



PREDICCIÓN DE LA DEMANDA ENERGÉTICA EN HUILA: COMPARACIÓN DE MODELOS PREDICTIVOS*

Jorge Eduardo Barón Méndez**
Diego Alejandro Manrique Cabezas***

Recibido: 16 de noviembre de 2025- Aprobado: 14 de marzo de 2026

DOI: <https://doi.org/10.22395/seec.v29n67a5326>

RESUMEN

Este estudio analiza la demanda energética del departamento del Huila entre 2023 y 2025 mediante modelos estadísticos y de aprendizaje automático, con el propósito de evaluar su capacidad predictiva y aportar evidencia para la comprensión de la dinámica reciente del consumo eléctrico regional. La base de datos utilizada incluye series diarias y horarias del consumo eléctrico, lo que permitió examinar patrones estacionales, tendencias de crecimiento y variaciones de la demanda asociadas a distintos comportamientos temporales. En el análisis se implementaron modelos ARIMA, SARIMAX, regresión lineal, Support Vector Regression (SVR), Gaussian Process Regression (GPR) y Random Forest, y se evaluó su desempeño mediante métricas de error como MAE, RMSE, MAPE, MASE y R^2 . Los resultados muestran que los modelos estadísticos tradicionales, en particular ARIMA y SARIMAX, presentaron un desempeño más limitado frente a la complejidad y no linealidad de la serie analizada. En contraste, los modelos de aprendizaje automático alcanzaron mejores niveles de ajuste, y se destacaron Random Forest, con un RMSE de 14.137 unidades y un R^2 de 0,67, lo que lo posiciona como el modelo de mejor desempeño entre los evaluados para la predicción horaria. En este sentido, los hallazgos permiten identificar el potencial de los enfoques de aprendizaje automático para fortalecer ejercicios de pronóstico de demanda energética a escala regional.

PALABRAS CLAVE

Predicción, aprendizaje automático, energías renovables, modelos estadísticos, transición energética.

CÓDIGO JEL

Q41; C53; Q42

CONTENIDO

Introducción, 1. Marco teórico, 2. Metodología, 3. Resultados y discusión, 4. Conclusiones, Referencias.

* Artículo de investigación derivado del trabajo de grado titulado "Predicción de la demanda energética en el departamento del Huila: Una comparación de modelos de *machine learning* y métodos estadísticos", presentado como requisito de grado para obtener el título de magíster en Economía, de la Universidad de Manizales.

** Profesional en Finanzas y Comercio Internacional, Universidad del Rosario, Bogotá, Colombia. Magíster en Economía, Universidad de Manizales, Manizales, Colombia. Correo electrónico: jbaronmendez@gmail.com. Orcid: <https://orcid.org/0009-0007-9371-8386>

*** Economista, Universidad del Quindío, Armenia, Colombia. Magíster en Economía, Universidad de Manizales, Manizales, Colombia. Magíster en Ingeniería – Automatización Industrial, Universidad Nacional de Colombia, Manizales, Colombia. PhD (C) en Ingeniería Automática, Universidad Nacional de Colombia, Manizales, Colombia. Profesor e investigador, Departamento de Administración y Economía, Universidad Autónoma de Manizales, Manizales, Colombia. Correo electrónico: diegoa.manriquec@autonoma.edu.co. Orcid: <https://orcid.org/0000-0002-6493-3657>

ENERGY DEMAND FORECASTING IN HUILA: A COMPARISON OF PREDICTIVE MODELS

ABSTRACT

This study analyzes the energy demand of the department of Huila between 2023 and 2025 using statistical and machine learning models, with the aim of evaluating their predictive capacity and providing evidence for understanding the recent dynamics of regional electricity consumption. The database used includes daily and hourly electricity consumption series, which allowed the examination of seasonal patterns, growth trends, and demand variations associated with different temporal behaviors. The analysis implemented ARIMA, SARIMAX, linear regression, Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Random Forest models, and their performance was evaluated using error metrics such as MAE, RMSE, MAPE, MASE, and R^2 . The results show that traditional statistical models, particularly ARIMA and SARIMAX, exhibited more limited performance in dealing with the complexity and nonlinearity of the analyzed series. In contrast, machine learning models achieved better levels of fit, with Random Forest standing out, obtaining an RMSE of 14.137 units and an R^2 of 0.67, positioning it as the best-performing model among those evaluated for hourly forecasting. In this regard, the findings highlight the potential of machine learning approaches to strengthen regional-scale energy demand forecasting exercises.

KEYWORDS

Forecasting; machine learning; renewable energy; statistical models; energy transition.

JEL CLASSIFICATION

Q41; C53; Q42

CONTENTS

Introduction; 1. Theoretical framework; 2. Methodology; 3. Results and discussion; 4. Conclusions; References.

PREDIÇÃO DA DEMANDA ENERGÉTICA EM HUILA: COMPARAÇÃO DE MODELOS PREDITIVOS

RESUMO

Este estudo analisa a demanda energética do departamento de Huila entre 2023 e 2025 por meio de modelos estatísticos e de aprendizado de máquina, com o propósito de avaliar sua capacidade preditiva e fornecer evidências para a compreensão da dinâmica recente do consumo elétrico regional. A base de dados utilizada inclui séries diárias e horárias do consumo elétrico, o que permitiu examinar padrões sazonais, tendências de crescimento e variações da demanda associadas a diferentes comportamentos temporais. Na análise, foram implementados modelos ARIMA, SARIMAX, regressão linear, Support Vector Regression (SVR), Gaussian Process Regression (GPR) e Random Forest, e seu desempenho foi avaliado por meio de métricas de erro como MAE, RMSE, MAPE, MASE e R^2 . Os resultados mostram que os modelos estatísticos tradicionais, em particular ARIMA e SARIMAX, apresentaram desempenho mais limitado diante da complexidade e da não linearidade da série analisada. Em contraste, os modelos de aprendizado de máquina alcançaram melhores níveis de ajuste, destacando-se o Random Forest, com um RMSE de 14,137 unidades e um R^2 de 0,67, o que o posiciona como o modelo de melhor desempenho entre os avaliados para a previsão horária. Nesse sentido, os achados permitem identificar o potencial das abordagens de aprendizado de máquina para fortalecer exercícios de previsão da demanda energética em escala regional.

PALAVRAS-CHAVE

Predição; aprendizado de máquina; energias renováveis; modelos estatísticos; transição energética.

CÓDIGO JEL

Q41; C53; Q42

CONTEÚDO

Introdução; 1. Marco teórico; 2. Metodologia; 3. Resultados e discussão; 4. Conclusões; Referências.

INTRODUCCIÓN

El departamento del Huila posee condiciones climáticas favorables y un alto potencial solar, especialmente en su zona norte, lo cual lo convierte en un territorio con ventajas competitivas para el desarrollo de energías renovables (Ibarra *et al.*, 2024). A pesar de ello, la demanda energética local continúa siendo abastecida principalmente a través del Sistema Interconectado Nacional, lo que refleja una dependencia estructural de la oferta externa (Torres *et al.*, 2017). La tendencia creciente del consumo plantea desafíos para la seguridad energética regional, ante los cuales se vuelve imperativa la diversificación de la matriz mediante la incorporación de fuentes renovables, especialmente tecnologías fotovoltaicas, dado que este proceso permitiría no solo atender la creciente demanda, sino también avanzar hacia un sistema más sostenible, resiliente y competitivo (Pineda *et al.*, 2020).

En esta línea, la energía eléctrica es un recurso fundamental para el desarrollo económico y social de cualquier región, al incidir directamente en la productividad, el bienestar de la población y la sostenibilidad del crecimiento (Zhang *et al.*, 2023). Sin embargo, el modelo energético global aún se sostiene en gran medida sobre fuentes no renovables –como petróleo, gas natural y carbón–, que representan aproximadamente el 50 % de la energía producida en el mundo (Zucaro y Agostinho, 2025). Esta dependencia conlleva importantes desafíos ambientales, especialmente por las emisiones de gases de efecto invernadero, la contaminación atmosférica y el agotamiento de recursos naturales, que contribuyen al cambio climático y ponen en riesgo la sostenibilidad de los sistemas energéticos (Li *et al.*, 2026).

En este contexto, el análisis riguroso de la demanda energética se convierte en un componente estratégico para la gestión eficiente del sector eléctrico (Alshater *et al.*, 2025). Los pronósticos precisos permiten planificar la infraestructura de generación y distribución, optimizar la programación de despacho y garantizar la estabilidad del suministro (Ahmed *et al.*, 2025). Además, aportan información fundamental para la formulación de políticas públicas, el diseño de tarifas, la integración de fuentes renovables y la implementación de medidas de eficiencia energética (Kaloop *et al.*, 2025). Esta visión de planificación energética basada en datos favorece una transición gradual hacia sistemas más sostenibles y resilientes, que respondan a las tendencias de electrificación y crecimiento económico sin comprometer la seguridad energética (Hasan *et al.*, 2025).

La energía solar fotovoltaica emerge como una alternativa técnica y ambientalmente viable para abordar estos desafíos. Aprovechando la radiación solar –una fuente limpia, inagotable y ampliamente disponible en el Huila– es posible diversificar la matriz energética, reducir las emisiones contaminantes y fortalecer

la independencia de los sistemas locales frente a la volatilidad del mercado de combustibles fósiles (Pineda *et al.*, 2020). La integración de sistemas solares también genera beneficios económicos directos, como el desarrollo de cadenas industriales, la generación de empleo y el aumento de la competitividad regional, al disminuir los costos marginales de generación y reducir la necesidad de importación energética (Ibarra *et al.*, 2024).

Por su parte, la planificación energética se erige como una herramienta fundamental para garantizar la continuidad, confiabilidad y seguridad del suministro eléctrico (Li *et al.*, 2025). En Colombia, esta tarea incluye la formulación de pronósticos confiables de corto, mediano y largo plazo que permitan anticipar la demanda y equilibrarla con la oferta disponible (Torres *et al.*, 2017). La regulación vigente –establecida por la Comisión de Regulación de Energía y Gas (en adelante, CREG)– exige a los operadores de red presentar proyecciones semanales de energía y potencia horaria, cuya precisión afecta directamente los costos operativos, la estabilidad de la red y la calidad del servicio; de hecho, pronósticos deficientes generan riesgos de redespacho, sobrecargas o energía no suministrada, con lo que se aumentan los costos y se reduce la confiabilidad del sistema (CREG, 2024).

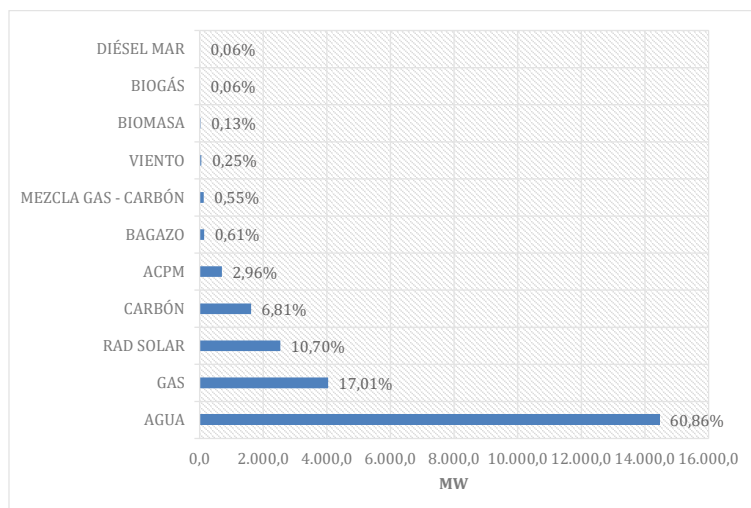
La crisis del apagón de 1992-1993 –ocasionada por el fenómeno de El Niño y la reducción de los niveles de los embalses hidroeléctricos– evidenció la vulnerabilidad estructural del sistema eléctrico colombiano y la falta de previsión en la gestión energética (CREG, 2024). Este episodio fue un punto de inflexión que impulsó profundas reformas regulatorias orientadas a modernizar el sector, diversificar la matriz de generación y fortalecer la gestión técnica del sistema. Desde entonces, la planificación energética se ha consolidado como una función crítica para el adecuado funcionamiento del mercado eléctrico nacional (Torres *et al.*, 2017).

