatura exterior exhibe variaciones más pronunciadas, de 24.5 a

31.5 °C. Este patrón refleja claramente la efectividad del sistema de control de temperatura del edificio.

La temperatura interior registrada está ligeramente por encima de la consigna, que tiene un promedio 26.8 °C, según la base de datos proporcionada por la Universidad Nacional de Singapur. Este fenómeno puede explicarse por la necesidad predominante de enfriar el espacio, no de calentarlo, dado el clima tropical de Singapur. Además, es importante destacar de nuevo que, durante la noche, el sistema de climatización no está activo, lo que podría contribuir a que la temperatura interior suba ligeramente por encima del setpoint establecido.

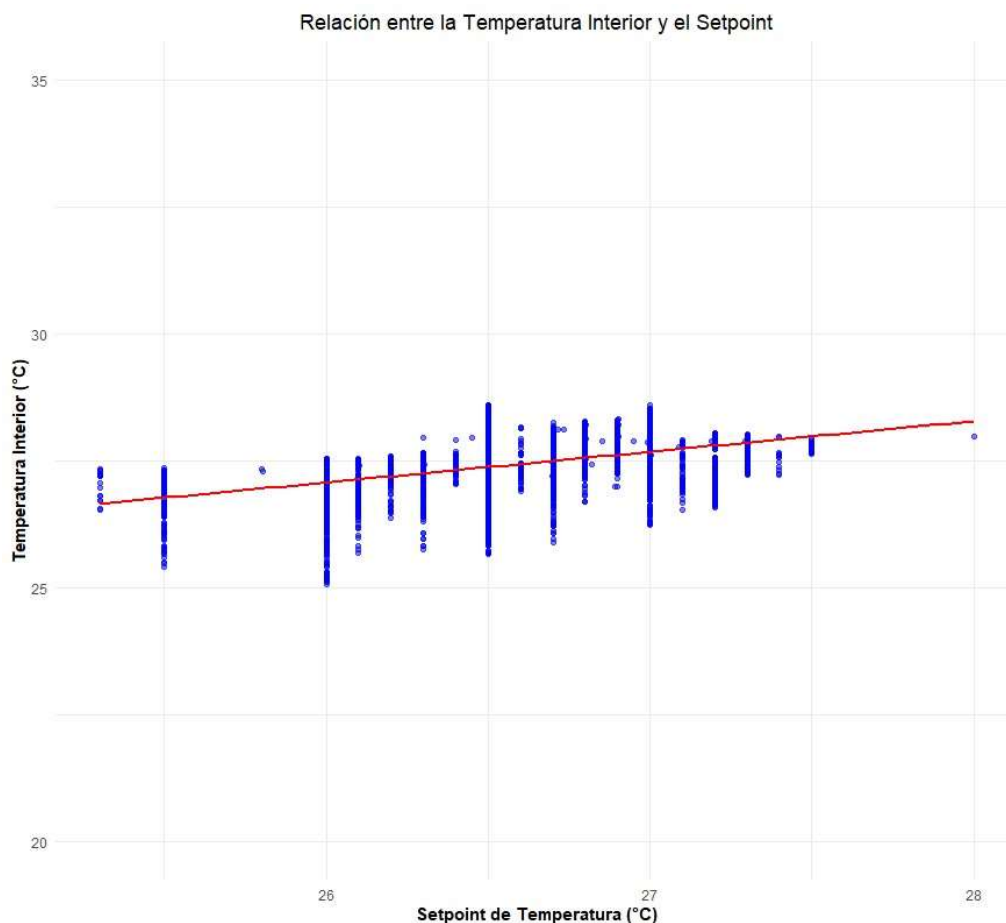


Ilustración 7. Relación entre la temperatura interior y su consigna

6.4.4 Primeras conclusiones

Al centrar nuestro análisis en la temperatura interior de un edificio con control climático, es difícil discernir los impactos directos de las variables climáticas externas sobre esta. Se ha utilizado una tabla de correlaciones de Spearman, debido a la no normalidad de las variables, que ilustra este fenómeno de manera clara. Observamos que la temperatura interior presenta correlaciones débiles o insignificantes con todas las variables seleccionadas, indicando una influencia mínima o nula de estos

factores externos sobre el clima interior del edificio. En contraste, el consumo de energía del sistema de climatización muestra correlaciones considerables con varias de estas variables, subrayando su papel crucial en la regulación de la temperatura interior frente a las fluctuaciones externas.

Esta diferencia en las correlaciones resalta cómo el sistema de climatización compensa activamente las variaciones en las condiciones climáticas externas, manteniendo así un ambiente interior estable. La tabla de correlaciones proporciona evidencia cuantitativa de esta dinámica, ofreciendo una perspectiva valiosa sobre la efectividad del control climático del edificio.

	Temperatura interior	Energía consumida en la climatización	Consigna temperatura interior	Radiación solar global	Temperatura exterior	Ocupantes
Temperatura interior	1	0,237773	0,343350	0,332827	0,249469	0,114643
Energía consumida en la climatización	0,237773	1	0,004698	0,809465	0,542317	0,574338

Tabla 17. Tabla de correlaciones para la temperatura interior y la energía consumida para refrigerar agua

Dado este contexto, se ha decidido cambiar el enfoque en la mejora de los modelos desarrollados anteriormente, orientando los esfuerzos hacia la creación de nuevos modelos que se centren en predecir el consumo de energía utilizado en la climatización del edificio, en lugar de solo la temperatura interior. Este cambio de enfoque no solo se alinea mejor con las características del edificio de la que se dispone la base de datos, sino que también ofrece varios beneficios prácticos.

Predecir el consumo energético para la climatización puede llevar a una gestión más eficiente de los recursos energéticos, lo que resulta en un ahorro económico significativo y una reducción de la huella de carbono del edificio. Además, permite una planificación más efectiva en términos de mantenimiento y optimización de los sistemas de HVAC (calefacción, ventilación y aire acondicionado), asegurando que el edificio mantenga un ambiente confortable para sus ocupantes sin incurrir en gastos innecesarios de energía. Finalmente, modelos precisos de consumo energético son esenciales para la implementación de estrategias de respuesta a la demanda que pueden mejorar la sostenibilidad general de las infraestructuras urbanas al integrar de manera más efectiva fuentes de energía renovable y tecnologías de edificios inteligentes.

6.5 SEGUNDA FASE: CREACIÓN DE NUEVOS MODELOS

6.5.1 Método de construcción

Para la construcción de los modelos, se adoptará un enfoque mixto que combina expansiones manuales y técnicas automáticas. Inicialmente, se empleará un método de expansión manual, donde las variables serán añadidas progresivamente al modelo. Este proceso no solo se basará en la mejora observada en las métricas de R^2 y MAE, sino también en la significancia estadística de las nuevas variables añadidas y su efecto sobre las ya incluidas en el modelo. Este paso cuidadoso asegura que cada nueva inclusión aporte valor significativo y relevante.

Posteriormente, se complementará este enfoque manual con la aplicación de la función stepAIC en R. Este método automático realiza una selección de variables basada en el criterio de información de Akaike, equilibrando eficazmente la complejidad del modelo contra su precisión. La combinación de expansiones manuales y ajustes automáticos permitirá refinar el modelo para alcanzar un equilibrio óptimo entre simplicidad y capacidad predictiva, garantizando que el modelo final sea tanto manejable como efectivo.

6.5.2 Variables seleccionadas

Las tres primeras variables seleccionadas son las que presentan una correlación lo suficientemente fuerte con la energía consumida en la climatización.

- **Temperatura exterior:** la temperatura exterior se incluye en el modelo porque juega un papel crítico en la determinación de la transferencia de calor entre el exterior y el interior de un edificio, impactando directamente las demandas sobre el sistema de climatización. Al modelar la temperatura exterior, se puede anticipar y ajustar proactivamente la operación de los sistemas de climatización para responder a las condiciones ambientales externas.
- **Radiación solar:** la radiación solar es una variable crucial debido a su significativo impacto en la carga térmica interior de un edificio, especialmente en áreas expuestas a la luz solar directa. Modelar esta variable permite evaluar cómo la exposición solar contribuye al calentamiento interno y, por ende, al aumento del consumo de energía para climatización.
- **Ocupantes:** el número de ocupantes se incorpora en el modelo ya que cada persona genera calor sensible y latente, sino que también afecta la calidad del aire y los niveles de humedad, incrementando así la carga térmica del espacio.

Además, se añade una nueva variable que resulta más interesante que la temperatura interior al uso.

- **Diferencia entre setpoint y temperatura interior:** la diferencia entre el setpoint y la temperatura interior es seleccionada como variable clave debido a que proporciona una medida directa del esfuerzo requerido por el sistema de climatización para alcanzar el confort deseado. Esta variable es muy eficaz para modelar, ya que refleja con precisión la energía necesaria para ajustar la temperatura interna al nivel óptimo establecido por el setpoint, facilitando predicciones precisas sobre el consumo energético y la gestión eficiente del sistema HVAC. La adición de esta variable es una de las principales innovaciones del proyecto.

En la ilustración 8 se muestra la tabla de correlaciones para las variables seleccionadas. Es interesante observar cómo todas las relaciones entre la variable que se quiere predecir (“energía_agua_refrigerada”) y el resto de las variables seleccionadas muestra una correlación de más de 0,3. Además, en la representación gráfica puede observarse una relación suficientemente lineal como para que tenga sentido utilizar un modelo lineal para caracterizar la carga térmica. Por último, también observamos una relación lineal entre la carga térmica en el momento de la predicción y ella misma una hora antes, lo que justifica el uso de un modelo ARX lineal.