COMPOSICIÓN DE LA CAPACIDAD EFECTIVA NETA DE GENERACIÓN ELÉCTRICA EN COLOMBIA

La figura 1 evidencia la marcada dependencia que tiene el sistema energético colombiano de la generación hidroeléctrica, que representa el 60,86 % de la capacidad efectiva neta instalada. Este predominio del recurso hídrico ha sido históricamente una ventaja competitiva por su bajo costo marginal y carácter renovable, sin embargo, también constituye la principal vulnerabilidad estructural del sistema frente a fenómenos climáticos como El Niño, que reduce los niveles de los embalses y obliga a utilizar fuentes térmicas más costosas. En segundo lugar, se encuentra el gas natural con un 17,01 %, seguido por la radiación solar fotovoltaica con el 10,70 %, lo cual demuestra un avance importante en

la transición energética, aunque aún insuficiente para compensar los riesgos asociados a la variabilidad hidrológica. Las fuentes basadas en carbón (6,81 %) y ACPM (2,96 %) continúan cumpliendo un rol de respaldo, pero presentan impactos ambientales y costos crecientes.

Figura 1. Participación por tipo de generación en la capacidad efectiva neta de generación por MW y fuente de generación eléctrica en Colombia, 2025.



Fuente: elaboración propia con base en XM, (2026)

En ese sentido, las tecnologías emergentes como biomasa, biogás, eólica y mezclas gas-carbón aún tienen participaciones marginales inferiores al 1 %, lo que refleja la necesidad de robustecer la diversificación energética. El crecimiento de la energía solar, que ya supera el 10 %, indica un cambio estructural en la matriz, especialmente a partir de las subastas de renovables y la mayor competitividad de las tecnologías fotovoltaicas. No obstante, el predominio hidroeléctrico plantea la urgencia de incorporar almacenamiento energético, sistemas híbridos y pronósticos avanzados de demanda y generación para evitar futuros episodios de racionamiento. En síntesis, la matriz colombiana transita hacia la diversificación, pero continúa siendo altamente dependiente del recurso hídrico, lo que exige estrategias de resiliencia frente al cambio climático y un mayor impulso a las renovables no convencionales.

Por su parte, la generación eléctrica en Colombia alcanza una capacidad efectiva neta total de 23.752,4 megavatios (MW) en 2025, como se presenta en la figura 1, distribuida entre diversas fuentes energéticas que reflejan la histórica dependencia

del país de los recursos hídricos y, al mismo tiempo, el avance progresivo hacia la incorporación de energías renovables no convencionales. A la fecha del estudio, el sistema eléctrico nacional cuenta con 492 generadores distribuidos en 25 departamentos, lo que evidencia una cobertura territorial amplia y una diversificación creciente de la infraestructura de generación, impulsada por la expansión de proyectos solares, eólicos y de biomasa que complementan el predominio tradicional de la generación hidroeléctrica. La principal fuente de generación sigue siendo el agua, con 14.455 MW que representan el 60,86 % del total, y eso sustenta el papel dominante de las centrales hidroeléctricas en la oferta nacional. Le sigue el gas natural con 4.040 MW, equivalente al 17,01 %, empleado principalmente en plantas termoeléctricas que aportan respaldo al sistema en épocas de baja hidrología (Torres *et al.*, 2017). La radiación solar alcanza 2.540,7 MW (10,70 %), lo cual refleja un crecimiento significativo de los proyectos fotovoltaicos en regiones con alto potencial solar. El carbón aporta 1.618,7 MW (6,81 %), con lo que mantiene una participación relevante, aunque en declive por su impacto ambiental. Otras fuentes menores incluyen el ACPM con 702,5 MW (2,96 %), utilizado sobre todo en plantas de respaldo y zonas aisladas; el bagazo de caña con 145,9 MW (0,61 %), derivado del aprovechamiento de residuos agroindustriales; la mezcla gas-carbón con 130 MW (0,55 %), que combina ambas fuentes para optimizar costos y disponibilidad; el viento con 59,4 MW (0,25 %), que aunque aún incipiente, señala el inicio del aprovechamiento eólico; la biomasa con 30 MW (0,13 %); el biogás con 15,2 MW (0,06 %), y el diésel marino con 15 MW (0,06 %), ambos con participación marginal, pero importantes en el suministro local y diversificación energética. Esta composición evidencia un sistema predominantemente hidráulico, con una tendencia creciente hacia fuentes renovables y tecnologías limpias (Rojas, 2024).

Hoy, la sostenibilidad energética es un eje articulador de las políticas públicas y del desarrollo territorial. Estudios recientes destacan que la regulación nacional ha promovido la cooperación entre agentes del sector para avanzar hacia una matriz baja en carbono (Reyes *et al.*, 2023). De manera complementaria, regiones como el Huila, con alto potencial solar, requieren una planificación energética precisa para orientar inversiones en infraestructura renovable (Olivares, 2025). Por su parte, los modelos de pronóstico de demanda permiten equilibrar oferta y consumo, posibilitando la implementación de políticas de transición energética con criterios de eficiencia y sostenibilidad (López *et al.*, 2024). De otro lado, es importante destacar los beneficios ambientales y económicos de desarrollar sistemas fotovoltaicos regionales que favorecen la autonomía energética y reducen la dependencia de la generación hidroeléctrica (Rojas, 2024).

Desde la perspectiva económica, la planificación energética se sustenta en el principio de eficiencia, entendido como la asignación óptima de recursos para maximizar el bienestar social (Rocha, 2024). Este enfoque plantea que la previsión de la demanda contribuye a garantizar tarifas justas, sostenibilidad regional y uso racional de la infraestructura (Osorio, 2025). La literatura económica coincide en que eficiencia y equidad son dimensiones complementarias del bienestar, y que alcanzar un óptimo de Pareto requiere sistemas tarifarios coherentes con los costos reales y la eliminación de subsidios que distorsionan el consumo (Olivares, 2025).

Adicionalmente, diversos estudios demuestran que los pronósticos de demanda energética no solo permiten optimizar la generación y distribución, sino que también contribuyen a la reducción de emisiones y a la consolidación de sistemas energéticos más resilientes (Zucaro y Agostinho, 2025). Como señalan Olivares (2025) y Rojas (2024), disponer de modelos predictivos robustos resulta clave para alcanzar metas de sostenibilidad ambiental, asegurar el abastecimiento energético y adaptarse a las demandas de un sistema eléctrico en permanente transformación.

En este contexto, el objetivo de la presente investigación fue predecir el comportamiento de la demanda energética del departamento del Huila entre los años 2023 y 2025, con el fin de generar insumos técnicos que apoyen la toma de decisiones estratégicas en materia de planificación energética. Para ello, se utilizaron modelos estadísticos y de aprendizaje automático orientados a evaluar la capacidad predictiva de distintos enfoques frente a la dinámica temporal del consumo energético regional. Los resultados se plantean como una herramienta de apoyo para fortalecer el análisis técnico de la demanda y aportar evidencia útil para procesos de planificación energética en el ámbito departamental. Por su parte, este estudio realiza una aplicación empírica al caso del departamento del Huila a partir de información diaria y horaria de consumo energético, un ámbito poco abordado en la literatura regional. Así mismo, compara de manera sistemática el desempeño de modelos estadísticos tradicionales y modelos de aprendizaje automático, lo que permite identificar diferencias relevantes en su capacidad predictiva. Entre los principales hallazgos, se evidencia la superioridad del modelo Random Forest frente a las demás técnicas evaluadas, especialmente en la predicción horaria de la demanda.

1. MARCO TEÓRICO

La teoría de la demanda, formulada por Marshall (1890) y desarrollada posteriormente por autores como Hicks y Allen (1934) y Samuelson y Nordhaus (2010), establece que, *ceteris paribus*, existe una relación inversa entre el precio de un bien y la cantidad demandada, explicada por los efectos sustitución e ingreso (Wang, 2025). En este

marco, el precio, el ingreso del consumidor, los precios de bienes relacionados, las preferencias, las expectativas y el número de consumidores son factores que determinan la cantidad demandada (Ahmed *et al.*, 2025). En el caso de la energía, la demanda presenta particularidades que la diferencian de los bienes ordinarios, dado que su consumo no responde al deseo directo de obtener energía, sino a los servicios que esta proporciona, como iluminación, movilidad o calor, por lo que se considera una demanda derivada (Alshater *et al.*, 2025). La energía cumple además una doble función: es un bien final en el consumo doméstico y un insumo intermedio esencial en los procesos productivos. A corto plazo, la demanda de energía es inelástica, con elasticidades-precio cercanas a -0,05, debido a la rigidez tecnológica y la falta de sustitutos inmediatos; a largo plazo, esta elasticidad aumenta por la posibilidad de incorporar eficiencia energética, tecnologías alternativas y esquemas de generación distribuida basados en energías renovables de autoconsumo; la elasticidad ingreso es positiva, lo cual refleja que el consumo energético crece con el ingreso y el producto interno bruto. Factores adicionales como el clima, la política energética, la innovación tecnológica y las preferencias ambientales también inciden en su comportamiento (Al Kez *et al.*, 2024).

Desde una perspectiva de recursos, Hotelling (1931) explicó que los precios de los recursos no renovables deben aumentar a la tasa de interés, mientras Georgescu-Roegen destacó los límites ecológicos del crecimiento (Wang, 2025). La energía, además de ser un bien esencial, genera externalidades negativas por su producción y consumo, lo que justifica la intervención estatal mediante impuestos o regulación. De esta forma, las externalidades negativas representan fallas de mercado que deben ser internalizadas a través de impuestos correctivos con el fin de alinear los costos privados y sociales del consumo de energía (Al Kez *et al.*, 2024). La demanda de energía refleja la interacción entre variables económicas clásicas y factores estructurales vinculados al desarrollo, la tecnología y la sostenibilidad ambiental (López *et al.*, 2024).

Por su parte, la predicción de la demanda energética se ha convertido en un componente estratégico para la planificación y operación de los sistemas eléctricos a nivel mundial (Rao *et al.*, 2024). En un contexto marcado por la transición hacia fuentes renovables, la electrificación de la economía y la digitalización de los sistemas energéticos, los pronósticos precisos son indispensables para garantizar el equilibrio entre oferta y demanda (Alshater *et al.*, 2025). Las empresas de energía, operadores de sistemas e instituciones regulatorias utilizan modelos avanzados, desde métodos estadísticos tradicionales hasta técnicas de aprendizaje automático, para anticipar variaciones en los patrones de consumo y asegurar la estabilidad del suministro en diferentes horizontes temporales (López *et al.*, 2024).

A escala global, los sistemas eléctricos enfrentan desafíos sin precedentes debido al crecimiento sostenido de la demanda, impulsado por el desarrollo económico, la urbanización y el aumento de los dispositivos eléctricos conectados (Rao *et al.*, 2024). En países altamente industrializados, la integración masiva de energía eólica y solar ha incrementado la necesidad de predicciones energéticas en tiempo real, a causa de la variabilidad inherente de estas fuentes; de hecho, en respuesta, los operadores de red han incorporado técnicas de predicción probabilística y sistemas de control basados en inteligencia artificial para gestionar la intermitencia sin comprometer la seguridad energética (Cui *et al.*, 2024).

Los países europeos lideran el desarrollo de modelos de predicción energética integrados, debido a su ambiciosa agenda de descarbonización; de hecho, Alemania, Dinamarca y Reino Unido utilizan modelos híbridos que combinan predicciones meteorológicas, variables de demanda histórica, datos de generación distribuida y aprendizaje profundo (Kulisz *et al.*, 2024). Estos métodos hacen posible ajustar el despacho energético diario y minimizar los costos, reduciendo así las emisiones y la dependencia de combustibles fósiles. En estos sistemas, la predicción no solo es técnica, sino también económica, ya que permite gestionar mercados eléctricos altamente dinámicos (Rao *et al.*, 2024).

En contraste, países en desarrollo enfrentan retos adicionales, como la limitada disponibilidad de datos históricos de calidad, redes eléctricas menos automatizadas y carencias en la infraestructura digital (Mohammed *et al.*, 2024). En regiones como África, gran parte de la demanda energética no está formalmente registrada, lo que impide desarrollar modelos robustos. Sin embargo, la adopción de tecnologías móviles, redes inteligentes y medición avanzada está comenzando a cerrar estas brechas, y eso permite construir predicciones, incluso en infraestructuras emergentes (Ansong *et al.*, 2024).

A nivel latinoamericano, la predicción energética se ha convertido en un instrumento central para la planificación de sistemas eléctricos en expansión; la región cuenta con una matriz relativamente limpia gracias a la alta participación de energía hidroeléctrica, pero también enfrenta vulnerabilidades asociadas a fenómenos climáticos como El Niño y La Niña; esta situación ha impulsado el desarrollo de modelos sensibles a variables hidrológicas y meteorológicas que permiten anticipar disponibilidad de generación y coordinar la operación entre centrales térmicas y renovables en momentos críticos (Reyes *et al.*, 2023).