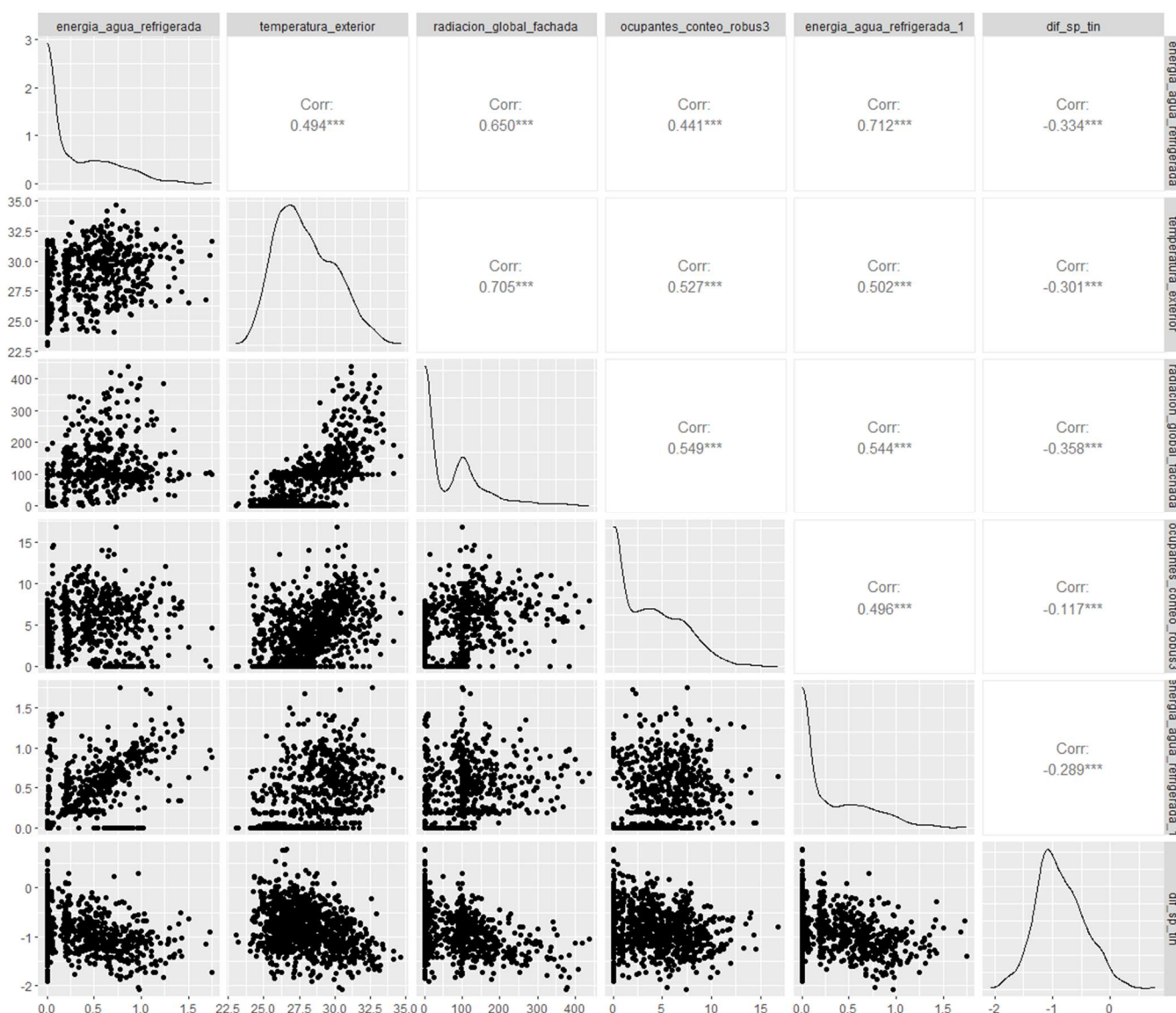


Ilustración 8. Tabla de correlaciones entre las variables seleccionadas

6.5.3 Otras consideraciones

Para desarrollar el modelo, se ha optado por un intervalo de muestreo y un horizonte de predicción de 60 minutos. Esto significa que el modelo recoge datos de cada variable cada 60 minutos y efectúa predicciones para el periodo siguiente de 60 minutos. Cabe destacar que, aunque el conjunto de datos original tiene un intervalo de muestreo de 5 minutos, se ha ajustado a un marco temporal deseado mediante el uso de una media móvil centrado de 11 puntos (equivalente a 60/5 minutos), calculando así el valor medio de cada variable usando los 5 valores previos y los 5 siguientes. Posteriormente, para sincronizar los datos con un muestreo cada hora en punto (00:00, 01:00, ..., 23:00), se seleccionan solamente las observaciones correspondientes a estas horas.

Además, se ha establecido que, como mucho, el modelo puede obtener información de las 2 últimas horas de cada variable, de esta forma se introduce el carácter autorregresivo del modelo al considerar también las 2 observaciones pasadas de temperatura interior y del resto de variables exógenas. No existe un criterio estandarizado para tomar esta decisión, pero el estudio de los antecedentes con aplicaciones similares a la que se lleva a cabo en el presente trabajo apuntan a que las variables con un retardo de más de dos horas no aportan información relevante a los modelos [27] [28].

El uso de variables actuales como la temperatura exterior, la radiación solar, la ocupación por conteo y la diferencia entre el setpoint y la temperatura interior para predecir la carga térmica futura está justificado por la capacidad de los sistemas modernos de gestión de edificios para proyectar estas condiciones hacia adelante. Se asume que tecnologías avanzadas, como datos meteorológicos de satélites y modelos climáticos locales, proporcionan predicciones confiables que permiten utilizar las condiciones actuales como representaciones precisas de las futuras. Similarmente, sistemas de predicción de ocupación basados en patrones históricos y análisis en tiempo real aseguran que la variable de ocupación refleje niveles futuros acertadamente. La variable de la diferencia entre el setpoint y la temperatura interior se utiliza bajo la premisa de que los ajustes en la temperatura son programados con antelación, lo que permite una gestión energética proactiva y eficiente. Este enfoque facilita un control más eficaz del sistema de climatización, optimizando tanto el consumo energético como el confort dentro del edificio.

6.5.4 Métricas de error

En el desarrollo de los modelos para este estudio, se utilizarán dos métricas principales para evaluar el rendimiento y la precisión de las predicciones: el coeficiente de determinación (R^2) y el Error Absoluto Medio (MAE).

Al combinar R^2 y MAE, se obtiene una visión holística del rendimiento del modelo, evaluando tanto la precisión como la utilidad de las predicciones, lo que es fundamental para la implementación práctica en sistemas de predicción de la energía consumida en edificios. Estas métricas ayudan a equilibrar la complejidad del modelo con su capacidad predictiva, asegurando que los modelos no solo sean rigurosos en términos estadísticos, sino también viables y eficientes desde una perspectiva de ingeniería aplicada.

El MAE se ha elegido sobre el error cuadrático medio RMSE por varias razones clave en el contexto de este proyecto de investigación. Primero, el MAE proporciona una medida más directa y comprensible de los errores medios, ya que no eleva los errores al cuadrado. Esto significa que el MAE es menos sensible a los valores atípicos, que son errores muy grandes; no los penaliza tanto como el RMSE. En investigaciones donde se exploran nuevos modelos predictivos, esta característica es importante porque permite que el impacto de los valores atípicos se mantenga proporcional y no domine la evaluación de la calidad del modelo.

Además, en un entorno de investigación, los efectos de esos errores pronunciados no son necesariamente desfavorables, pues pueden proporcionar información valiosa sobre las limitaciones del modelo o circunstancias excepcionales en los datos.

En concreto se busca un R^2 superior a 0,6 y un MAE inferior a 0,15 kWh.

6.5.5 Modelado

6.5.5.1 Modelo inicial

Para comenzar la construcción del modelo de predicción eficaz para la energía consumida en la climatización, se ha establecido un modelo inicial que sirva como punto de partida para análisis y refinamientos posteriores. El modelo inicial propuesto incluye dos variables explicativas críticas: la diferencia entre el setpoint y la temperatura interior, y la temperatura exterior. Esta elección de variables está orientada a capturar los principales impulsores del consumo energético en un sistema de climatización de un modo simplificado pero efectivo. El modelo inicial se representa en la ecuación 26.

$$Q_t = \beta_0 + \beta_1 DSPT_{in,t} + \beta_2 T_{ext,t} \quad (26)$$

Se presupone que este modelo no va a ser capaz de predecir de manera precisa la carga térmica, pero sirve como punto de partida para observar la significatividad de las variables seleccionadas.

En la tabla 18, se da un resumen de las características principales del modelo.

Se puede observar que el modelo, aunque tanto en su totalidad como por cada variable por separado sea significativo muestra un R^2 muy bajo y un MAE elevado. Para una mejor comprensión del lector cabe destacar que la energía consumida en la climatización toma valores de 0 a aproximadamente 1,75 kWh, por lo que un error medio absoluto de 0,23 kWh sigue siendo significativamente alto.

Variables	Coeficientes	p-valor	p-valor (modelo)	R^2	R^2 Ajustado	MAE (kWh)
Intercepto	-2,002184	2,00E-16	2,20E-16	0,282	0,2806	0,23043
Temperatura exterior	0,075518	2,00E-16				
Diferencia SP y Tin	-0,170023	3,10E-14				

Tabla 18. Resumen de los aspectos más relevantes del modelo inicial

En este modelo no tiene sentido la aplicación de la función stepAIC, ya que como se observa en la tabla todas las variables se consideran significativas.

En la ilustración 9, se representa gráficamente una comparativa entre la energía consumida predicha y la energía consumida real.

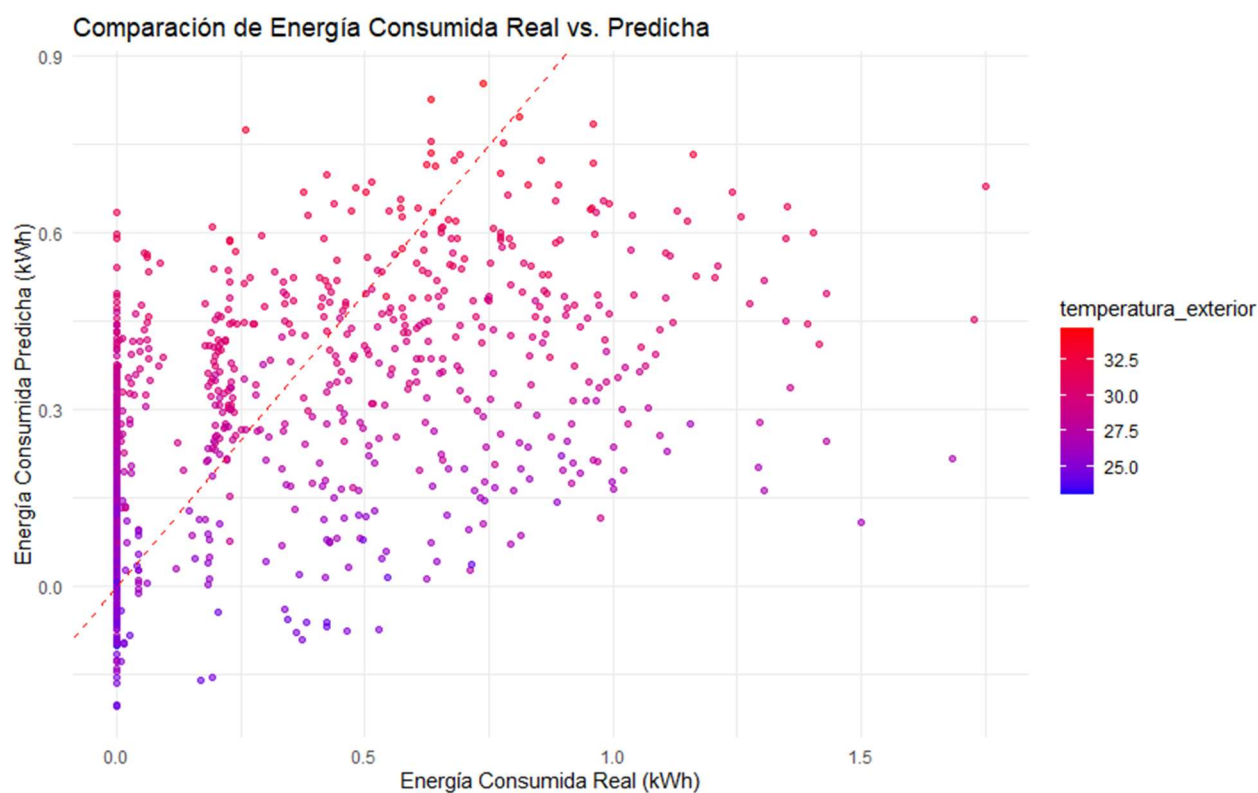


Ilustración 9. Representación gráfica de la energía consumida para la climatización medida y la energía consumida predicha para el modelo inicial.

6.5.5.2 Modelos añadiendo variables en el momento t

A partir de este primer modelo se van a ir añadiendo nuevas variables que hagan que el modelo vaya obteniendo un mejor ajuste. La primera de las variables introducidas en la radiación solar, obteniendo un modelo como el representado en la ecuación 19:

$$Q_t = \beta_0 + \beta_1 DSPT_{in,t} + \beta_2 T_{ext,t} + \beta_3 R_t \quad (27)$$

En la tabla 17 se da un resumen de las características principales del modelo.

Variables	Coeficientes	p-valor	p-valor (modelo)	R ²	R ² Ajust	MAE (kWh)
Intercepto	-0,2690397	0,0707	2,20E-16	0,436	0,4347	0,184679
Temperatura exterior	0,0104966	0,058				
Diferencia SP y Tin	-0,0942176	2,97E-06				
Radiación solar	0,0024092	2,00E-16				

Tabla 19. Resumen de los aspectos más relevantes del modelo inicial y radiación solar

La inclusión de la variable de radiación solar en el modelo ha resultado ser altamente beneficiosa, evidenciado por una mejora significativa en las métricas clave de desempeño del modelo.

Específicamente, el coeficiente de determinación (R²) del modelo mejoró en un 54%, indicando que una mayor proporción de la variabilidad en la energía consumida ahora puede ser explicada con la adición de la radiación solar. Además, el Error Absoluto Medio (MAE) experimentó una reducción del 19.8%, lo que sugiere que las predicciones del modelo se han vuelto más precisas.

Adicionalmente, al evaluar la significancia estadística de las variables incluidas en el modelo mediante los p-valores, se observa un cambio notable en la importancia de la temperatura exterior. Antes percibida como altamente significativa, la temperatura exterior pierde relevancia con la introducción de la radiación solar, lo que puede indicar una correlación entre estas variables que influye en su capacidad predictiva individual. De manera similar, el intercepto del modelo también pierde significancia, lo cual sugiere que las variaciones en la energía consumida están siendo más efectivamente capturadas por las variables del modelo en lugar de ser atribuidas al término de intercepto.

Dado estos hallazgos, la utilización de la función stepAIC se justifica plenamente para este modelo. Este método de selección de variables podría ayudar a identificar y eliminar aquellas que no aportan significativamente a la predicción del modelo, optimizando así la simplicidad y efectividad de este.

Sin embargo, una vez aplicado el método stepAIC el modelo devuelto sigue siendo el mismo por lo cual queda validado como mejor modelo hasta ahora. Se continúa construyendo a partir de este.

La siguiente y última variable introducida para el momento de tiempo t es la ocupación por conteo.

$$Q_t = \beta_0 + \beta_1 DSPT_{in,t} + \beta_2 T_{ext,t} + \beta_3 R_t + \beta_4 O_t \quad (28)$$

En la tabla 20 se da un resumen de las características principales del modelo.

Variables	Coeficientes	p-valor	p-valor (modelo)	R ²	R ² Ajustado	MAE (kWh)
Intercepto	-0,1289511	0,391	2,20E-16	0,447	0,4453	0,179751
Temperatura exterior	0,0039128	0,489				
Diferencia SP y Tin	-0,1058833	1,48E-07				
Radiación solar	0,0021963	2,00E-16				
Ocupación por conteo	0,0141031	2,36E-06				

Tabla 20. Resumen de los aspectos más relevantes del modelo inicial y radiación solar + ocupación por conteo

La inclusión de la variable "ocupantes por conteo" en el modelo ha generado una mejora leve. Aunque el incremento en el coeficiente R² no es destacable y podría atribuirse meramente a la inclusión de esta nueva variable más que a un aumento real en la capacidad explicativa del modelo, el R² ajustado, que toma en cuenta la penalización por añadir variables adicionales, aún muestra una ligera mejora comparado con el modelo anterior. Por otro lado, la métrica del Error Absoluto Medio (MAE) sí ha experimentado una mejora más notable, con una reducción específica del 2,7%.

En el modelo revisado, al igual que en los intentos anteriores, se observa de nuevo una pérdida de significatividad tanto en la variable temperatura exterior como en el intercepto. Esto reitera la posibilidad de redundancia o de superposición en la explicación que estas variables aportan al modelo. Dada esta situación, se justifica nuevamente el uso de la función stepAIC para explorar si simplificar el modelo podría mejorar su eficiencia y precisión. Al aplicar stepAIC, la variable de temperatura exterior ha sido descartada del modelo, lo que sugiere que su contribución no era esencial para la capacidad predictiva del modelo con respecto a las otras variables incluidas.

El modelo resultante se representa en la ecuación 29:

$$Q_t = \beta_0 + \beta_1 DSPT_{in,t} + \beta_2 R_t + \beta_3 O_t \quad (29)$$

En la tabla 21 se da un resumen de las características principales del modelo.

Variables	Coeficientes	p-valor	p-valor (modelo)	R ²	R ² Ajustado	MAE (kWh)
Intercepto	-0,0257991	0,18	2,2E-16	0,447	0,4455	0,179722
Diferencia SP y Tin	-0,1072749	8,76E-08				
Radiación solar	0,00225	2,00E-16				
Ocupación por conteo	0,0146083	4,63E-07				

Tabla 21. Resumen de los aspectos más importantes del modelo óptimo utilizando solo variables exógenas en el momento t

El análisis revela que el modelo óptimo, que utiliza solamente variables exógenas en el momento actual t, incorpora todas las variables seleccionadas excepto la temperatura exterior. Esta exclusión es clave, dado que la eliminación de la temperatura exterior no afecta adversamente las métricas de rendimiento del modelo; tanto el R² como el MAE se mantienen estables. Además, al prescindir de esta variable, el modelo resulta menos complejo.

6.5.5.3 Modelos añadiendo variables con un paso de tiempo (t-1)

A partir del modelo óptimo que utiliza variables exógenas en el momento actual t, se planea avanzar hacia la incorporación de un enfoque más dinámico al añadir las mismas variables exógenas, pero observadas una hora antes (en t-1). Este enriquecimiento del modelo, donde las variables se introducen con un paso de tiempo hacia atrás, facilitará la transición de un modelo de regresión lineal múltiple convencional a uno de tipo ARX (AutoRegressive with eXogenous inputs).

Este modelo ARX permite capturar no solo el impacto instantáneo de las variables exógenas sobre la variable dependiente, sino también cómo las condiciones pasadas afectan el comportamiento presente del sistema.

La primera nueva variable que se añade es la energía consumida para la climatización en el instante actual (t-1). Al hacer esto, estamos implementando un enfoque que refleja la naturaleza auto correlativa de los sistemas de climatización. Este enfoque permite que el modelo utilice el consumo energético más reciente para predecir el consumo futuro, aprovechando la información sobre cómo las condiciones pasadas pueden informar o influir en las necesidades energéticas futuras.

Incorporar el consumo de energía en t-1 ofrece una base directa para estimar la cantidad de energía que podría requerirse en el próximo intervalo. Este paso es crucial porque la demanda de energía en

sistemas de climatización no solo es influenciada por las condiciones externas e internas actuales, sino también por el rendimiento y las demandas del sistema en el pasado reciente.

$$Q_t = \beta_0 + \beta_1 DSPT_{in,t} + \beta_2 R_t + \beta_3 O_t + \beta_4 Q_{t-1} \quad (30)$$

En la tabla 22 se da un resumen de las características principales del modelo.

Variables	Coeficientes	p-valor	p-valor (modelo)	R ²	R ² Ajustado	MAE (kWh)
Intercepto	-0,0073684	0,65175	2,2E-16	0,6081	0,6067	0,133248
Diferencia SP y Tin	-0,001602	0,00314				
Radiación solar	-0,0506202	2,00E-16				
Ocupación por conteo	0,0015361	0,52891				
Energía consumida climatización	0,5050068	2,00E-16				

Tabla 22. Resumen de los aspectos más importantes del modelo óptimo con solo variables exógenas en el momento t y energía consumida para la climatización en t-1

La inclusión del consumo de energía para la climatización de la habitación en el instante t-1 en el modelo ha resultado en una notable mejora de su desempeño. Específicamente, se observa un incremento del 36% en el coeficiente de determinación R², lo que indica que una mayor proporción de la variabilidad en la demanda de energía ahora puede explicarse con la adición de esta información temporal. Además, esta modificación ha conseguido una reducción significativa del 26% en el MAE, demostrando que las predicciones del modelo son considerablemente más precisas con la integración de datos del consumo energético previo.

También cabe mencionar que la variable de ocupación por conteo parece haber perdido su significatividad al incorporar Q_{t-1}, lo que justifica hacer uso de la función stepAIC para ver si puede ser eliminada definitivamente.

Después de llevar a cabo esta técnica se determina que sí puede ser eliminada quedando el modelo hasta ahora óptimo como se representa en la ecuación 30:

$$Q_t = \beta_0 + \beta_1 DSPT_{in,t} + \beta_2 R_t + \beta_3 Q_{t-1} \quad (31)$$

En la tabla 23, se da un resumen de las características principales del modelo:

Variables	Coeficientes	p-valor	p-valor (modelo)	R ²	R ² Ajus	MAE (kWh)
Intercepto	-0,01126	1,51E-02	2,2E-16	0,608	0,6069	0,133057
Diferencia SP y Tin	-0,05219	1,69E-02				
Radiación solar	0,001509	9,78E-05				
Energía consumida climatización	0,5006	2,25E-02				

Tabla 23. Resumen de los aspectos más importantes del modelo óptimo con solo variables exógenas en el momento t y energía consumida para la climatización en t-1 después de stepAIC

La aplicación del método de selección de variables en el modelo ha sido claramente exitosa, dado que se ha logrado eliminar una variable, simplificando así la complejidad del modelo sin sacrificar la precisión. La eliminación no solo ha reducido la redundancia y el riesgo de sobreajuste, sino que, además, ha mantenido o incluso mejorado ligeramente las métricas clave de rendimiento. Tanto el coeficiente de determinación (R²) como el Error Absoluto Medio (MAE) han permanecido prácticamente iguales, lo que demuestra que las variables restantes en el modelo conservan la capacidad de explicar la variabilidad de los datos y predecir con efectividad. Este equilibrio entre simplicidad y eficacia es indicativo de un modelo bien ajustado y optimizado.

En las ilustraciones 10 y 11, se muestra una comparación del modelo inicial y el modelo optimizado hasta el momento respectivamente.

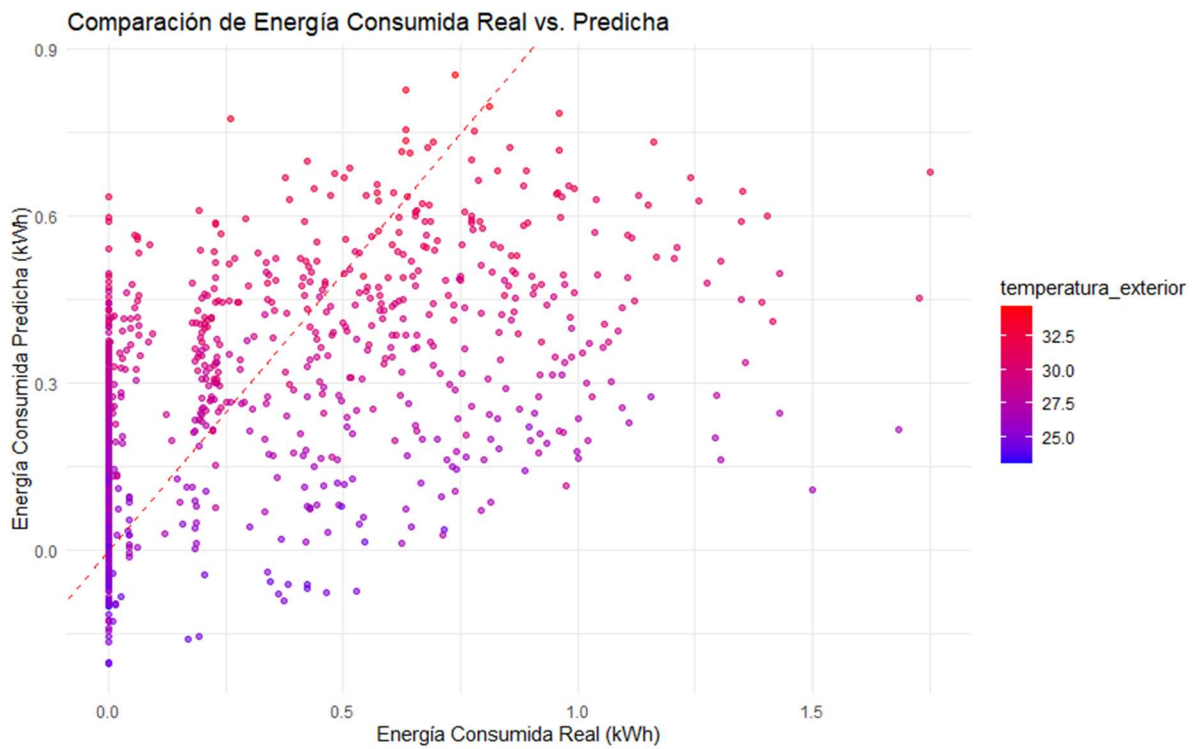


Ilustración 10. Representación gráfica de la energía consumida para la climatización medida y la energía consumida predicha para el modelo inicial.