En el caso colombiano, la predicción de la demanda energética es una obligación regulatoria y una necesidad operativa. El Sistema Interconectado Nacional está

estrechamente regulado por la CREG, que exige reportes de demanda y disponibilidad de energía a los agentes del mercado. Debido a la alta participación de energías hidroeléctricas, los modelos en Colombia deben incorporar, además de las dinámicas de uso industrial y residencial, las variables hidrológicas y climáticas que impactan los niveles de embalses y, por ende, la disponibilidad de generación (CREG, 2024).

En las últimas décadas, Colombia ha avanzado en la integración de fuentes renovables no convencionales, especialmente en energía solar y eólica. Este cambio ha incrementado la incertidumbre en la operación del sistema energético, haciendo que la predicción en tiempo real adquiera mayor relevancia. Modelos basados en *machine learning* y técnicas híbridas se están comenzando a utilizar en operadores de red como XM para integrar pronósticos de generación renovable con curvas de carga de usuarios. Así, se ajustan los balances del sistema con mayor precisión y se reducen costos de operación (CREG, 2024).

En el ámbito regional, departamentos colombianos como Valle del Cauca, Atlántico, Córdoba y Huila están experimentando cambios significativos en sus perfiles de demanda debido al crecimiento industrial, la electrificación de procesos productivos y el desarrollo de comunidades energéticas. En este contexto, la predicción local de demanda energética se convierte no solo en un ejercicio técnico, sino también en un insumo para la formulación de políticas de transición energética, incentivos fiscales y planificación de infraestructura de generación y distribución (Olivares, 2025).

En suma, la predicción energética a nivel mundial, latinoamericano y colombiano es una herramienta crítica para garantizar la sostenibilidad, seguridad y eficiencia de los sistemas eléctricos. Impulsa la integración de renovables, reduce los costos del sistema, optimiza la infraestructura y orienta la política pública hacia modelos más resilientes (Ansong *et al.*, 2024). El caso colombiano muestra que, si bien existen retos asociados a la dependencia hidroeléctrica y la variabilidad climática, el uso adecuado de modelos predictivos avanzados puede fortalecer la resiliencia del sistema y catalizar la transición energética en regiones con alto potencial renovable, como el Huila.

2. METODOLOGÍA

2.1 Composición de la base de datos

La base de datos utilizada corresponde al registro histórico de demanda energética del departamento del Huila entre enero de 2023 y noviembre de 2025. El conjunto de datos presenta una resolución temporal diaria con desagregación horaria completa,

organizada en 24 columnas que representan el consumo energético registrado en cada hora del día. Cada fila corresponde a un día calendario, lo que permite reconstruir tanto la curva de carga horaria, como el comportamiento agregado del consumo total por jornada. Esta estructura ofrece una base adecuada para la modelación de series temporales, al permitir identificar patrones estacionales intradiarios, variaciones semanales y dinámicas de mediano plazo en la demanda.

Además de las variables de consumo, la base incorpora información asociada a la fecha de registro y al código del distribuidor, lo que facilita la trazabilidad de los datos dentro del Sistema Interconectado Nacional. A partir de las 24 columnas horarias se construyó una variable adicional de consumo total diario, obtenida mediante la suma de los valores registrados en cada día. Este indicador permitió evaluar la evolución estructural de la demanda a lo largo del periodo analizado y sirvió como variable objetivo en los ejercicios de predicción agregada. Durante la etapa de revisión y depuración se verificó que la base no presentaba valores faltantes en el rango temporal estudiado, por lo cual no fue necesario aplicar procedimientos de imputación. Asimismo, se realizó una inspección general de consistencia para validar la continuidad temporal de los registros y la coherencia de los valores observados.

Como parte del proceso de preparación de los datos, se derivaron variables calendarizadas a partir de la fecha, entre ellas, año, mes, día, día de la semana y hora. Estas variables se incorporaron como predictores en los modelos de aprendizaje automático y regresión, dado que la demanda energética presenta un comportamiento fuertemente dependiente de la estructura temporal del consumo, incluyendo ciclos horarios, diferencias entre días hábiles y fines de semana, así como patrones asociados a la estacionalidad mensual. La selección de estas variables se fundamenta en su capacidad para capturar regularidades temporales de manera directa y en la disponibilidad homogénea de esta información para todo el horizonte de análisis. No obstante, se reconoce que la inclusión de variables exógenas adicionales, como temperatura, condiciones climáticas, actividad económica o eventos especiales, podría fortalecer futuros ejercicios de modelación al incorporar factores explicativos complementarios de la demanda.

La configuración del conjunto de datos permitió trabajar bajo diferentes enfoques de modelación según la naturaleza de cada técnica empleada. En particular, la serie agregada de consumo total diario se utilizó para los modelos ARIMA, mientras que para SARIMAX se analizaron series por hora específica con el fin de capturar comportamientos diferenciados a lo largo del día. Por su parte, para los modelos de aprendizaje automático, como Random Forest, Gaussian Process Regression y Support Vector Regression, se construyó una estructura

expandida tipo panel día-hora, en la cual cada observación combina información temporal y nivel de consumo. Esta estrategia hizo posible comparar el desempeño de enfoques tradicionales y no lineales bajo una misma base empírica, respetando la lógica de funcionamiento de cada familia de modelos.

En cuanto al proceso de validación, los modelos fueron evaluados mediante una separación temporal entre datos de entrenamiento y prueba, preservando el orden cronológico de la serie para evitar filtraciones de información futura hacia el ajuste de los modelos. Esta decisión metodológica resulta consistente con la naturaleza de los problemas de predicción en series temporales, en los cuales no es adecuado realizar particiones aleatorias convencionales. Sobre esta base, el desempeño predictivo se valoró utilizando métricas complementarias como MAE, RMSE, MAPE, MASE y R^2 , lo que permitió contrastar la precisión, estabilidad y capacidad explicativa de los distintos modelos bajo criterios homogéneos.

De esta forma, respecto al ajuste de los modelos de aprendizaje automático, la selección de hiperparámetros se orientó a identificar configuraciones que mejoraran el desempeño predictivo sin comprometer la capacidad de generalización. Aunque el estudio prioriza la comparación de resultados entre técnicas, resulta pertinente explicitar que el ajuste se realizó a partir de criterios de desempeño sobre el conjunto de entrenamiento y validación, buscando configuraciones consistentes con la estructura temporal de la demanda.

2.2 Modelos de Regresión y Aprendizaje Automático para la predicción del consumo energético en el departamento del Huila

Con el objetivo de identificar relaciones y patrones en los datos horarios de consumo energético del departamento del Huila, esta investigación evalúa un conjunto de modelos estadísticos y de aprendizaje automático ampliamente utilizados en problemas de predicción de series temporales. En particular, se emplean enfoques autorregresivos y técnicas de regresión supervisada con el fin de analizar la relación entre la variable dependiente, correspondiente a la demanda energética, y un conjunto de variables independientes derivadas de la estructura temporal de los datos. Los modelos seleccionados para este estudio fueron ARIMA, SARIMAX, regresión lineal, Random Forest, Gaussian Process Regression y Support Vector Regression, debido a que representan enfoques complementarios con diferentes capacidades para modelar tendencias, estacionalidad, no linealidad y variaciones complejas en la demanda (Bulungu y Kumar, 2024).

La inclusión de modelos como ARIMA y SARIMAX se justifica porque constituyen referentes clásicos en la literatura de predicción energética y series temporales,

especialmente útiles para capturar componentes autorregresivos, patrones estacionales y estructuras lineales de dependencia temporal (Bulungu y Kumar, 2024). Su incorporación permite establecer una línea base robusta de comparación frente a métodos más flexibles.

En cuanto a los modelos de aprendizaje automático, Random Forest fue seleccionado por su capacidad para capturar relaciones no lineales, interacciones complejas entre variables y estructuras de decisión robustas frente a datos con alta variabilidad. Gaussian Process Regression se incorporó por su flexibilidad probabilística y por su potencial para modelar relaciones no lineales, además de ofrecer una aproximación útil para estimar comportamientos suaves en la serie (Pérez *et al.*, 2025). Support Vector Regression, por su parte, ha sido ampliamente utilizado en problemas de predicción energética debido a su capacidad para manejar relaciones no lineales en espacios de alta dimensión y su buen desempeño en contextos con estructuras complejas y tamaños de muestra moderados (Zhu *et al.*, 2024).

La selección de estos modelos responde, por tanto, a un criterio comparativo y metodológico: incluir técnicas tradicionales ampliamente aceptadas en la literatura y contrastarlas con enfoques de aprendizaje automático que permiten capturar mejor la complejidad no lineal de la demanda energética (Pérez *et al.*, 2025). De esta manera, el estudio no busca agotar todas las alternativas disponibles, sino construir una comparación entre modelos representativos que permita evaluar diferencias de desempeño y establecer una base analítica consistente para el caso del Huila.

No obstante, se reconoce que en la literatura reciente sobre predicción de demanda energética han cobrado relevancias técnicas más avanzadas basadas en redes neuronales profundas, como Artificial Neural Networks (ANN) y Long Short-Term Memory (LSTM), debido a su capacidad para modelar dependencias temporales complejas y patrones no lineales de largo plazo (Ubal *et al.*, 2023). Sin embargo, estas técnicas no fueron incorporadas en el presente estudio debido a que el propósito principal fue comparar modelos estadísticos clásicos y métodos de aprendizaje automático supervisado con estructuras metodológicas más controlables e interpretables dentro del alcance de la investigación. Además, la inclusión de modelos profundos exige decisiones adicionales relacionadas con arquitectura, entrenamiento, regularización y requerimientos computacionales, lo que habría ampliado sustancialmente el alcance metodológico del trabajo. En este sentido, su ausencia no implica desconocer su relevancia, sino delimitar el estudio a un conjunto de modelos adecuados para una comparación inicial y consistente del comportamiento predictivo de la demanda energética regional.

2.3 Modelo Autorregresivo Integrado de Media Móvil (ARIMA)

Se utiliza un modelo ARIMA para detectar correlaciones estacionarias dentro de los datos de series temporales. La representación matemática se expresa mediante la siguiente ecuación, en la cual, por simplicidad, se presume que tanto los componentes autorregresivos como los de media móvil poseen rezagos idénticos:

$$y_t = \eta + \sum_{p=1}^{p'} \bar{\theta}_{p'} y_t - \rho' + \sum_{p'=1}^{p'} \bar{\theta}_{p'} \zeta_t - \rho' \quad [1]$$

En esta formulación, $y_t \in \mathbb{R}$ y representa la serie temporal en tiempo t , mientras que, $\eta \in \mathbb{R}$ representa una constante; el conjunto de observaciones que abarca los rezagos pertinentes se denota como $y = \{y_t \in \mathbb{R} : t \in \mathbb{N}\}$; además, los rezagos de la serie se representan mediante $\hat{X} = \{y_{t-p'} \in \mathbb{R} : t \in \mathbb{N}, p' \in P'\}$, mientras que los coeficientes asociados de los rezagos del componente autorregresivo se organizan en una matriz o vector de parámetros $\bar{\theta} = \{\bar{\theta}_{p'} \in \mathbb{R} : p' \in P'\}$, donde $P' \in \mathbb{N}$ representa la cantidad de rezagos considerados.

Por el contrario, $X' = \{\zeta_{t-p'} \sim \mathcal{N}(0, \vartheta_{t,p}^2) : t \in \mathbb{N}, p' \in P'\}$ denota un proceso de ruido blanco y $\theta' = \{\theta'_{p'} \in \mathbb{R} : p' \in P'\}$ reúne los componentes de media móvil.

Por su parte, el modelo ARIMA puede expresarse en notación matricial de la siguiente manera:

$$y = \hat{X} \bar{\theta} + X' \theta' + 1\eta = X\theta \quad [2]$$

donde $X = [\hat{X}, X', 1]$ y $X \in \mathbb{R}^{N \times P}$. El modelo de coeficientes es representado por $\theta = [\bar{\theta}, \theta', \eta]^T$ donde $\theta \in \mathbb{R}^P$.