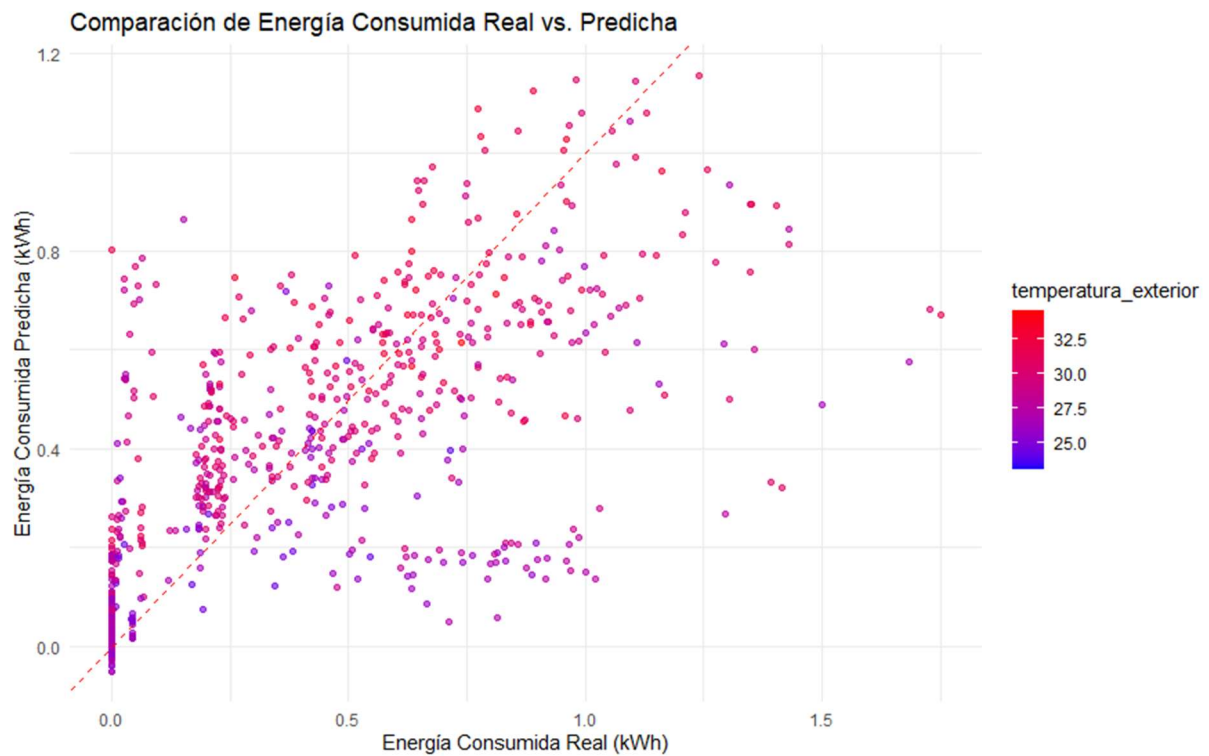


Ilustración 11. Representación gráfica de la energía consumida para la climatización medida y la energía consumida predicha para el modelo ARX optimizado.

6.5.5.4 Modelo final optimizado

Este proceso de inclusión de variables, aplicando hasta dos periodos de retraso (pasos anteriores), se ha realizado de forma exhaustiva para todas las variables seleccionadas. No obstante, con el objetivo de no incurrir en redundancias y dado que los ajustes más significativos ya han sido descritos, se ha decidido enfocarse directamente en el modelo óptimo. Este modelo ideal integra exclusivamente las variables que contribuyen a mejorar la precisión de las predicciones sin aumentar innecesariamente la complejidad del modelo.

Durante este análisis, añadir distintas variables con retrasos no ha logrado reducir el Error Absoluto Medio por debajo de los niveles ya obtenidos. A pesar de que la inclusión de nuevas variables eleva el coeficiente R^2 hasta un máximo de 0.62, esta mejora no es suficientemente significativa en comparación con un modelo más simple que únicamente incorpora la radiación solar y la diferencia entre el setpoint y la temperatura interior en el momento de la predicción, así como el consumo de energía para la climatización una hora antes de la predicción. Por lo tanto, se concluye que este último constituye el modelo óptimo, ya que captura de manera más eficaz y eficiente la energía que será consumida para la climatización en la próxima hora.

En la ilustración 12 se representa el con un intervalo de confianza del 95% que la predicción del modelo estará aleja $[-0.44; 0.44]$ kWh lo que es un intervalo bastante amplio. El motivo principal de este intervalo tan grande es la presencia de una considerable cantidad de outliers que están interfiriendo en una mejor predicción.

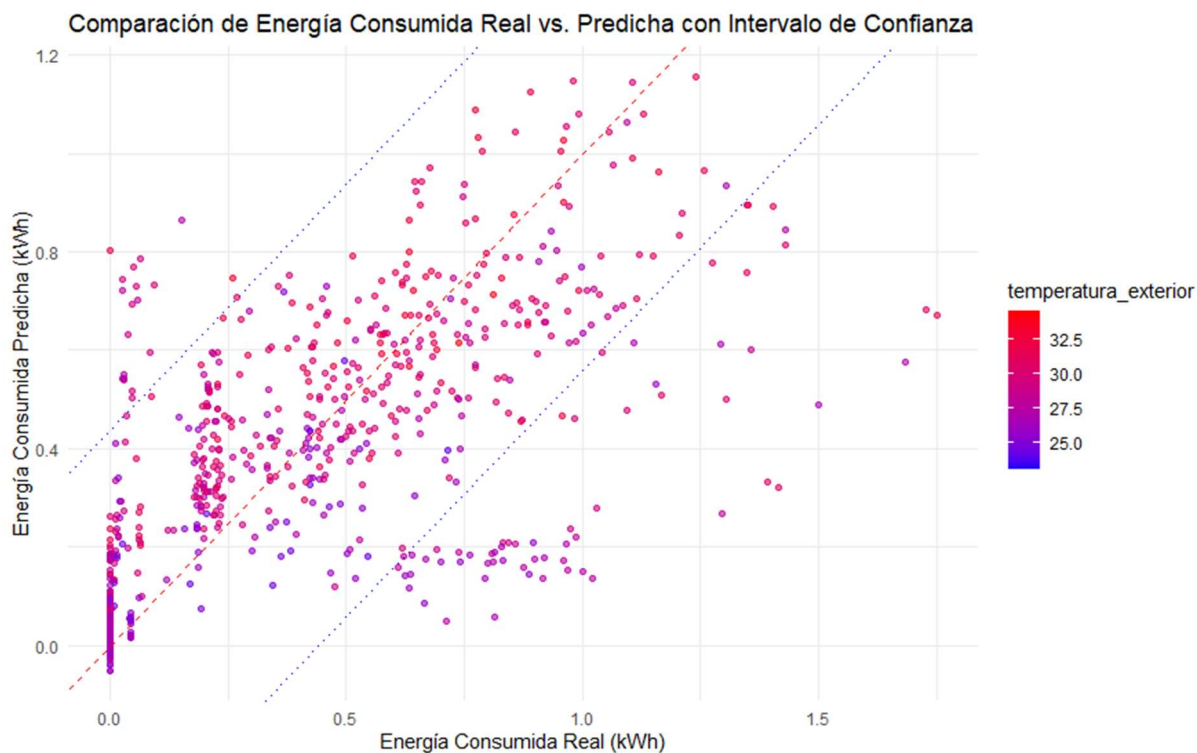


Ilustración 12. Representación gráfica de la energía consumida para la climatización medida y la energía consumida predicha para el modelo ARX optimizado, con intervalos de confianza

6.5.6 Análisis de los residuos

Antes de proceder con la exploración y visualización gráfica de los residuos del modelo, se ha realizado una prueba de Durbin-Watson para verificar la independencia de estos, obteniendo un resultado de 1,98. Este valor, cercano a 2, sugiere que no hay una correlación significativa entre los residuos consecutivos, indicando que la autocorrelación no es un problema importante en este modelo. Este paso es fundamental para asegurar la fiabilidad del análisis de regresión. Con esta confirmación, podemos proceder con confianza a explorar visualmente los residuos en relación con todas las variables exógenas involucradas, buscando identificar patrones sistemáticos o tendencias que puedan indicar problemas específicos con el modelo.

Confirmada la ausencia de autocorrelación en los residuos, la última fase del análisis se centra en explorar cómo y cuándo el modelo podría estar fallando en sus predicciones. Al graficar los residuos frente a cada una de las variables independientes, como la temperatura exterior, la radiación solar, la ocupación por conteo, y la diferencia entre el setpoint y la temperatura interior, se busca identificar cualquier patrón sistemático o tendencias que puedan indicar problemas específicos con el modelo. Para facilitar esta exploración, se han creado gráficos de dispersión de los residuos contra cada variable exógena, complementados con líneas de tendencia suavizadas. Esta técnica de análisis gráfico no solo permite una identificación visual intuitiva de los patrones en los datos, sino que también proporciona una base para posibles ajustes o refinamientos del modelo.

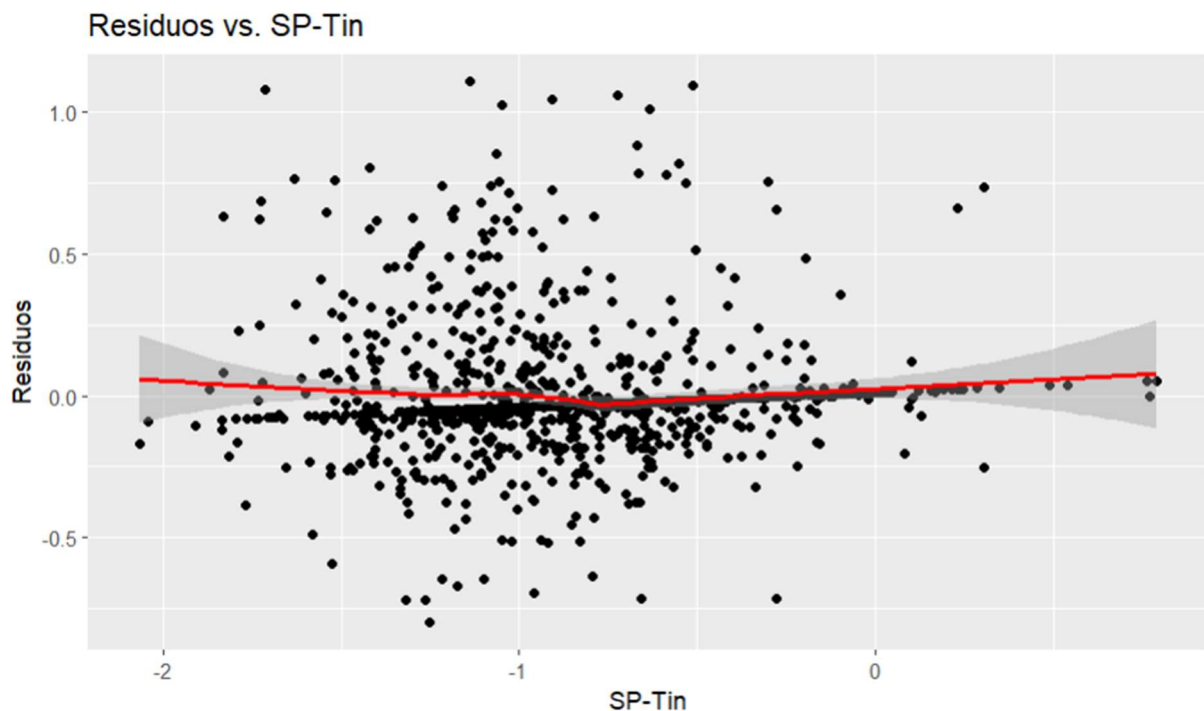


Ilustración 13. Gráfico de dispersión de los residuos frente a la diferencia entre la consigna y la temperatura interior

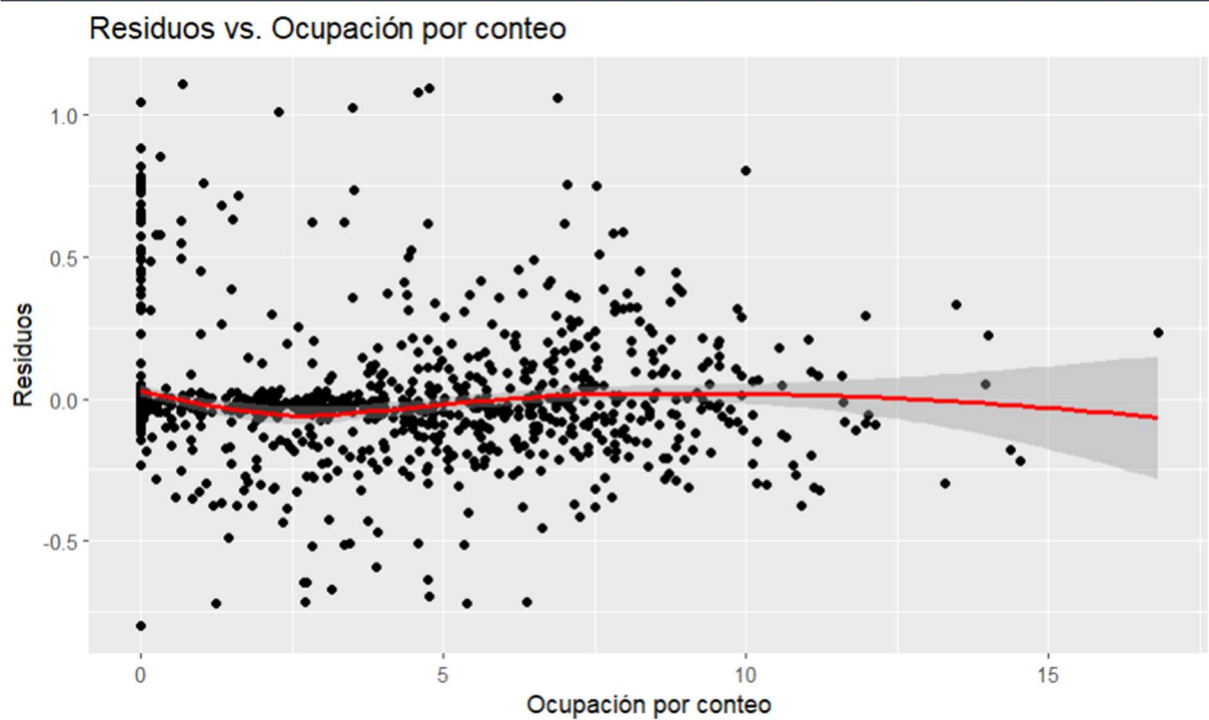


Ilustración 14. Gráfico de dispersión de los residuos frente a la ocupación por conteo

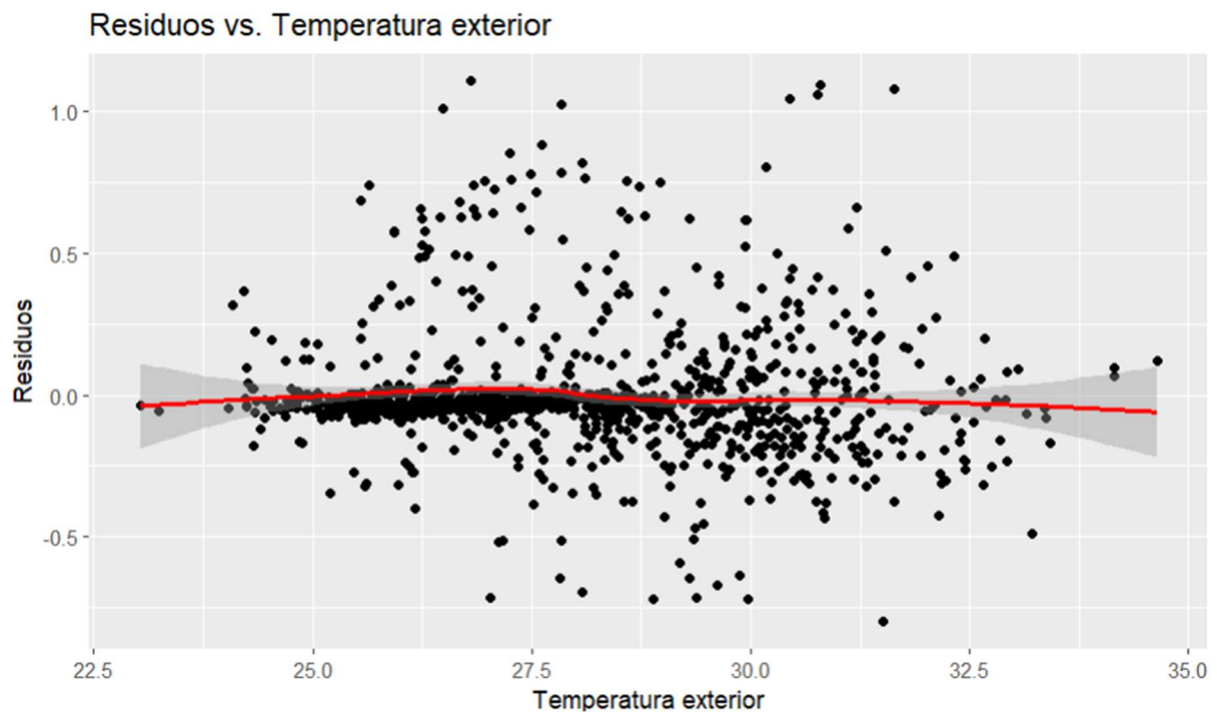


Ilustración 15. Gráfico de dispersión de los residuos frente a la temperatura exterior

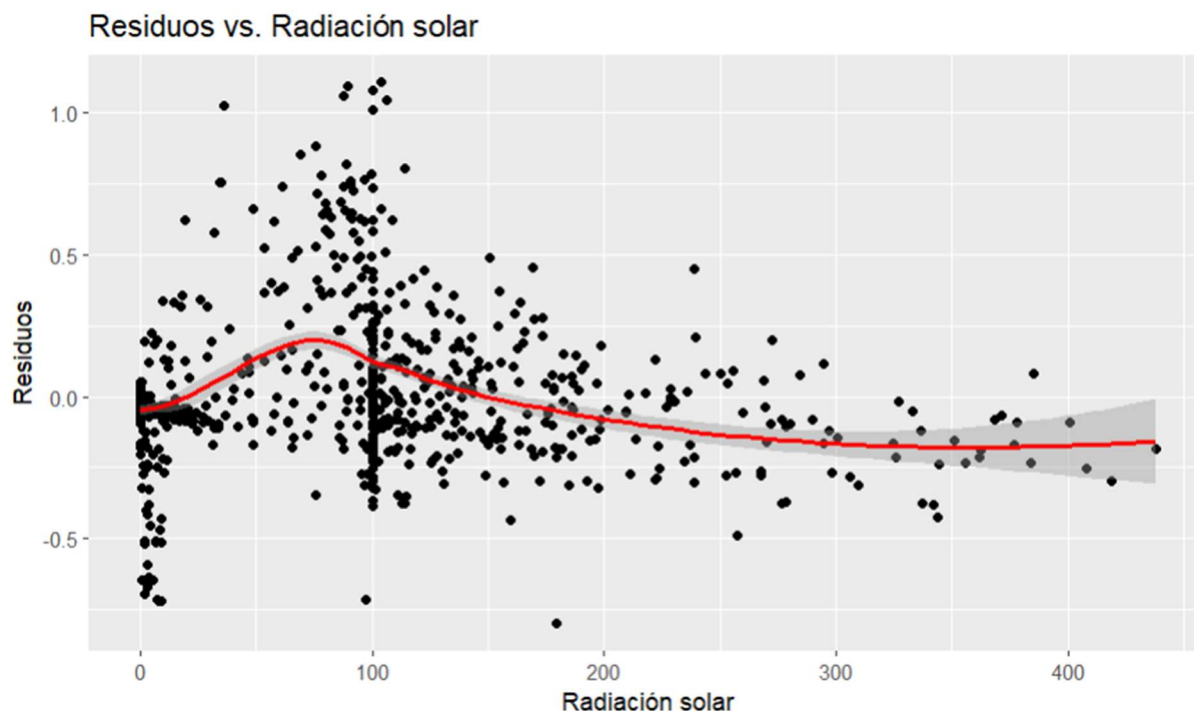


Ilustración 16. Gráfico de dispersión de los residuos frente a la radiación solar

De todos estos gráficos resalta el que relaciona la radiación solar y los residuos. Este gráfico revela dos patrones de error claros que coinciden con los protocolos programados de activación y desactivación del sistema de climatización. El primer patrón emerge cuando la radiación solar es 0 W/m², que normalmente ocurre alrededor de las 20:30. Este es el momento programado para que el sistema se apague automáticamente, no necesariamente porque no se requiera climatización, sino como parte de la estrategia de gestión energética del edificio. Los errores observados en este punto sugieren que el modelo no está adecuadamente calibrado para anticipar este cese programado en el uso de energía, lo que puede llevar a predicciones inexactas de la demanda energética durante la noche.

El segundo patrón se nota cuando la radiación solar alcanza cerca de 90-100 W/m², aproximadamente a las 7:30 de la mañana, momento en el cual el sistema de climatización se activa según lo programado para responder a las necesidades de climatización del día. Los errores detectados en este punto indican que el modelo puede no estar preparado para modelar el incremento abrupto en el consumo de energía que se produce cuando el sistema comienza a operar en respuesta a las condiciones externas cambiantes.

En las ilustraciones 17 y 18 se puede ver muy claramente la relación explicada. Las líneas punteadas en rojo indican las 7:30 y las 20:30 para cada día de la semana observándose una correlación muy clara.

También se observa como la radiación solar se estabiliza en alrededor de 100 W/m² cuando en esas primeras horas de radiación, lo que coincide a la perfección con los patrones observados.

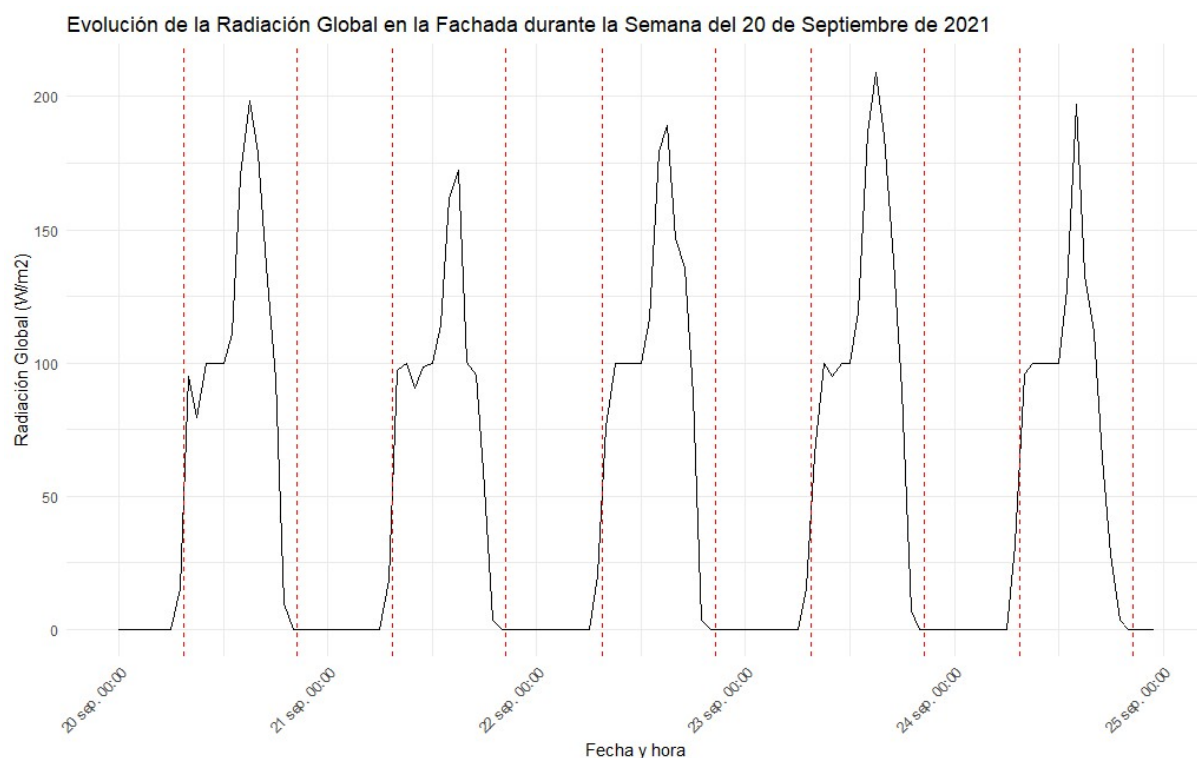


Ilustración 17. Evolución de la radiación para la semana del 20 septiembre de 2021. Las líneas rojas punteadas se colocan sobre las 7:30 y 20:30.