Por su parte, cuando el enfoque anteriormente mencionado se adapta para incluir patrones estacionales, el modelo ARIMA se convierte en un Modelo Autorregresivo Integrado de Media Móvil Estacional (SARIMA). Este enfoque integral abarca tanto las correlaciones estacionarias como los cambios estacionales presentes en los datos. La forma general, asumiendo que los componentes estacionales presentan el mismo rezago que los componentes no estacionales, se expresa de la siguiente manera:

$$y_t = \eta + \sum_{p'=1}^{p'} \bar{\theta}_{p'} y_{t-p'} + \sum_{p'=1}^{p'} \theta'_{p'} \zeta_{t-p'} + \sum_{s=1}^{p'} \hat{\theta}_s y_{t-s} + \sum_{s=1}^{R'} \hat{\theta}_s \zeta_{t-s} \quad [3]$$

A continuación, la representación basada en matrices es la siguiente:

$$y = \hat{X} \bar{\theta} + X' \theta' + \check{X} \check{\theta} + \hat{X} \hat{\theta} + 1\eta \quad [4]$$

donde $\check{X} = \{y_{t-s} \in \mathbb{R} : t \in \mathbb{N}, s \in P'\}$ y $\check{\theta} = \{\check{\theta}_s \in \mathbb{R} : s \in P'\}$ representan el componente autorregresivo estacional, con $P' \in \mathbb{N}$ indicando el número de rezagos; del mismo modo, $\hat{X} = \{z_{t-s} \in \mathbb{R} : t \in \mathbb{N}, s \in R'\}$ y $\hat{\theta} = \{\hat{\theta}_s \in \mathbb{R} : s \in R'\}$ representan el proceso de ruido blanco de media móvil estacional y sus coeficientes correspondientes. En ambos casos, $s \in \mathbb{N}$ denota la cantidad de períodos utilizados para representar la estacionalidad.

Asimismo, el modelo SARIMA puede representarse en su versión concisa de la siguiente manera: $y = X\theta$, con $X = [\bar{X}, X', \check{X}, \hat{X}, 1]$, donde $X \in \mathbb{R}^{N \times P}$ simultáneamente, el vector de parámetros arroja: $\theta = [\bar{\theta}, \theta', \check{\theta}, \hat{\theta}, \eta]^T$, donde $\theta \in \mathbb{R}^P$.

Cabe destacar que los modelos ARIMA y SARIMA suelen optimizarse mediante el método de Máxima Verosimilitud (MLE, por sus siglas en inglés), en el que el proceso de estimación de parámetros se lleva a cabo, en muchos casos, a través del algoritmo L-BFGS-B (*Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Box constraints*). Este es un método cuasi-Newton especialmente adecuado para problemas de alta dimensión con restricciones acotadas (Hyndman y Athanasopoulos, 2018).

2.4 Modelo de Mínimos Cuadrados Ordinarios (OLS)

Esta metodología presupone una correlación lineal entre las variables exógenas $X = \{x_n \in \mathbb{R}^P : n \in N\}$, compuesta por N muestras y P características o variables, y una variable objetivo $y = \{y_n \in \mathbb{R}^P : n \in N\}$. A continuación, se establece una relación lineal de la siguiente forma: $y = X\theta$, donde el vector de coeficientes $\theta = \{\theta_p \in \mathbb{R}_p : p \in P\}$ representa los parámetros que caracterizan la influencia lineal de cada característica sobre la variable objetivo. Un método comúnmente utilizado para estimar los coeficientes de Mínimos Cuadrados Ordinarios (OLS) es el uso de la pseudoinversa de Moore–Penrose, lo cual conduce a la solución en forma cerrada:

$$\theta = (X^T X)^{-1} X^T y \quad [5]$$

Una extensión regularizada del método de Mínimos Cuadrados Ordinarios (OLS) puede derivarse resolviendo el siguiente problema de optimización:

$$\theta^* = \arg \min_{\theta} \|y - X\theta\|_2^2 + \lambda_1 \|\theta\|_1 + \lambda_2 \|\theta\|_2^2 \quad [6]$$

Cabe destacar que cuando $\lambda_1 \in \mathbb{R}^+$ y $\lambda_2 = 0$, OLS se aplica la regularización L1, conocida como *Least Absolute Shrinkage and Selection Operator* (LASSO). Del mismo modo, cuando $\lambda_1, \lambda_2 \in \mathbb{R}^+$ la combinación de las regularizaciones L1 y L2 –comúnmente conocida como regresión *Elastic Net*– surge como la formulación resultante.

2.5 Random Forest (RF)

Sea $y \in \mathbb{R}^N$ el vector objetivo y $X \in \mathbb{R}^{N \times P}$ la matriz de características exógenas de entrada. Una predicción no lineal basada en bosques aleatorios (RF) se define como:

$$y_n = \frac{1}{J} \sum_{j=1}^J \Phi^j(X, \Theta^{(j)}) \theta^{(j)} \quad [7]$$

donde el árbol j -ésimo, $\Phi^{(j)}(X, \Theta^{(j)}) = \{\varphi_{n,l}^{(j)}(x_n, \Theta^{(j)}) \in \{0,1\} : n \in \mathcal{N}, l \in L_j\}$ codifica si la observación x_n está asociada con un nodo del árbol. Por el contrario, los parámetros vinculados a cada nodo se denotan como $\theta^{(j)} = \{\theta_l^{(j)} \in \mathbb{R} : l \in L_j\}$, representando los coeficientes asociados a cada nodo en los respectivos árboles:

Entonces, la expresión: $y = \frac{1}{J} \Phi(X, \Theta) \theta$, donde el conjunto $\Theta = \{\Theta^{(j)} \in \mathbb{R}^{(L_j-1) \times P} : j \in J\}$ delimita las particiones del espacio de entradas generadas por cada árbol. La matriz consolida las funciones indicadoras de todos los árboles, es decir: $\Phi(X, \Theta) = \{\Phi^{(j)}(X, \Theta^{(j)})\}_{j=1}^J$ y $\theta = \{\theta^{(j)}\}_{j=1}^J$ representa la agregación de los coeficientes correspondientes.

El modelo de Bosques Aleatorios (RF, por sus siglas en inglés) determina los criterios óptimos de división y los umbrales de decisión utilizando el algoritmo de Árboles de Clasificación y Regresión (CART), el cual particiona iterativamente los datos con el objetivo de minimizar el siguiente error de clasificación o regresión.

$$\theta^*, \Theta^* = \arg \min_{\theta, \Theta} y - \frac{1}{J} \Phi(X, \Theta) \theta^2 \quad [8]$$

2.6 Support Vector Regression (SVR)

Dada una muestra de entrada $X_n \in \mathbb{R}^P$, la función de predicción no lineal del modelo SVR se formula de la siguiente manera:

$$f(x_n) = \varphi(x_n) \omega \quad [9]$$

donde $\omega = \{\omega_q \in \mathbb{R} : q \in Q\}$ es un coeficiente de vectores y $\varphi: \mathbb{R}^P \rightarrow \mathbb{R}^Q$ denota la función que transforma los datos a un espacio de mayor dimensión, con $Q \rightarrow \infty$. Esta transformación conlleva una carga computacional considerable, la cual se mitiga mediante el truco del kernel, que permite calcular el producto interno en el espacio transformado sin realizar explícitamente dicha transformación. Así, para dos muestras, la función kernel se define de la siguiente manera:

$$k(x_n, x_n) = \{\phi(x_n), \phi(x_n)\} \mathcal{H} = \phi(x_n) \phi(x_n)^T \in \mathbb{R} \quad [10]$$

La matriz resultante se representa como $K_{xx} = \{k((x_n, x_n): n, n' \in N)\}$ representa una transformación no lineal del espacio de entrada a un espacio de características de mayor dimensión y ω es el vector de pesos aprendido durante el entrenamiento. Posteriormente, se emplea una formulación dual de optimización para derivar la función de predicción del modelo SVR, lo que permite una computación eficiente en espacios de características de alta dimensión:

$$y = K_{xx} (\alpha - \alpha') \quad [11]$$

En este contexto, $\alpha = \{\alpha_n \in \mathbb{R}: n \in N\}$ y $\alpha' = \{\alpha'_n \in \mathbb{R}: n \in N\}$ representan los coeficientes de soporte. El modelo SVR utiliza el enfoque de Optimización Secuencial Mínima (SMO, por sus siglas en inglés) para resolver un problema de programación cuadrática, el cual se formula de la siguiente manera:

$$\alpha^*, \alpha'^* = \arg \max_{\alpha, \alpha'} -\frac{1}{2} (\alpha - \alpha')^T K_{xx} (\alpha - \alpha') \quad [12]$$

$$-\epsilon 1(\alpha + \alpha') + y^T (\alpha + \alpha') \quad [13]$$

$$\text{s.t } 1^T (\alpha - \alpha') = 0, \quad [14]$$

$$0 \leq \alpha, \alpha' \leq C \quad [15]$$

donde $\epsilon, C \in \mathbb{R}^+$

En consecuencia, la formulación del problema de minimización asociado a cada modelo de regresión varía en función de la estrategia de modelado seleccionada –por ejemplo, lineal versus no lineal– y del tipo de regularización aplicada. Estos elementos influyen significativamente tanto en la complejidad computacional de la solución como en la capacidad del modelo para generalizar frente a datos no observados. Además, la relevancia de cada variable de entrada suele reflejarse en los parámetros del modelo, cuya interpretación depende del marco supervisado específico utilizado.

En el caso de los modelos lineales ARIMA y SARIMA, la relevancia de las variables se infiere a partir de los coeficientes autorregresivos estacionales y no estacionales, los cuales encapsulan la influencia de las observaciones pasadas sobre la dinámica temporal de la serie. Por su parte, la regresión por mínimos cuadrados ordinarios (OLS) permite una cuantificación explícita de la influencia que cada variable exógena ejerce sobre la variable dependiente, mediante los coeficientes estimados. Este

enfoque contrasta con los modelos autorregresivos, que se centran en capturar las dependencias temporales internas de la serie.

En el caso de los Bosques Aleatorios (RF), la interpretabilidad se logra mediante el análisis de la importancia de las variables, la cual suele medirse a través de la reducción de la impureza de los nodos o del cambio en la precisión predictiva que resulta de la inclusión o exclusión de cada variable (Li *et al.*, 2026). Por el contrario, el modelo de *Support Vector Regression* (SVR) no ofrece inherentemente un mecanismo directo para evaluar la importancia de las variables, debido a su dependencia de la optimización basada en márgenes y funciones kernel (Alizamir *et al.*, 2024). Como resultado, se requieren técnicas auxiliares de interpretabilidad –como la descomposición mediante valores de Shapley o el análisis de sensibilidad– para estimar la contribución individual de cada variable (Molnar, 2020).

2.7 Supervised Kernel-Based Relevance Analysis using Gaussian Process

Considerando una entrada que involucra variables exógenas $X \in \mathbb{R}^{N \times p}$ y una variable objetivo $y \in \mathbb{R}^N$, la regresión mediante Procesos Gaussianos (GP) establece un priori sobre el vector de coeficientes $w = \{w_q \sim N(w_q | 0, \Sigma_q, q): q \in Q\}$, donde $\Sigma = \{\Sigma_q, q \in Q\}$ representa una matriz de covarianzas. La media y la varianza de la función de predicción $f(x)$ se calculan como:

$$m(x_n) = \mathbb{E}_w[f(x_n)] = \varphi(x_n) \mathbb{E}_w[W] = 0 \quad [16]$$

Y,

$$k(x_n, x_n') = \text{cov}(f(x_n), f(x_n')) = \mathbb{E}_w [f(x_n) f(x_n')^T] = \phi(x_n) \Sigma \phi(x_n')^T \quad [17]$$

Como resultado,

$$p(f(x)) = N(f(X) | 0, K_{XX}) \quad [18]$$

donde la notación $\mathcal{N}(\cdot | \cdot, \cdot)$ representa una función de probabilidad basada en una distribución gaussiana. En la regresión basada en Procesos Gaussianos (GP), las salidas observadas se modelan como realizaciones ruidosas de una función latente subyacente. Suponiendo ruido gaussiano aditivo, el modelo se expresa como: $y_n = f(x_n) + \varepsilon_n$, donde $\varepsilon_n \sim \mathcal{N}(\varepsilon_n | 0, \sigma_\varepsilon^2)$ representa un ruido gaussiano independiente e idénticamente distribuido. En consecuencia, la verosimilitud marginal de las salidas observadas, y condicionada a la matriz de entrada X , sigue una distribución normal multivariada:

$$p(y) = \mathcal{N}(y \mid 0, K_{xx} + \sigma_\epsilon^2 \mathbf{I}) \quad [19]$$

donde K_{xx} es la matriz de covarianza definida por el kernel del Proceso Gaussiano (GP). Y $\sigma_\epsilon^2 \mathbf{I}$ representa la covarianza del ruido de observación.