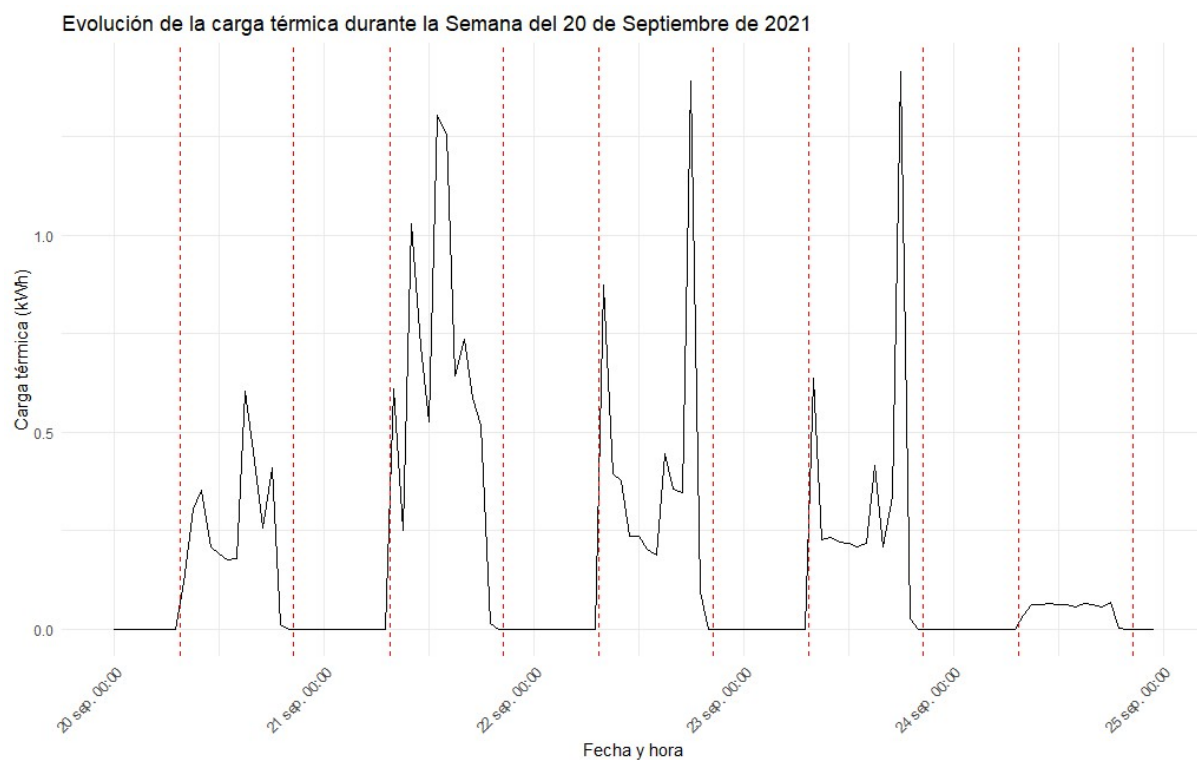


Ilustración 18. Evolución de la carga térmica para la semana del 20 septiembre de 2021. Las líneas rojas punteadas se colocan sobre las 7:30 y 20:30.

7. CONCLUSIONES Y TRABAJO FUTURO

El presente trabajo se ha abordado en dos etapas fundamentales por los descubrimientos encontrados a medida que se ha ido desarrollando la investigación inicial.

La **primera etapa** se ha enfocado en mejorar los modelos ARX de predicción de la temperatura interior en edificios, basados en unos modelos de referencia para la habitación 4 del edificio SDE4 de la Universidad Nacional de Singapur. Se ha realizado un análisis detallado para depurar variables no significativas. No obstante, **se ha llegado a la conclusión** de que, para el contexto específico del edificio de la Universidad Nacional de Singapur, **predecir la temperatura interior no resulta óptimo**.

Esta decisión se basa principalmente en **dos factores**:

- El edificio está climáticamente controlado, manteniendo la temperatura interior dentro de un rango constante de 25 a 28 grados Celsius.
- La temperatura una hora antes resulta ser un predictor demasiado dominante para la temperatura de la hora siguiente, lo que limita la utilidad de desarrollar un modelo predictivo bajo estas circunstancias.

Ante esta situación, se ha optado por **reorientar** la investigación **hacia el desarrollo de modelos que predicen la energía necesaria para la climatización** para el próximo período de una hora, en lugar de la temperatura interior. Este enfoque permite abordar directamente el consumo de energía y la eficiencia del sistema de climatización, relacionándolo con las variables climáticas exteriores y la señal de ocupación por conteo.

Para refinar aún más el modelado de la carga térmica, se ha **incorporado** una **nueva variable exógena** significativa que mide la **diferencia entre la consigna recibida por el controlador y la temperatura interior actual**. Esta nueva variable ha demostrado ser significativa para mejorar la precisión y relevancia del modelo de predicción.

El desarrollo del modelo ha culminado con la creación de una versión optimizada que equilibra eficazmente la complejidad y la precisión. Este balance se ha logrado mediante un método mixto de extensión, combinando el ajuste manual de variables con la selección automática basada en el criterio de información de Akaike, utilizando la función stepAIC de R. El modelo finalizado exhibe un coeficiente de determinación **R² de 0.606** y un **Error Medio Absoluto (MAE) de 0.133 kWh**, indicativos de un desempeño correcto.

Un análisis detallado de los residuos revela que, aunque el modelo es robusto, aún **podría mejorarse ajustando** cómo maneja **las transiciones de encendido y apagado repentinas del sistema de control térmico** del edificio. Estas transiciones, que reflejan los ciclos operativos diarios y que no responden a las dinámicas ambientales externas e internas, si se modelan con mayor precisión, podrían elevar significativamente la exactitud del modelo.

El trabajo futuro de este proyecto tiene varias vertientes:

- Uno de los principales objetivos para la continuación de este estudio es desarrollar y perfeccionar técnicas que permitan modelar de manera precisa los momentos de encendido y apagado del sistema de climatización. Estos momentos son críticos porque representan picos de demanda energética que no siempre son anticipados correctamente por los modelos convencionales. Una mejor caracterización de estos eventos podría resultar en una gestión más eficiente de la energía y en una reducción de costos operativos significativa.
- Para verificar la aplicabilidad y robustez de los modelos desarrollados, sería interesante aplicarlos en otras habitaciones del mismo edificio o en edificios con características similares. Esta estrategia permitirá evaluar la capacidad de los modelos para generalizarse y funcionar de manera efectiva en distintos entornos arquitectónicos que compartan similitudes en tamaño, uso (como comercial, residencial, o institucional) y condiciones climáticas. Esta aplicación ampliada también facilitará la identificación de ajustes necesarios para optimizar los modelos según las particularidades de cada nuevo contexto.
- Para explorar la mejora en la precisión y efectividad de los modelos desarrollados se propone utilizar muestras temporales más reducidas, como intervalos de 30 minutos, en futuras investigaciones. Esta aproximación podría permitir capturar mejor las variaciones a corto plazo en las condiciones ambientales y la respuesta dinámica de los sistemas de climatización. Al analizar datos con una granularidad temporal más alta, es posible que los modelos puedan ajustarse de manera más precisa a las fluctuaciones rápidas

8. VALORACION ETICA DEL PROYECTO

El presente trabajo, centrado en el desarrollo de modelos de predicción de carga térmica en edificaciones, ha sido conscientemente dirigido por un enfoque que equilibra la búsqueda de innovación técnica con un firme compromiso hacia los principios éticos que sustentan la práctica profesional y la investigación en ingeniería. Enfrentar el desafío de la eficiencia energética en edificios no solo requiere soluciones técnicas avanzadas, sino también una reflexión profunda sobre su impacto social y ambiental.

ODSs:

Este proyecto se ha alineado intencionalmente con los Objetivos de Desarrollo Sostenible de las Naciones Unidas [37], enfatizando la importancia de la energía asequible y no contaminante, así como la necesidad de promover ciudades y comunidades sostenibles. Mejorando la eficiencia con la que los edificios consumen energía, el trabajo ayuda a reducir la cantidad de recursos energéticos necesarios y, por ende, contribuye a disminuir la emisión de gases contaminantes. Este es un paso fundamental hacia la mitigación del impacto humano en el cambio climático y hacia la promoción de un desarrollo urbanístico que respeta los límites de nuestro entorno natural.

La implantación de los modelos desarrollados en sistemas de climatización no solo podría aumentar la eficiencia operativa, sino que también promueve un consumo de energía más consciente y medido. La implementación de estos sistemas puede reducir significativamente los costos operativos y el desperdicio de recursos, lo que está directamente alineado con la promoción de patrones de consumo y producción responsables. Por tanto, el proyecto no solo se enfoca en la innovación técnica, sino que también aborda la necesidad de sostener prácticas que benefician la economía y el medio ambiente a largo plazo.

Código deontológico de los ingenieros:

La adopción del código deontológico de los ingenieros guía todas las acciones dentro de este proyecto, resaltando la integridad, responsabilidad y respeto por la sociedad. Además, en concordancia con las directrices éticas que se derivan del código deontológico de los ingenieros, este proyecto se ha esforzado en salvaguardar el bienestar y los derechos de todos los involucrados, manteniendo siempre una perspectiva de respeto y justicia. Siguiendo este código, se ha asegurado que todas las prácticas involucradas respeten la integridad, la autonomía y el bienestar de las personas afectadas por la implementación y resultados del estudio. Este compromiso va más allá de la mera observancia de las leyes vigentes; se extiende a una constante búsqueda de mejora y beneficio comunitario, priorizando la seguridad, salud y bienestar sobre cualquier beneficio tecnológico o científico aislado.

En lo que respecta al uso de inteligencia artificial, en particular las tecnologías generativas para la redacción y mejora de contenidos se han aplicado rigurosamente las tecnologías para asegurar que la información redactada sea de máxima calidad y fácil comprensión, pero siempre bajo la revisión exhaustiva de expertos. Esto demuestra un uso ético de la IA, donde no sustituye la generación de ideas o la toma de decisiones críticas, sino que funciona como una herramienta de apoyo para mejorar la eficacia comunicativa.

En términos de ética investigativa y transparencia, este estudio emplea R, un entorno de programación de código abierto, destacando el compromiso con la transparencia y la accesibilidad en la investigación. El uso de R no sólo facilita la investigación replicable y transparente, sino que también promueve la democratización del conocimiento. Todos los scripts y análisis realizados están disponibles para su escrutinio y uso por parte de la comunidad científica, reafirmando nuestro compromiso con un conocimiento abierto y accesible que cualquiera puede verificar y utilizar para fomentar nuevas investigaciones.

Además, este proyecto también refleja un profundo respeto por los principios de sostenibilidad tecnológica. Al proporcionar libre acceso a los métodos y resultados de la investigación, se fomenta un entorno de colaboración y crecimiento compartido que no sólo beneficia a la comunidad científica sino también al público en general. Este enfoque no sólo ayuda a mitigar los riesgos de dependencia tecnológica y desigualdad en el acceso a la información, sino que también subraya la importancia de una sociedad informada y empoderada en la era digital.

Este amplio compromiso con la ética profesional, la transparencia y la responsabilidad social asegura que el proyecto no solo cumpla con los estándares científicos y técnicos, sino que también sirva como un modelo de integridad en la investigación, preparando el camino para futuras innovaciones que sean éticamente sólidas y socialmente responsables.

Es también crucial reflexionar sobre la implementación de las tecnologías desarrolladas, considerando sus potenciales impactos éticos. La autonomía de los sistemas controlados por modelos predictivos precisa ser gestionada cuidadosamente para evitar que la dependencia de la automatización eclipse la toma de decisiones humanas, especialmente en situaciones que podrían requerir juicios éticos complejos. Además, es imperativo que cualquier implementación de nuevas tecnologías sea acompañada de políticas que garanticen la equidad y el acceso igualitario, asegurando que los beneficios de la innovación sean compartidos ampliamente y no solo limitados a entornos privilegiados.

En conclusión, este estudio no solo cumple con los estándares científicos y técnicos, sino que también se erige como un modelo de cómo la integridad y la ética pueden y deben integrarse en la investigación para fomentar innovaciones que sean sostenibles, éticamente sólidas y socialmente responsables, preparando el camino para un futuro donde la tecnología y la humanidad avanzan juntas hacia un desarrollo más sostenible.

9. RIESGOS

En el desarrollo de modelos predictivos de carga térmica en edificaciones, existen diversos riesgos que deben ser cuidadosamente gestionados para asegurar la precisión y aplicabilidad del proyecto.