Dado un nuevo punto de entrada x^* , la regresión basada en Procesos Gaussianos (GP) calcula la distribución de probabilidad predictiva condicionando la distribución conjunta gaussiana previa con base en los datos observados. Esto conduce a una expresión en forma cerrada:

$$p(y^* \mid x^*, X, y) = \mathcal{N}(y^* \mid \mu^*, \sigma_*^2) \quad [20]$$

con la media y varianza predictiva definidas como:

$$\mu^* = k_*^T (\mu K_{xx} + \sigma_\epsilon^2 \mathbf{I})^{-1} y, \quad [21]$$

$$\sigma_*^2 = k(x_*, x_*) - k_*^T (K_{xx} + \sigma_\epsilon^2 \mathbf{I})^{-1} k_*; \quad [22]$$

donde $k_* = \{k(x_*, x_n) : n \in N\}$

Por su parte, cabe señalar que la regresión GP típicamente determina los hiperparámetros del kernel maximizando la verosimilitud marginal (también conocida como *evidence*) de los datos observados. Este proceso permite equilibrar el ajuste del modelo y su complejidad de manera probabilística y fundamentada.

Para resolver este problema de optimización, se emplea el algoritmo L-BFGS-B, lo cual implica minimizar la siguiente función objetivo:

$$v^* = \arg \min_v \left\{ \frac{1}{2} \log |k_{xx}(v) + \sigma_\epsilon^2 \mathbf{I}| + \frac{1}{2} y^T (k_{xx}(v) + \sigma_\epsilon^2 \mathbf{I})^{-1} y \right\} \quad [23]$$

donde la notación $k_{xx}(v)$ resalta la dependencia del kernel respecto al conjunto de hiperparámetros (v). Con el fin de habilitar una regresión no lineal y supervisada dentro de un marco probabilístico que aprovecha las transformaciones basadas en kernels y la regularización de pesos informada por prioris, se fortalece el análisis de relevancia de las características de entrada para mejorar la interpretabilidad del modelo basado en Procesos Gaussianos (GP).

Específicamente, se emplea un kernel exponencial cuadrado –también conocido como función de base radial (RBF)– con *longitudes de escala (lengthscales)* distintas asignadas a cada dimensión de entrada:

$$k(x_n, x_n | \tilde{\sigma}_f^2, \{\tilde{\sigma}_p^2\}_{p=1}^p) = \tilde{\sigma}_f^2 \exp \left(\sum_{p=1}^p \frac{-(x_{np} - x_{n'p})^2}{2\tilde{\sigma}_p^2} \right) \quad [24]$$

donde $x_{np} \in x_n$ y $\tilde{\sigma}_f^2, \tilde{\sigma}_p^2 \in \mathbb{R}^+$. Luego, cada *longitud de escala* puede ser optimizada según la Ecuación (24) para cuantificar la relevancia de la característica p-ésima en relación con las predicciones del modelo de Proceso Gaussiano (GP). En este contexto, el valor de relevancia supervisada de la característica p se define como $\lambda_p \in \mathbb{R}^+$ definido como el recíproco de $\tilde{\sigma}_p$ dando como resultado:

$$\tilde{\lambda}_p = \frac{\tilde{\sigma}_p^{-1}}{\max(\tilde{\sigma}_p^{-1})} \quad [25]$$

donde un valor reducido de $\tilde{\sigma}_p$ indica una mayor capacidad del modelo para detectar fluctuaciones locales en la característica asociada. Por el contrario, un valor más elevado indica un efecto más difuso, lo que significa una influencia disminuida en la predicción.

3. RESULTADOS Y DISCUSIÓN

La estructura de generación eléctrica en Colombia evidencia una distribución altamente diferenciada entre los departamentos, como se aprecia en la tabla 1, resultado de la variabilidad regional en recursos naturales, la disponibilidad de infraestructura y las decisiones históricas de inversión. Antioquia se consolida como el principal núcleo generador, con 6.958,8 MW que representan el 29,3 % de la capacidad efectiva neta nacional, sustentada mayoritariamente en centrales hidroeléctricas, complementadas por plantas térmicas, sistemas solares y cogeneración. Le siguen Cundinamarca (10,65 %) y Atlántico (10,00 %), cuyas matrices combinan generación térmica y solar, con una participación creciente de fuentes renovables no convencionales.

Tabla 1. Comparativo departamental de capacidad efectiva neta y tecnologías de generación en Colombia, 2025

Departamento	Tipo de generación	Generadores	Capacidad efectiva neta [MW]	Participación (%)
ANTIOQUIA	térmica, hidráulica, solar, cogenerador	81	6.958,7894	29,30
ARAUCA	térmica	5	29,6	0,12
ATLÁNTICO	térmica, solar, eólica	38	2.374,152	10,00
BOGOTÁ. D.C.	térmica, hidráulica	4	25,5	0,11

Departamento	Tipo de generación	Generadores	Capacidad efectiva neta [MW]	Participación (%)
BOLÍVAR	térmica, solar, cogenerador	17	950,235	4,00
BOYACÁ	térmica, hidráulica	9	1.388,6	5,85
CALDAS	hidráulica, solar, térmica	16	845,78	3,56
CASANARE	térmica, solar	15	354,59	1,49
CAUCA	térmica, solar, hidráulica, cogenerador	23	509,55	2,15
CESAR	térmica, solar	20	339,804	1,43
CÓRDOBA	térmica, solar, hidráulica	24	1.195,33	5,03
CUNDINAMARCA	térmica, solar, hidráulica	32	2.530,665	10,65
HUILA	solar, hidráulica	8	956,72	4,03
LA GUAJIRA	térmica, solar, eólica	10	353,456	1,49
MAGDALENA	térmica, solar	7	210,13	0,88
META	térmica, solar, cogenerador	19	225,26	0,95
NARIÑO	hidráulica, solar	10	30,07	0,13
NORTE DE SANTANDER	térmica, solar	10	419,39	1,77
PUTUMAYO	térmica, hidráulica	4	7,468	0,03
QUINDÍO	hidráulica	4	4,28	0,02
RISARALDA	solar, hidráulica, cogene- rador	6	53,29	0,22
SANTANDER	térmica, solar, hidráulica	23	1.962,24675	8,26
SUCRE	térmica, solar	9	75,7879	0,32
TOLIMA	térmica, solar, hidráulica	40	530,164	2,23
VALLE DEL CAUCA	térmica, solar, hidráulica, cogenerador	58	1.421,5258	5,98
Total		492	23.752,38385	100

Fuente: Elaboración propia con base en XM, (2026)

En este contexto, el departamento del Huila ocupa un lugar estratégico dentro del sistema interconectado colombiano. Con una capacidad efectiva neta de 956,72 MW, equivalente al 4,03 % del total nacional, presenta una matriz de generación basada casi exclusivamente en recursos renovables: 99 % hidráulica y 1 % solar. A diferencia de territorios con fuerte dependencia térmica –como Atlántico, Córdoba o Santander–, la oferta energética huilense se sustenta en una fuente limpia, de bajo impacto en emisiones y con menores riesgos de exposición a la volatilidad de los precios internacionales del gas y el carbón.

Esta composición genera ventajas económicas y ambientales, pero también implica vulnerabilidades operativas. La alta dependencia hidroeléctrica convierte al departamento en un territorio especialmente sensible a fenómenos climáticos como El Niño, que afectan los niveles de embalse y la estabilidad del suministro. La incorporación reciente de proyectos fotovoltaicos, aunque aún incipiente, demuestra que el Huila cuenta con condiciones favorables para diversificar su matriz y fortalecer su resiliencia energética mediante un mayor aprovechamiento del potencial solar. En comparación con otros departamentos de tamaño similar, como Boyacá (5,85 %) o Córdoba (5,03 %), el Huila exhibe un modelo de generación más sostenible, pero con un margen significativo para ampliar su capacidad instalada mediante proyectos solares de mayor escala. Esto adquiere especial relevancia ante la evidencia presentada en los modelos de predicción energética: la demanda en el Huila muestra una tendencia creciente, estructural y predecible, con picos cada vez más altos y un incremento sostenido en el consumo total diario.

La combinación de crecimiento progresivo de la demanda –confirmado por modelos como Random Forest y Gaussian Process Regression, que se presentarán a continuación– con una matriz fuertemente renovable implica un desafío doble para la planificación energética. No basta con contar con fuentes limpias, es necesario anticipar con precisión los niveles de carga futuros para garantizar que la infraestructura de generación y distribución sea capaz de abastecer el aumento sostenido del consumo.

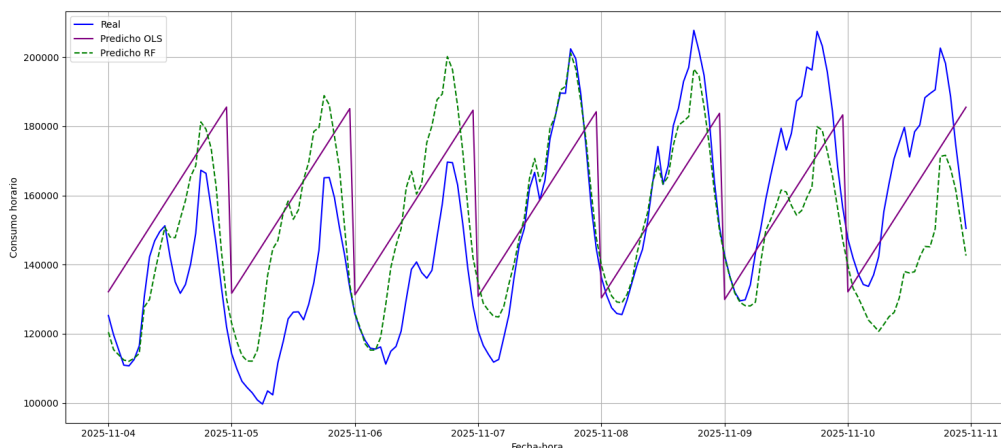
En consecuencia, la capacidad del Huila para consolidarse como referente en transición energética depende directamente de su capacidad para predecir su demanda con rigor técnico, ajustar la expansión de su infraestructura renovable y diseñar esquemas de operación flexibles que mitiguen la variabilidad climática. Predecir la demanda no es un ejercicio técnico aislado, sino un insumo esencial para orientar inversiones, planear ampliaciones en la red, definir incentivos regulatorios y asegurar la seguridad de suministro frente al aumento progresivo del consumo eléctrico en la región.

3.1 Resultados de la predicción

El comportamiento horario de la demanda energética en el departamento del Huila revela una estructura estable, cíclica y con señales claras de crecimiento sostenido. El análisis de la serie horaria muestra que el sistema presenta picos recurrentes durante las horas de mayor actividad económica y residencial, con valores máximos concentrados entre las 18:00 y 21:00 horas, como se presenta en la figura 2. Este comportamiento evidencia la coexistencia de dos fuentes primarias de consumo: la actividad productiva diurna y el uso residencial nocturno. En contraste, las horas valle se encuentran entre la medianoche y las cinco de la mañana, con niveles mínimos de demanda que se mantienen dentro de rangos estables, lo cual confirma la existencia de un consumo base asociado a cargas permanentes del sistema.

La evolución diaria del consumo energético en el departamento del Huila muestra una tendencia ascendente en la envolvente superior de la demanda, es decir, un aumento progresivo en los picos máximos del sistema a lo largo del tiempo. Este comportamiento sugiere una expansión paulatina del consumo y evidencia que el sistema aún presenta margen operativo para absorber incrementos adicionales de demanda. En términos económicos, esta dinámica puede estar asociada al crecimiento de la actividad agroindustrial, la ampliación de la electrificación rural, la expansión urbana en zonas periurbanas y el mayor uso de bienes y servicios intensivos en energía eléctrica.

Figura 2. Consumo horario real vs. OLS y vs. RF (última semana)



Fuente: elaboración propia de los autores con base en los resultados, (2026)

Los resultados de predicción muestran diferencias claras entre los modelos evaluados. Los enfoques estadísticos tradicionales, en particular ARIMA y SARIMAX, presentan un desempeño más limitado, con errores más altos y baja capacidad explicativa, lo que indica dificultades para capturar adecuadamente la complejidad temporal del consumo energético. En contraste, los modelos de aprendizaje automático alcanzan mejores niveles de ajuste, y se destaca Random Forest como la técnica con mejor desempeño predictivo, al explicar cerca del 67 % de la variabilidad observada y reducir de forma importante los errores frente a los modelos convencionales.

Este resultado sugiere que la demanda energética del Huila responde más a relaciones no lineales entre variables temporales, como la hora del día, el día de la semana y la estacionalidad, que a estructuras autorregresivas simples. En consecuencia, la modelación horaria mediante enfoques flexibles ofrece una mejor aproximación al comportamiento real del consumo, especialmente en contextos en los que confluyen patrones heterogéneos asociados a actividades agroindustriales, comerciales, residenciales y de servicios.

En términos aplicados, la capacidad de predecir la demanda horaria con mayor precisión representa una oportunidad para fortalecer la planificación energética departamental. Entre las posibles aplicaciones se encuentran la gestión inteligente de la demanda, el diseño de esquemas tarifarios diferenciados, la incorporación de generación distribuida en horarios estratégicos y la anticipación de requerimientos de infraestructura en zonas con crecimiento sostenido del consumo.

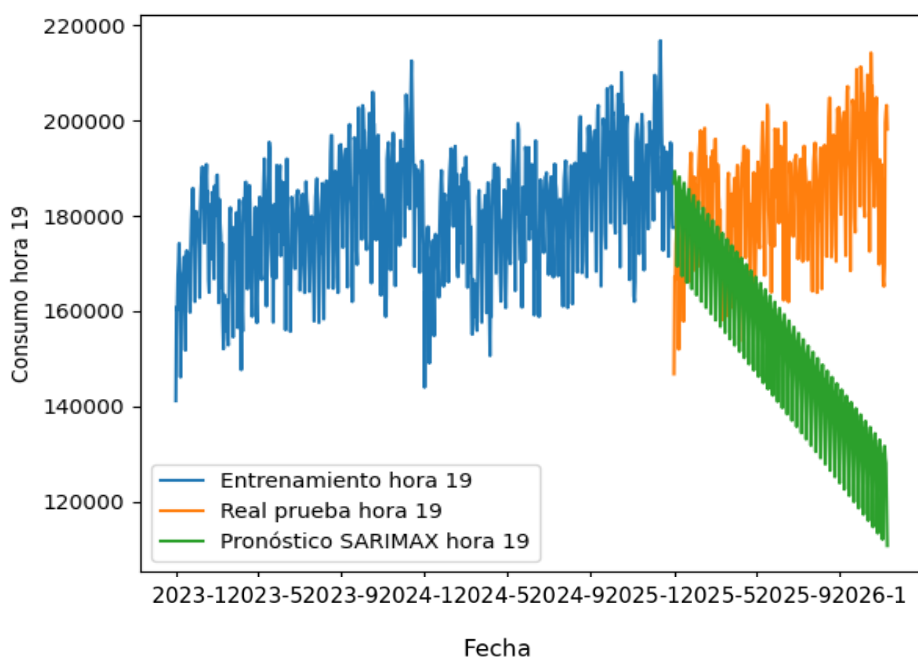
De este modo, la evidencia observada permite concluir que el Huila presenta un patrón de crecimiento energético estable y predecible, con condiciones favorables para avanzar hacia una gestión más eficiente del sistema. No obstante, las implicaciones en materia de infraestructura, transición energética o políticas sectoriales deben entenderse como posibles aplicaciones derivadas de los resultados del modelo, más que como efectos demostrados directamente por el análisis realizado.

El comportamiento histórico del consumo energético en la hora 19 presenta un patrón cíclico bien definido, con estacionalidad diaria y semanal, así como una tendencia creciente durante el período 2023-2025. En la etapa de entrenamiento, la serie muestra oscilaciones relativamente estables, aunque con variaciones asociadas a la dinámica entre días hábiles, fines de semana y momentos de mayor actividad económica. En el período de prueba se conserva esta estructura estacional, pero se observa un aumento en la envolvente superior, lo que sugiere una continuidad en el crecimiento de la demanda para esa franja horaria. Además, el pronóstico generado

por SARIMAX no logra reproducir adecuadamente este comportamiento. Aunque incorpora una estructura estacional semanal, el modelo presenta una trayectoria descendente que se aparta visiblemente de la serie observada, subestimando de manera sistemática la demanda real. Esto indica que la especificación utilizada no logra capturar ni la tendencia creciente ni la complejidad temporal del consumo en esta hora.

Así, este resultado es consistente con las métricas obtenidas, que evidencian un bajo desempeño predictivo del modelo, reflejado en errores elevados y un coeficiente R^2 negativo. Por su parte, la figura 3 muestra que, para la hora 19, el modelo SARIMAX resulta insuficiente para representar la dinámica del consumo energético, especialmente en un contexto de crecimiento sostenido y variaciones no lineales en la demanda.

Figura 3. Consumo hora 19: real vs. Sarimax

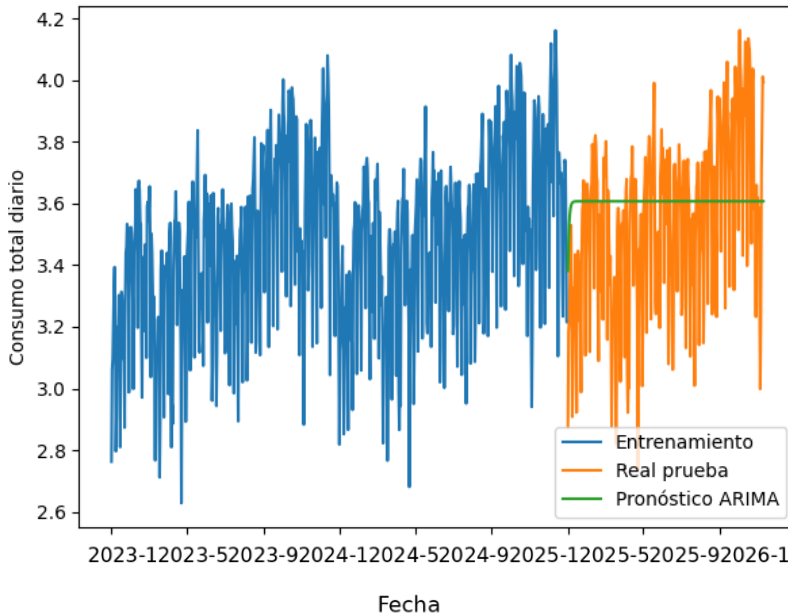


Fuente: elaboración propia de los autores con base en los resultados, (2026)

La figura 4 compara el consumo total diario observado con las predicciones generadas por el modelo ARIMA. La serie real, tanto en el período de entrenamiento como en el de prueba, presenta oscilaciones diarias marcadas y una tendencia creciente en los niveles máximos y mínimos de consumo, lo que sugiere una

expansión progresiva de la demanda energética en el departamento del Huila. En contraste, el modelo ARIMA produce una predicción prácticamente constante a partir del inicio del período de prueba, convergiendo hacia un valor cercano al promedio histórico. Este comportamiento indica que el modelo no logra captar ni la tendencia creciente ni la variabilidad diaria observada en la serie.

Figura 4. Consumo total diario: real vs. ARIMA

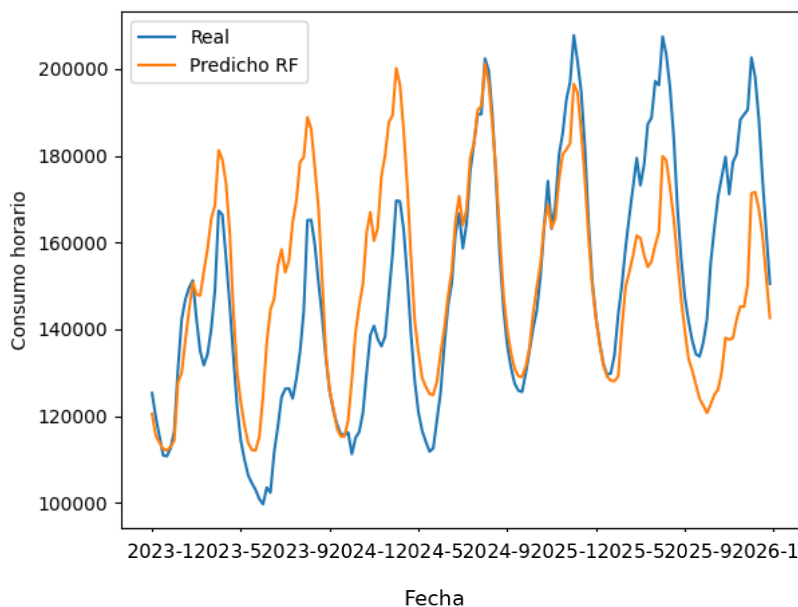


Fuente: elaboración propia de los autores con base en los resultados, (2026)

Aunado a lo anterior, este resultado es consistente con las métricas obtenidas, que muestran errores elevados y un coeficiente R^2 negativo, lo que evidencia un bajo desempeño predictivo. Así, la figura 4 sugiere que el modelo ARIMA utilizado resulta insuficiente para representar adecuadamente la dinámica del consumo total diario en un contexto de crecimiento sostenido de la demanda.

La figura 5 compara la serie horaria real de consumo energético con la estimación generada por el modelo Random Forest durante la última semana del conjunto de prueba. El resultado muestra que el modelo reproduce con buena precisión la forma general del perfil de carga, capturando la periodicidad diaria y la ubicación de los principales picos y valles de consumo.

Figura 5. Consumo horario: real vs. RF (última semana)



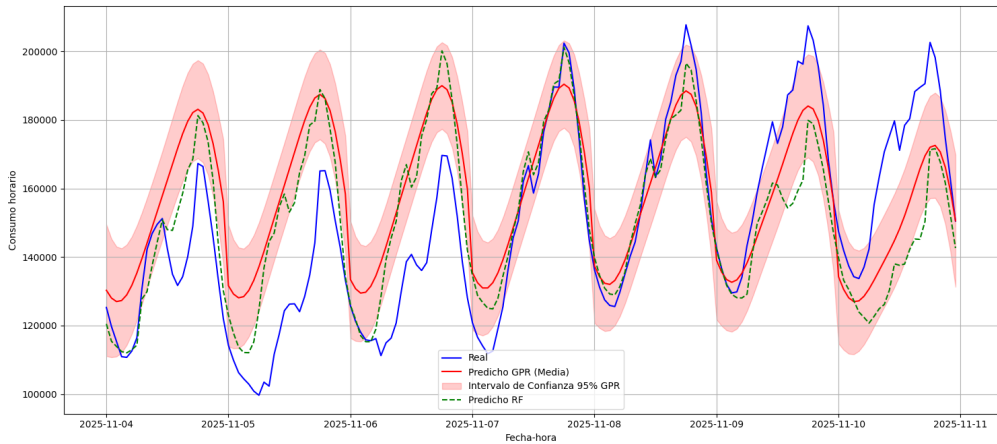
Fuente: elaboración propia de los autores con base en los resultados, (2026)

En términos generales, la demanda presenta un ciclo horario bien definido, con incrementos desde la mañana, máximos entre las 18:00 y las 21:00 y descensos durante la madrugada. El modelo logra seguir este patrón de manera consistente, aunque presenta pequeñas desviaciones en algunos picos máximos, en los que tiende a subestimar o sobreestimar ligeramente los valores observados.

Aun con estas diferencias puntuales, el desempeño del modelo es estable a lo largo del período evaluado y consistente con las métricas obtenidas, lo que confirma su capacidad para capturar la estructura no lineal del consumo energético horario. Además, la figura 5 evidencia que Random Forest ofrece una aproximación sólida para la predicción de la demanda energética en el departamento del Huila.

La figura 6 presenta la comparación entre el consumo energético horario real y las predicciones generadas por dos modelos de aprendizaje automático: Gaussian Process Regression (GPR) y Random Forest (RF). En el caso de GPR, además, se incorpora un intervalo de confianza del 95 %, lo que permite no solo evaluar la predicción puntual, sino también analizar el nivel de incertidumbre asociado a la estimación. Esta representación resulta útil para examinar la capacidad de ambos modelos en un contexto de demanda energética con alta variabilidad temporal.

Figura 6. Consumo horario: real vs. RF y vs. GPR



Fuente: elaboración propia de los autores con base en los resultados, (2026)

En términos generales, ambos modelos logran reproducir adecuadamente la forma del perfil de carga horario, capturando la estructura cíclica del consumo y sus principales variaciones a lo largo del día. Tanto GPR como RF representan de manera razonable los picos de mayor demanda y los períodos de menor consumo, lo que confirma la existencia de patrones recurrentes y no lineales en la serie energética del departamento del Huila.

El modelo Random Forest presenta una mayor cercanía a los valores observados, y mantiene una buena sincronía con la serie real y un ajuste más preciso en la mayoría de las franjas horarias. Aunque se observan pequeñas desviaciones en algunos picos de consumo, su comportamiento general es estable y consistente, lo que respalda su buen desempeño en términos de predicción puntual.