- **Riesgo de Inexactitud en la Predicción:** uno de estos riesgos es la posibilidad de que los modelos no alcancen la precisión deseada, posiblemente debido a la subestimación de la complejidad de las dinámicas térmicas del edificio o las influencias ambientales. Para mitigar este riesgo, sería interesante realizar pruebas continuas y extensivas durante las etapas de desarrollo, utilizando conjuntos de datos variados para entrenar y validar el modelo, de los que actualmente no se dispone.
- **Carga Computacional Excesiva:** otro riesgo significativo es la carga computacional excesiva que pueden requerir los modelos avanzados, lo que limitaría su aplicabilidad práctica. La optimización de algoritmos para una mejor eficiencia computacional, junto con el uso de software y hardware especializados para el procesamiento de datos, puede reducir este riesgo.
- **Generalización del Modelo:** la posibilidad de que el modelo esté demasiado ajustado a las condiciones específicas de un único edificio también presenta un riesgo para la generalización de los resultados. Para aumentar la robustez y la generalización del modelo, sería interesante incluir datos de otros edificios con diferentes características y climas, y realizar pruebas cruzadas en estos variados entornos, asegurará que el modelo mantenga su validez y utilidad en un rango más amplio de aplicaciones.

Además de los riesgos relacionados con los resultados y la eficacia del proyecto, los riesgos operativos como el incumplimiento de plazos pueden retrasar la finalización del proyecto.

- **Incumplimiento de Plazos:** establecer un cronograma realista con revisiones periódicas y mantener una comunicación regular con el tutor facilita la gestión efectiva del tiempo y la rápida identificación y resolución de cualquier retraso.
- **Dependencia de Tecnología Específica:** la dependencia de tecnologías específicas en este caso, el software R o de un ordenador con suficiente potencia computacional, es otro riesgo operativo, que se puede mitigar manteniendo la documentación de los procedimientos y códigos utilizados bien organizada en la nube, facilitando así la transición a nuevas plataformas si es necesario.
- **Pérdida de información:** la pérdida de datos es un riesgo considerable que puede ser contrarrestado implementando un sistema robusto de copias de seguridad y medidas de seguridad cibernética. Utilizar almacenamiento en la nube junto con soluciones locales, y seguir prácticas recomendadas de seguridad informática, protegerá los datos de fallos técnicos, errores humanos o ataques externos, asegurando la integridad y disponibilidad de la información crítica del proyecto.

10. PLANIFICACIÓN Y PRESUPUESTO

10.1 PLANIFICACIÓN

En este apartado se incluye un diagrama de Gantt, herramienta para la visualización de la planificación y el seguimiento de las distintas etapas del proyecto.

El diagrama de Gantt presentado refleja la planificación detallada del proyecto desde su inicio el 16 de octubre hasta su conclusión el 20 de junio. Cada fila del gráfico representa una tarea específica dentro del proyecto, y se subdivide en varias subtareas e hitos importantes, lo cual facilita la organización y la asignación de responsabilidades.

Las principales características mostradas en el diagrama incluyen:

- **Nombre de la tarea:** indica la actividad específica o conjunto de actividades que deben realizarse.
- **Duración:** representada en el gráfico por la longitud de las barras horizontales, muestra el tiempo estimado para completar cada tarea.
- **Fecha de comienzo y finalización:** cada barra comienza en la fecha de inicio de la tarea y termina en la fecha de finalización prevista, proporcionando una visión clara del cronograma.
- **Recursos asignados:** señala quiénes participan en cada etapa del proyecto, diferenciando entre roles como el del estudiante y el director.

Es importante destacar que, aunque en el diagrama de Gantt algunas tareas aparecen con una duración en días significativamente mayor que otras, esto no implica necesariamente una mayor carga de trabajo. La distribución del número de horas trabajadas está diseñada para ser bastante equitativa entre las distintas tareas, independientemente de su duración en días. Esto se debe a que algunas tareas, aunque prolongadas, requieren menos horas diarias de dedicación, mientras que otras pueden necesitar un esfuerzo intensivo en un período más corto.

El número total de horas ocupadas por el estudiante en cada uno de los bloques es el siguiente:

- **Definición de objetivos:** 20 horas
- **Revisión bibliográfica:** 100 horas
- **Revisión y mejora del modelo de referencia:** 75 horas
- **Desarrollo del nuevo modelo predictivo:** 80 horas
- **Conclusiones y trabajo futuro:** 25 horas

El número total de horas ocupadas por el director en cada uno de los bloques es el siguiente:

- **Definición de objetivos:** 8 horas
- **Revisión bibliográfica:** 5 horas
- **Revisión y mejora del modelo de referencia:** 10 horas
- **Desarrollo del nuevo modelo predictivo:** 10 horas
- **Conclusiones y trabajo futuro:** 7 horas

El diagrama de Gantt está representado en las ilustraciones 19 y 20.

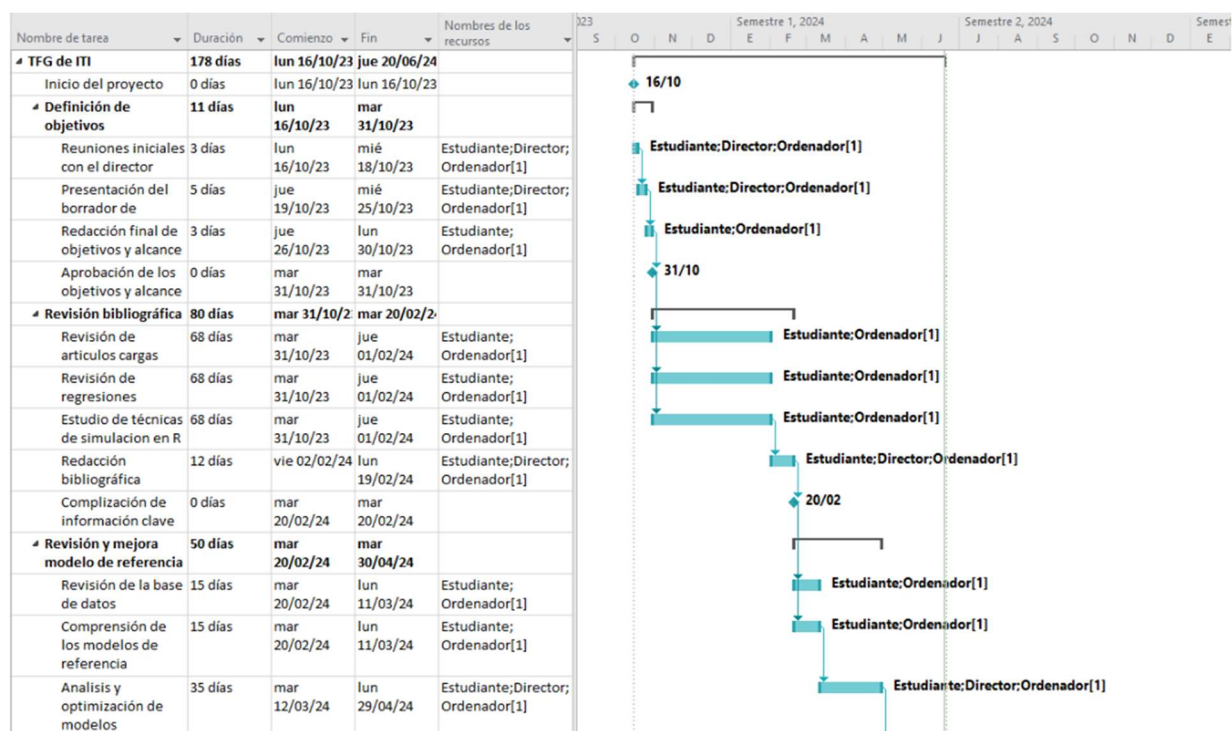


Ilustración 19. Diagrama de Gantt para la planificación del proyecto

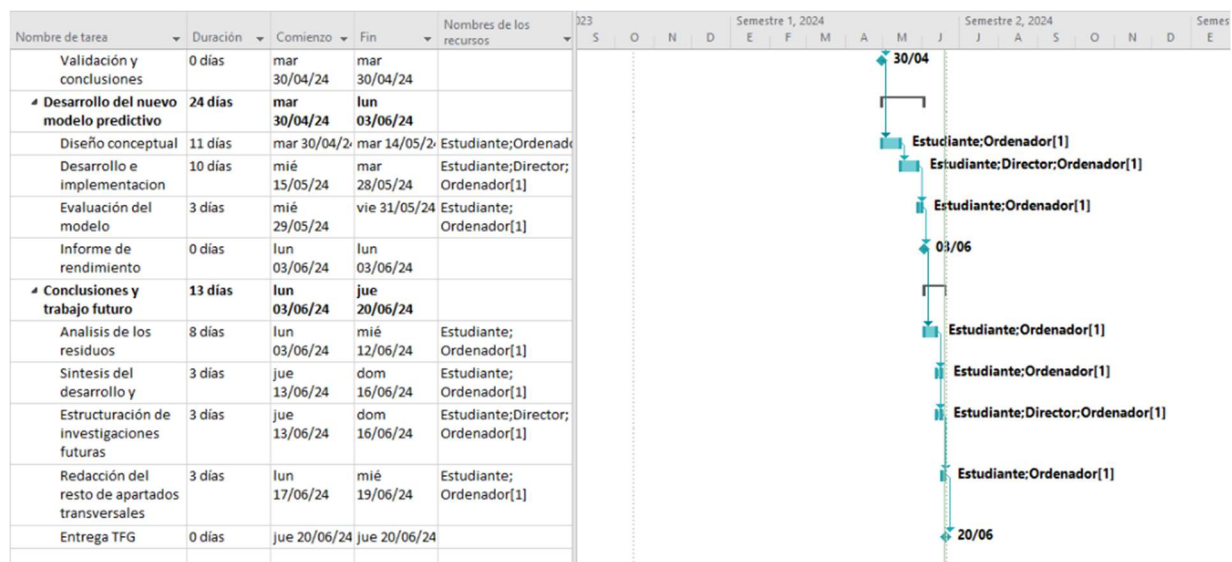


Ilustración 20. Diagrama de Gantt para la planificación del proyecto

10.2 PRESUPUESTO

Para el cálculo del presupuesto del proyecto, se han asignado los siguientes costes a los recursos que ha requerido.

Recursos humanos:

- Coste por hora trabajada del estudiante: 36€/h
- Coste por hora trabajada del director: 100€/h

Empleo de recursos humanos:

ESTADÍSTICAS DE RECURSOS

Estado de trabajo de todos los recursos de trabajo.

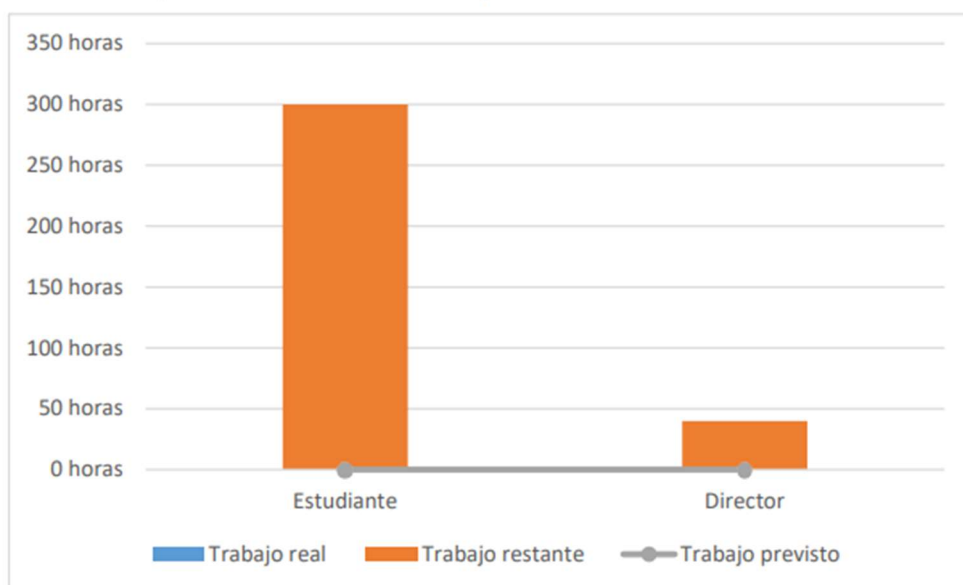


Ilustración 21. Información general sobre las horas empleadas por estudiante y director

Recursos físicos:

- Ordenador personal. Su coste se ha calculado tomando una inversión inicial de 900 euros y suponiendo una amortización en 3 años.