Por su parte, Gaussian Process Regression ofrece una estimación más suavizada de la serie. Su principal fortaleza radica en la posibilidad de cuantificar la incertidumbre mediante la banda de confianza, la cual se amplía en los momentos de mayor variabilidad y se reduce cuando el comportamiento del consumo es más predecible. El hecho de que gran parte de los valores observados se mantenga dentro del intervalo del 95 % indica que el modelo presenta una calibración probabilística adecuada.

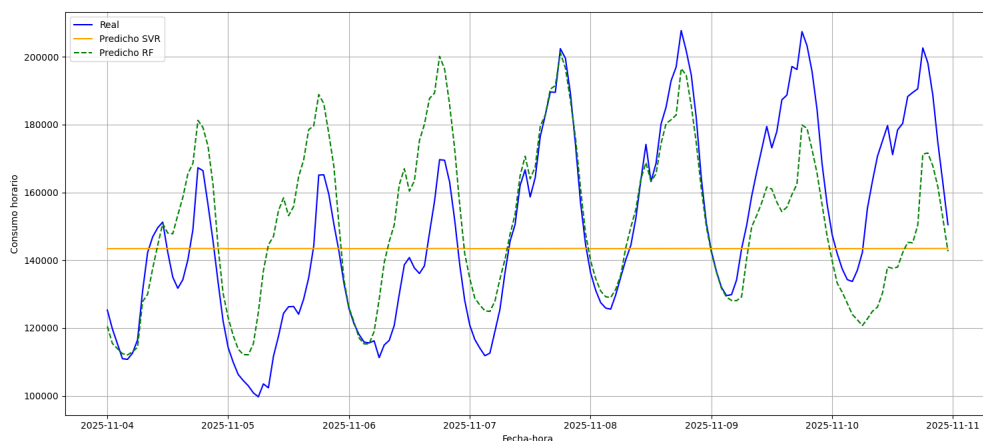
No obstante, al comparar ambas técnicas, se observa que GPR tiende a suavizar en exceso la señal: subestima algunos máximos y sobrestima ciertos mínimos. Esto explica que, aunque ofrece una representación probabilística valiosa, su precisión puntual sea inferior a la obtenida por Random Forest. En este sentido, mientras

RF resulta más conveniente para aplicaciones que requieren mayor exactitud en el pronóstico, GPR aporta ventajas en escenarios en los cuales también interesa evaluar la incertidumbre de la predicción.

De hecho, la comparación refuerza la idea de que la demanda energética horaria en el Huila responde a patrones no lineales que pueden ser modelados de manera más efectiva mediante técnicas de aprendizaje automático. Entre los modelos evaluados, Random Forest destaca por su mayor precisión predictiva, mientras que Gaussian Process Regression aporta una perspectiva complementaria al incorporar explícitamente la incertidumbre asociada al pronóstico.

La figura 7 compara el consumo energético horario real con las predicciones generadas por dos modelos de aprendizaje automático: Support Vector Regression (SVR) y Random Forest. La serie observada presenta el comportamiento típico de una curva de carga eléctrica, con oscilaciones diarias marcadas, picos de consumo en las horas de la noche y mínimos recurrentes durante la madrugada. Este patrón se mantiene de forma relativamente estable a lo largo del período analizado, lo que confirma la existencia de una estructura temporal definida en la demanda energética del departamento del Huila.

Figura 7. Consumo horario: real vs. SVR y vs. RF (última semana)



Fuente: elaboración propia de los autores con base en los resultados, (2026)

El modelo SVR muestra un ajuste limitado frente a esta dinámica. En la figura, su predicción se mantiene cercana a un valor casi constante, lo que indica una baja capacidad para reproducir la variabilidad real de la serie y para captar los cambios cíclicos propios del consumo horario. Este comportamiento es consistente con

sus resultados cuantitativos, caracterizados por errores elevados y un bajo poder explicativo, lo que sugiere que la configuración utilizada no fue suficiente para representar adecuadamente la complejidad no lineal de la demanda.

En contraste, el modelo Random Forest logra aproximarse de forma mucho más precisa al comportamiento observado. Su predicción conserva la forma general de la curva de carga, reproduce los ciclos diarios y mantiene coherencia temporal con los valores reales, aunque con ligeras diferencias en algunos picos máximos. Este desempeño concuerda con las métricas obtenidas, que muestran una reducción importante del error y una mayor capacidad explicativa en comparación con SVR y con los modelos estadísticos evaluados previamente.

La comparación entre ambos modelos evidencia que la demanda energética horaria en el Huila responde a patrones no lineales y a estructuras temporales complejas, que son capturadas con mayor eficacia por Random Forest. En este sentido, la figura confirma que, para el problema abordado, este modelo ofrece una aproximación más robusta y adecuada para ejercicios de predicción horaria con potencial aplicación en la planificación y gestión energética regional.

Los resultados presentados en la tabla 2 evidencian diferencias significativas en la capacidad predictiva de los modelos probados. Los modelos estadísticos clásicos (ARIMA y SARIMAX) presentan el peor desempeño relativo. ARIMA, aplicado a la serie diaria agregada, muestra un coeficiente de determinación negativo y un RMSE superior a 280.000 unidades, lo que indica que su rendimiento es equivalente –o inferior– a un pronóstico constante basado en la media histórica. SARIMAX, aun incorporando estacionalidad semanal exógena, evidencia una divergencia aún mayor, con un R^2 de -10,15 y un error escalado (MASE) de 3,97, demostrando que su estructura autorregresiva no logra capturar la dinámica real del consumo energético horario.

Tabla 2. Resumen métricas de desempeño de los modelos

Modelo	MAE	RMSE	MAPE (%)	R^2	MASE
ARIMA (diario)	218.374,79	280.491,79	6,39	-0,02	1,18
SARIMAX (hora 19)	34.131,97	41.046,84	17,97	-10,15	3,97
Random Forest (horario)	9.910,59	14.137,15	6,57	0,67	1,49
Gaussian Process (horario)	11.244,38	14.753,16	7,59	0,64	1,69
Support Vector Regressor (horario)	21.226,11	25.614,97	13,81	-0,09	3,19
Linear Regression (horario)	14.490,59	17.860,73	9,79	0,47	2,18

Fuente: elaboración propia de los autores con base en los resultados, (2026)

En contraste, los modelos de aprendizaje automático exhiben comportamientos sustancialmente más precisos. El modelo Random Forest obtiene los mejores resultados en casi todas las métricas (MAE, RMSE y R^2), con un desempeño que indica que logra explicar el 67 % de la variabilidad observada y reduce los errores absolutos en más de un orden de magnitud respecto al ARIMA diario. Este resultado confirma que la demanda energética presenta patrones no lineales y dependencias complejas que son mejor capturadas por métodos basados en particionamiento del espacio de predicción.

El modelo Gaussian Process Regression, aunque ligeramente menos preciso que Random Forest, demuestra ser consistente en términos probabilísticos. Su banda de confianza conserva la mayor parte de la observación real dentro del rango del 95 %, lo que lo hace apropiado para aplicaciones de gestión del riesgo energético, simulación de escenarios o predicciones probabilísticas (no puntuales). El Support Vector Regressor, por su parte, ofrece un desempeño deficiente, con una clara incapacidad para aproximar la forma de la serie, lo que se traduce en un comportamiento casi constante y un R^2 negativo.

Finalmente, la regresión lineal establece un punto intermedio: aunque logra capturar parte de la variabilidad estructural de la serie, su R^2 de 0,47 confirma que la relación entre calendarios y consumo es no lineal, lo que justifica el uso de modelos más flexibles.

4. CONCLUSIONES

El análisis desarrollado sobre las series de consumo energético en el departamento del Huila permitió evaluar de manera comparativa la capacidad predictiva de distintos modelos estadísticos y de aprendizaje automático aplicados a la demanda diaria y horaria. En conjunto, los resultados muestran que el comportamiento del consumo energético regional no puede ser representado de forma suficiente mediante estructuras lineales o autorregresivas simples, sino que responde a patrones temporales más complejos, con componentes no lineales, estacionalidades diferenciadas y señales de crecimiento estructural. Esta constatación constituye uno de los principales hallazgos del estudio, ya que permite establecer que la predicción de la demanda energética en el caso huilense requiere enfoques metodológicos con mayor flexibilidad analítica que los ofrecidos por los modelos estadísticos convencionales.

En este marco, los modelos ARIMA y SARIMAX presentaron un desempeño considerablemente inferior al de las técnicas de aprendizaje automático. El modelo

ARIMA, aplicado al consumo total diario, mostró errores elevados y un coeficiente de determinación negativo, lo que indica que sus predicciones no aportan capacidad explicativa adicional frente a estimadores muy básicos. Por su parte, SARIMAX, utilizado para la predicción de la hora 19, también exhibió una baja capacidad de ajuste, incluso incorporando estacionalidad semanal y regresores de calendario. En ambos casos, los resultados sugieren que estos enfoques no logran captar adecuadamente ni la tendencia creciente del consumo ni la complejidad intradiaria de la serie energética analizada. Más que evidenciar la inutilidad de estos modelos, los hallazgos muestran su insuficiencia como herramienta única de pronóstico en un contexto en el cual la demanda presenta variaciones estructurales y relaciones no lineales persistentes.

La regresión lineal multivariante mostró un comportamiento intermedio. Sus resultados fueron superiores a los de los modelos autorregresivos, lo que confirma que variables temporales como año, mes, día, día de la semana y hora contienen información relevante para explicar parte de la dinámica de la demanda. Sin embargo, el nivel de ajuste alcanzado también evidencia que la estructura del consumo energético no puede reducirse a una relación lineal simple entre los predictores y la variable objetivo. En este sentido, la regresión lineal conserva utilidad como punto de comparación e incluso como referencia interpretativa básica, pero resulta insuficiente para construir proyecciones robustas en contextos en los que la variabilidad del consumo depende de interacciones más complejas.

Entre los modelos evaluados, Random Forest se posicionó como el algoritmo de mejor desempeño global. Sus métricas de error fueron notablemente inferiores a las de los modelos estadísticos y su capacidad explicativa fue la más alta entre las técnicas probadas. Este desempeño indica que el modelo logra capturar de manera efectiva una proporción importante de la variabilidad horaria del consumo energético, y mantiene coherencia temporal, estabilidad en la predicción y una buena reproducción de la forma general de la curva de carga. La superioridad de Random Forest puede explicarse por su capacidad para representar relaciones no lineales, segmentar el espacio de predicción y adaptarse a estructuras complejas sin requerir una parametrización explícita de tendencia o estacionalidad. En consecuencia, el estudio muestra que este enfoque constituye una herramienta técnicamente sólida para ejercicios de pronóstico energético regional.

Gaussian Process Regression también presentó un desempeño favorable, aunque ligeramente inferior al de Random Forest en términos de precisión puntual. No obstante, su valor metodológico es relevante, ya que incorpora explícitamente la cuantificación de incertidumbre a través de intervalos de confianza sobre la

predicción. Esto amplía la utilidad del modelo más allá del pronóstico puntual, y permite explorar escenarios de riesgo, evaluar rangos probables de comportamiento de la demanda y fortalecer análisis en contextos en los cuales la incertidumbre constituye una dimensión importante de la toma de decisiones. Así, mientras Random Forest destaca por su mayor precisión, Gaussian Process Regression ofrece una perspectiva complementaria al introducir una representación probabilística de la demanda energética.

En contraste, Support Vector Regressor mostró el peor desempeño entre los modelos de aprendizaje automático considerados. Su predicción tendió a aproximarse a valores casi constantes, sin reproducir de manera adecuada la variabilidad observada en la serie. Este comportamiento confirma que, bajo la configuración utilizada en el estudio, el modelo no logra capturar la complejidad del fenómeno analizado. Más que descartar por completo este tipo de técnica, el resultado indica que su aplicación a problemas de demanda energética requiere un proceso de ajuste más cuidadoso, posiblemente acompañado por modelos de ingeniería más sofisticados, selección de *kernels* especializados o transformaciones adicionales de la serie.

La comparación conjunta de resultados numéricos y gráficos permite concluir que la demanda energética del Huila presenta una estructura suficientemente regular como para ser modelada, pero también suficientemente compleja como para exigir enfoques de predicción flexibles. La serie no exhibe un comportamiento caótico o puramente aleatorio; por el contrario, muestra estabilidad relativa, ciclos identificables y una tendencia creciente en la demanda agregada y en ciertos picos horarios. Esta combinación de regularidad y complejidad explica por qué los modelos de aprendizaje automático, particularmente Random Forest, logran un mejor desempeño que los modelos estadísticos tradicionales. En ese sentido, el estudio aporta evidencia empírica de que la predictibilidad de la demanda energética regional puede mejorarse de forma sustancial mediante metodologías de modelación más adaptativas.

Desde una perspectiva aplicada, los hallazgos sugieren que los modelos con mejor desempeño podrían constituir insumos técnicos útiles para ejercicios de planificación de carga, análisis prospectivo de demanda y apoyo a decisiones operativas en el sistema energético regional. Sin embargo, es importante precisar que este trabajo demuestra la mayor capacidad predictiva de ciertos modelos sobre la base analizada, pero no demuestra por sí mismo la conveniencia de una reorganización inmediata de la matriz energética departamental ni la adopción directa de tecnologías específicas. Tales implicaciones deben entenderse como

posibilidades plausibles derivadas del estudio, mas no como conclusiones demostradas de manera directa. Cualquier decisión relacionada con expansión de infraestructura, incorporación de fuentes renovables, almacenamiento energético o rediseño de esquemas tarifarios requeriría estudios adicionales de factibilidad técnica, económica, regulatoria y territorial.

En este punto, conviene reconocer explícitamente las principales limitaciones de la investigación. En primer lugar, la modelación se construyó principalmente a partir de variables temporales derivadas del calendario, sin incorporar variables exógenas que podrían influir de manera importante en la demanda, como condiciones climáticas, temperatura, actividad económica, precios de la energía, eventos extraordinarios o transformaciones sectoriales en el territorio. En segundo lugar, el análisis se concentró en un conjunto acotado de modelos estadísticos y de aprendizaje automático, sin incluir arquitecturas más recientes basadas en redes neuronales profundas, como ANN o LSTM, que han mostrado resultados promisorios en la literatura reciente de predicción energética. En tercer lugar, aunque los modelos evaluados permiten medir capacidad predictiva, el estudio no aborda de manera directa dimensiones como costos de error, impacto operativo real, viabilidad de implementación en entornos institucionales o articulación con esquemas integrales de planificación energética. Estas limitaciones no invalidan los resultados obtenidos, pero sí delimitan su alcance interpretativo.

Precisamente por ello, una línea futura de trabajo consiste en ampliar el conjunto de variables explicativas para incorporar factores climáticos, económicos, operativos y territoriales que permitan enriquecer la explicación de la demanda y no solo su predicción a partir de patrones temporales. De igual manera, sería pertinente evaluar arquitecturas más avanzadas de aprendizaje profundo, así como estrategias de validación temporal más detalladas y procesos de optimización de hiperparámetros más exhaustivos. También resultaría valioso desarrollar estudios complementarios que articulen los pronósticos obtenidos con escenarios de planeación energética, análisis de expansión de red, integración de generación distribuida y evaluación de políticas públicas para la transición energética regional. De esta forma, futuras investigaciones podrán conectar de manera más directa la evidencia predictiva con decisiones estratégicas de mediano y largo plazo.

En síntesis, la investigación demuestra que la demanda energética del Huila puede predecirse con mayor precisión mediante modelos de aprendizaje automático que mediante enfoques estadísticos clásicos, y que Random Forest constituye, entre los modelos evaluados, la alternativa de mejor desempeño para la predicción horaria. Este hallazgo representa una contribución empírica relevante para el

análisis de demanda energética regional, al mostrar que la estructura del consumo eléctrico huilense combina crecimiento, regularidad temporal y no linealidad. Al mismo tiempo, el estudio deja claro que la utilidad de estos resultados depende de reconocer sus límites y de avanzar hacia investigaciones que integren mayor riqueza explicativa, análisis de incertidumbre y vínculos más directos con la planificación energética territorial.

REFERENCIAS

- Ahmed, A., Jakir, T., Hossain, N., Fahim, A., Hossain, A., Hoque, A., & Hasan, S. (2025). Predicting energy consumption in hospitals using machine learning: A data-driven approach to energy efficiency in the USA. *Journal of Computer Science and Technology Studies*, 7,1. DOI: <https://doi.org/10.32996/jcsts.2025.7.1.15>
- Alizamir, M., Wang, M., Adnan-Ikram, R., Kim, S., Ahmed, K., & Heddam, S. (2024). Developing an efficient explainable artificial intelligence approach for accurate reverse osmosis desalination plant performance prediction: Application of SHAP analysis. *Engineering Applications of Computational Fluid Mechanics*, 18(1), 2422060. <https://doi.org/10.1080/19942060.2024.2422060>
- Al Kez, D., Foley, A., Abdul, Z. K., & Furszyfer Del Rio, D. (2024). Energy poverty prediction in the United Kingdom: A machine learning approach. *Energy Policy*, 184, 113909. <https://doi.org/10.1016/j.enpol.2023.113909>
- Alshater, M. M., Kampoouris, I., Marashdeh, H., Atayah, O. F., & Banna, H. (2025). Early warning system to predict energy prices: The role of artificial intelligence and machine learning. *Annals of Operations Research*, 345(2), 1297–1333. <https://doi.org/10.1007/s10479-022-04908-9>
- Ansong, M., Nyang'onda, T. N., Musembi, R. J., & Richards, B. S. (2024). Very short-term solar irradiance forecasting for photovoltaic power integration with the grid: Potentials and challenges for Africa. In 2024 IEEE PES/IAS PowerAfrica (pp. 1–5). IEEE. <https://doi.org/10.1109/PowerAfrica61624.2024.10759405>
- Bulungu, D. M., & Kumar, A. (2024). Forecasting the economic growth of Sverdlovsk region: A comparative analysis of machine learning, linear regression and autoregressive models. *Journal of Applied Economic Research*, 23(3), 674–695. <https://doi.org/10.15826/vestnik.2024.23.3.027>
- Comisión de Regulación de Energía y Gas. (2024). Informe al Congreso: Segundo semestre de 2024. CREG.
- Cui, X., Lee, M., Koo, C., & Hong, T. (2024). Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy-efficient buildings. *Energy and Buildings*, 309, 113997. <https://doi.org/10.1016/j.enbuild.2024.113997>
- Hasan, M. S., Tarequzzaman, M., Moznuzzaman, M., & Ahad Juel, M. A. (2025). Prediction of energy consumption in four sectors using support vector regression optimized with genetic algorithm. *Heliyon*, 11(2), e41765. <https://doi.org/10.1016/j.heliyon.2025.e41765>

- Hicks, J. R., & Allen, R. G. D. (1934). A reconsideration of the theory of value. Part I. *Economica*, 1(1), 52–76. <https://doi.org/10.2307/2548574>
- Hyndman, R. y Athanasopoulos, G. (2018). *Pronóstico: principios y práctica*. OTexts.
- Hotelling, H. (1931). The economics of exhaustible resources. *Journal of Political Economy*, 39(2), 137–175. <https://doi.org/10.1086/254195>
- Ibarra, N., Duque, J., Velandia, Y. y Salas, L. (2024). Energías renovables en la caficultura colombiana: Estrategias para la sostenibilidad y la eficiencia energética. Universidad EAN.
- Kaloop, M., Ahmad, F., Samui, P., Elbeltagi, E., Hu, J.-W., & Wefki, H. (2025). Predicting energy consumption of residential buildings using metaheuristic-optimized artificial neural network technique in early design stage. *Building and Environment*. 274, 112749 y DOI: <https://doi.org/10.1016/j.buildenv.2025.112749>
- Kulisz, M., Kujawska, J., Cioch, M., Cel, W., & Pizoń, J. (2024). Comparative analysis of machine learning methods for predicting energy recovery from waste. *Applied Sciences*, 14(7), 2997. <https://doi.org/10.3390/app14072997>
- Li, Y., Zhang, W., Zhao, B., Sharp, B., & Nie, J. (2026). Does energy poverty affect subjective well-being? Evidence from a cross-country analysis. *Applied Economics*, 58(1), 141–156. <https://doi.org/10.1080/00036846.2024.2449208>
- López, F., Torre, J., Ríos, L., & Ruvulvaba, L. (2024). Enhancing electricity demand prediction in Mexico: A comparative analysis of forecasting models using conformal prediction. *Revista de Gestão Social e Ambiental*, 18(12), e010644. <https://doi.org/10.24857/rgsa.v18n12-235>
- Marshall, A. (1890). *Principios de economía*. Macmillan and Company.
- Mohammed, D., Ebrahim Ali, D. M. T., Motuzienė, V., & Džiugaitė-Tuménienė, R. (2024). AI-driven innovations in building energy management systems: A review of potential applications and energy savings. *Energies*, 17(17), 4277. <https://doi.org/10.3390/en17174277>
- Molnar, C. (2020). *Interpretable machine learning*. Lulu.com.
- Olivares, F. (2025). Análisis de técnicas en ciencia de datos aplicadas a la matriz energética renovable en Colombia. Repositorio UNAD. <https://repository.unad.edu.co/handle/10596/70349>
- Osorio, W. (2025). Uso del *machine learning* en la predicción de la demanda de energía en Colombia. Repositorio UNAD. <https://repository.unad.edu.co/handle/10596/70599>
- Pérez-Rosero, D. A., Manrique-Cabezas, D. A., Triana-Martínez, J. C., Álvarez-Meza, A. M., & Castellanos-Domínguez, G. (2025). An explainable framework integrating local biplots and Gaussian processes for unemployment rate prediction in Colombia. *Computation*, 13(5), 116. <https://doi.org/10.3390/computation13050116>
- Pineda, C., Arrieta, I. y Quitián, M. (2020). Generando oportunidades en la calidad de vida de los habitantes de Villavieja en el departamento del Huila a través de energías renovables. Repositorio UNAD. <https://repository.unad.edu.co/handle/10596/35251>
- Rao, A., Kumar, S., & Karim, S. (2024). Accelerating renewables: Unveiling the role of green energy markets. *Applied Energy*, 366, 123286. <https://doi.org/10.1016/j.apenergy.2024.123286>

- Reyes, R., Turriago, A., Cárdenas, M. y Buitrago, J. (2023). Análisis de políticas públicas para la adopción de energías renovables no convencionales en Colombia. Cuadernos Latinoamericanos de Administración, 19, 36. DOI: <https://doi.org/10.18270/cuaderlam.v19i36.4052>
- Rocha, A. (2024). Modelo de pronóstico de demanda de energía eléctrica a corto plazo basado en redes neuronales recurrentes [Tesis de grado, Escuela Politécnica Nacional]. <https://bibdigital.epn.edu.ec/handle/15000/25750>
- Rojas, J. (2024). Pronóstico de demanda de energía eléctrica en un mercado de comercialización en Colombia [Tesis de grado, Universidad de Antioquia]. Disponible en: <https://hdl.handle.net/10495/40395>
- Samuelson, P. A. y Nordhaus, W. D. (2010). Macroeconomía con aplicaciones en Latinoamérica. McGraw-Hill.
- Torres, C., Rojas, A., Higuera, D., Hernández, J. y Calle, J. (2017). Planeamiento estratégico del sector de las fuentes no convencionales de energía renovable en Colombia. [Tesis de grado, Pontificia Universidad Católica del Perú]. <http://hdl.handle.net/20.500.12404/8788>
- Ubal, C., Di-Giorgi, G., Contreras-Reyes, J. E., & Salas, R. (2023). Predicting the long-term dependencies in time series using recurrent artificial neural networks. Machine Learning and Knowledge Extraction, 5(4), 1340–1358. <https://doi.org/10.3390/make5040068>
- Wang, W. (2025). Theories for analyzing issues of household energy consumption. En Household energy consumption in China. Consumo energético doméstico en China. Cambio climático y transición energética. Springer, Singapur. https://doi.org/10.1007/978-981-96-7941-6_2
- XM SA ESP. (s. f.). *GitHub*. GitHub. Recuperado el 1 de abril de 2026, de <https://github.com/XM-SA-ESP>
- Zhang, Y., Dong, X., Wang, X., Zhang, P., Liu, M., Zhang, Y., & Xiao, R. (2023). The relationship between the low-carbon industrial model and human well-being: A case study of the electric power industry. Energies, 16(3), 1357. <https://doi.org/10.3390/en16031357>
- Zhu, H., Hao, H., & Lu, C. (2024). Enhanced support vector machine-based moving regression strategy for response prediction and reliability estimation of complex structure. Aerospace Science and Technology, 155, 109634. <https://doi.org/10.1016/j.ast.2024.109634>
- Zucaro, A., & Agostinho, F. (2025). Urban sustainability: Challenges and opportunities for resilient and resource-efficient cities. Frontiers in Sustainable Cities, 7, 1556974. <https://doi.org/10.3389/frsc.2025.1556974>