En la tabla 24 se observa el cálculo detallado del presupuesto con el coste por hora y las horas requeridas en el caso del estudiante y el director, y de la amortización en el caso de ordenador personal.

Presupuesto total:

Recurso	Tiempo empleado	Coste por hora	Amortización anual	Coste total
Estudiante	300 horas	36 €/h	-	10.800 €
Director	40 horas	100€/h	-	4.000 €
Ordenador	8 meses	-	300€/año	200€
TOTAL				15.000€

Tabla 24. Desglose del presupuesto por empleo de recursos

11. BIBLIOGRAFÍA

- [1] «DEUSTO SUSTAINABLE RESEARCH GROUP | Investigación | Deusto». Accedido: 13 de junio de 2024. [En línea]. Disponible en: <https://www.deusto.es/en/home/research/whos-who/research-teams/deusto-sustainable-research-group-2/equipoinvestiga>
- [2] M. Moran, «Energía», Desarrollo Sostenible. Accedido: 13 de junio de 2024. [En línea]. Disponible en: <https://www.un.org/sustainabledevelopment/es/energy/>
- [3] U. Nations, «Net Zero Coalition | Naciones Unidas», United Nations. Accedido: 13 de junio de 2024. [En línea]. Disponible en: <https://www.un.org/es/climatechange/net-zero-coalition>
- [4] «2020 Buildings GSR_FULL REPORT.pdf». Accedido: 12 de junio de 2024. [En línea]. Disponible en: https://globalabc.org/sites/default/files/inline-files/2020%20Buildings%20GSR_FULL%20REPORT.pdf
- [5] «Building sector can play key role in combating global warming, UN report says | UN News». Accedido: 12 de junio de 2024. [En línea]. Disponible en: <https://news.un.org/en/story/2007/03/213932>
- [6] M. Elnour *et al.*, «Neural network-based model predictive control system for optimizing building automation and management systems of sports facilities», *Applied Energy*, vol. 318, n.º C, 2022, Accedido: 12 de junio de 2024. [En línea]. Disponible en: <https://ideas.repec.org/a/eee/appene/v318y2022ics0306261922005268.html>
- [7] T. Cholewa *et al.*, «On the forecast control of heating system as an easily applicable measure to increase energy efficiency in existing buildings: Long term field evaluation», *Energy and Buildings*, vol. 292, p. 113174, ago. 2023, doi: 10.1016/j.enbuild.2023.113174.
- [8] H. Liu, J. Liang, Y. Liu, y H. Wu, «A Review of Data-Driven Building Energy Prediction», *Buildings*, vol. 13, n.º 2, Art. n.º 2, feb. 2023, doi: 10.3390/buildings13020532.
- [9] G. Xu y W. Wang, «China's energy consumption in construction and building sectors: An outlook to 2100», *Energy*, vol. 195, p. 117045, mar. 2020, doi: 10.1016/j.energy.2020.117045.
- [10] K. Song, N. Kwon, K. Anderson, M. Park, H.-S. Lee, y S. Lee, «Predicting hourly energy consumption in buildings using occupancy-related characteristics of end-user groups», *Energy and Buildings*, vol. 156, sep. 2017, doi: 10.1016/j.enbuild.2017.09.060.
- [11] J. Kim y K.-I. Kim, «Data-driven hybrid model and operating algorithm to shave peak demand costs of building electricity», *Energy and Buildings*, vol. 229, p. 110493, dic. 2020, doi: 10.1016/j.enbuild.2020.110493.
- [12] N. Dawood, «Short-Term Prediction of Energy Consumption in Demand Response for Blocks of Buildings: DR-BoB Approach», *Buildings*, vol. 9, n.º 10, Art. n.º 10, oct. 2019, doi: 10.3390/buildings9100221.

- [13] M. J. Risbeck, M. Z. Bazant, Z. Jiang, Y. M. Lee, K. H. Drees, y J. D. Douglas, «Modeling and multiobjective optimization of indoor airborne disease transmission risk and associated energy consumption for building HVAC systems», *Energy Build*, vol. 253, p. 111497, dic. 2021, doi: 10.1016/j.enbuild.2021.111497.
- [14] Z. Li, H. Ye, N. Liao, R. Wang, Y. Qiu, y Y. Wang, «Impact of COVID-19 on electricity energy consumption: A quantitative analysis on electricity», *International Journal of Electrical Power & Energy Systems*, vol. 140, p. 108084, sep. 2022, doi: 10.1016/j.ijepes.2022.108084.
- [15] C. Fan, L. Xiao, y Y. Zhao, «A short-term building cooling load prediction method using deep learning algorithms», *Applied Energy*, vol. 195, pp. 222-233, jun. 2017, doi: 10.1016/j.apenergy.2017.03.064.
- [16] A. Mohammed, A. Alshibani, O. Alshamrani, y M. Hassanain, «A regression-based model for estimating the energy consumption of school facilities in Saudi Arabia», *Energy and Buildings*, vol. 237, p. 110809, abr. 2021, doi: 10.1016/j.enbuild.2021.110809.
- [17] D. Yan, J. Xia, W. Tang, F. Song, X. Zhang, y Y. Jiang, «DeST — An integrated building simulation toolkit Part I: Fundamentals», *Build. Simul.*, vol. 1, n.º 2, pp. 95-110, jun. 2008, doi: 10.1007/s12273-008-8118-8.
- [18] «A review of data-driven approaches for prediction and classification of building energy consumption - ScienceDirect». Accedido: 16 de junio de 2024. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S136403211731362X?via%3Dihub>
- [19] A. Afram y F. Janabi-Sharifi, «Review of modeling methods for HVAC systems», *Applied Thermal Engineering*, vol. 67, n.º 1, pp. 507-519, jun. 2014, doi: 10.1016/j.applthermaleng.2014.03.055.
- [20] C. A. Thilker, P. Bacher, D. Cali, y H. Madsen, «Identification of non-linear autoregressive models with exogenous inputs for room air temperature modelling», *Energy and AI*, vol. 9, p. 100165, ago. 2022, doi: 10.1016/j.egyai.2022.100165.
- [21] G. J. Ríos-Moreno, M. Trejo-Perea, R. Castañeda-Miranda, V. M. Hernández-Guzmán, y G. Herrera-Ruiz, «Modelling temperature in intelligent buildings by means of autoregressive models», *Automation in Construction*, vol. 5, n.º 16, pp. 713-722, 2007, doi: 10.1016/j.autcon.2006.11.003.
- [22] L. Mba, P. Meukam, y A. Kemajou, «Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region», *Energy and Buildings*, vol. 121, pp. 32-42, jun. 2016, doi: 10.1016/j.enbuild.2016.03.046.
- [23] F. Ferracuti *et al.*, «Data-driven models for short-term thermal behaviour prediction in real buildings», *Applied Energy*, vol. 204, pp. 1375-1387, oct. 2017, doi: 10.1016/j.apenergy.2017.05.015.
- [24] M. Lumbreras, G. Diarce, K. Martin, R. Garay-Martinez, y B. Arregi, «Unsupervised recognition and prediction of daily patterns in heating loads in buildings», *Journal of Building Engineering*, vol. 65, p. 105732, abr. 2023, doi: 10.1016/j.jobee.2022.105732.

- [25] Z. Dong, J. Liu, B. Liu, K. Li, y X. Li, «Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification», *Energy and Buildings*, vol. 241, p. 110929, mar. 2021, doi: 10.1016/j.enbuild.2021.110929.
- [26] M. Lumbreras *et al.*, «Data driven model for heat load prediction in buildings connected to District Heating by using smart heat meters», *Energy*, vol. 239, p. 122318, oct. 2021, doi: 10.1016/j.energy.2021.122318.
- [27] M. Eguizabal, R. Garay-Martinez, y I. Flores-Abascal, «Simplified model for the short-term forecasting of heat loads in buildings», *Energy Reports*, vol. 8, pp. 79-85, dic. 2022, doi: 10.1016/j.egyr.2022.10.224.
- [28] I. García de Eulate Pascual, «Desarrollo de modelos de caracterización a corto plazo de la evolución térmico-energética de edificios», 2023. doi: 10.5281/zenodo.8320272.
- [29] Z. D. Tekler, E. Ono, Y. Peng, S. Zhan, B. Lasternas, y A. Chong, «ROBOD, room-level occupancy and building operation dataset», *Build. Simul.*, vol. 15, n.º 12, pp. 2127-2137, dic. 2022, doi: 10.1007/s12273-022-0925-9.
- [30] «Regresión lineal», *Wikipedia, la enciclopedia libre*. 24 de mayo de 2024. Accedido: 13 de junio de 2024. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Regresi%C3%B3n_lineal&oldid=160320830
- [31] «Regresión no lineal», *Wikipedia, la enciclopedia libre*. 5 de diciembre de 2023. Accedido: 13 de junio de 2024. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Regresi%C3%B3n_no_lineal&oldid=155836069
- [32] «Introduction to the Fundamentals of Time Series Data and Analysis | Aptech». Accedido: 13 de junio de 2024. [En línea]. Disponible en: <https://www.aptech.com/blog/introduction-to-the-fundamentals-of-time-series-data-and-analysis/>
- [33] C. Shetty, «Time Series Models», Medium. Accedido: 13 de junio de 2024. [En línea]. Disponible en: <https://towardsdatascience.com/time-series-models-d9266f8ac7b0>
- [34] «SPSS Modeler Subscription». Accedido: 13 de junio de 2024. [En línea]. Disponible en: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=networks-neural-model>
- [35] «Criterio de información de Akaike», *Wikipedia, la enciclopedia libre*. 3 de enero de 2024. Accedido: 14 de junio de 2024. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Criterio_de_informaci%C3%B3n_de_Akaike&oldid=156586167
- [36] J. Duffie y W. Beckman, *Solar Engineering of Thermal Processes*. 2006. doi: 10.1002/9781118671603.

- [37] M. J. Gamez, «Objetivos y metas de desarrollo sostenible», Desarrollo Sostenible. Accedido: 17 de junio de 2024. [En línea]. Disponible en: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>
- [38] V. Subramanyam, Md. Ahiduzzaman, y A. Kumar, «Greenhouse gas emissions mitigation potential in the commercial and institutional sector», *Energy and Buildings*, vol. 140, pp. 295-304, abr. 2017, doi: 10.1016/j.enbuild.2017.02.007.
- [39] D. D. Kim y H. S. Suh, «Heating and cooling energy consumption prediction model for high-rise apartment buildings considering design parameters», *Energy for Sustainable Development*, vol. 61, pp. 1-14, abr. 2021, doi: 10.1016/j.esd.2021.01.001.
- [40] S. D. Atalay, «Performance analyses of statistical approaches for modeling electricity consumption of a commercial building in France», *Energy and Buildings*, ene. 2019, Accedido: 12 de junio de 2024. [En línea]. Disponible en: https://www.academia.edu/85613995/Performance_analyses_of_statistical_approaches_for_modeling_electricity_consumption_of_a_commercial_building_in_France
- [41] P. R. S. Jota, V. R. B. Silva, y F. G. Jota, «Building load management using cluster and statistical analyses», *International Journal of Electrical Power and Energy Systems*, vol. 8, n.º 33, pp. 1498-1505, 2011, doi: 10.1016/j.ijepes.2011.06.034.
- [42] E. Sala-Cardoso, M. Delgado-Prieto, K. Kampouropoulos, y L. Romeral, «Activity-Aware HVAC Power Demand Forecasting», *Energy and Buildings*, vol. 170, abr. 2018, doi: 10.1016/j.enbuild.2018.03.087.
- [43] H. Liu, J. Liang, Y. Liu, y H. Wu, «A Review of Data-Driven Building Energy Prediction», *Buildings*, vol. 13, n.º 2, Art. n.º 2, feb. 2023, doi: 10.3390/buildings13020532.
- [44] Y. Chen, M. Guo, Z. Chen, Z. Chen, y Y. Ji, «Physical energy and data-driven models in building energy prediction: A review», *Energy Reports*, vol. 8, pp. 2656-2671, nov. 2022, doi: 10.1016/j.egyr.2022.01.162.
- [45] R. M. Granados, «Modelos de regresión lineal múltiple».
- [46] M. Zhang, «Time Series: Autoregressive models AR, MA, ARMA, ARIMA».