

Maestría en

Inteligencia Artificial Aplicada

Trabajo previo a la obtención de título de Magister en Inteligencia Artificial Aplicada

AUTOR/ES:

Jácome Guanoluisa Richard Xavier

Molineros Negrete Pablo Alexander

Márquez Chávez Cristian Andrés

TUTOR/ES:

Avalos Silva Héctor Guillermo

Mora Cajas Karla Estefanía

Aplicación de Arquitecturas Híbridas de Aprendizaje
Profundo para Predecir Condiciones Climáticas Utilizando
Datos Históricos del Distrito Metropolitano de Quito

Certificación de autoría

Nosotros, **Richard Xavier Jácome Guanoluisa, Pablo Alexander Molineros Negrete, Cristian Andrés Márquez Chávez** declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

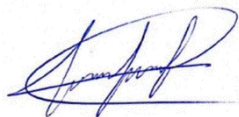
Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



Firma
Richard Xavier Jácome Guanoluisa



Firma
Pablo Alexander Molineros Negrete

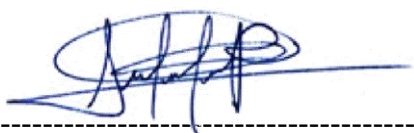


Firma
Cristian Andrés Márquez Chávez

Autorización de Derechos de Propiedad Intelectual

Nosotros, **Richard Xavier Jácome Guanoluisa, Pablo Alexander Molineros Negrete, Cristian Andrés Márquez Chávez** en calidad de autores del trabajo de investigación titulado *“Aplicación de Arquitecturas Híbridas de Aprendizaje Profundo para Predecir Condiciones Climáticas Utilizando Datos Históricos del Distrito Metropolitano de Quito”*, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

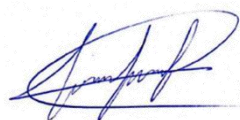
D. M. Quito, agosto de 2026



Firma
Richard Xavier Jácome Guanoluisa



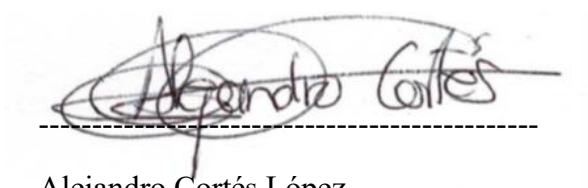
Firma
Pablo Alexander Molineros Negrete



Firma
Cristian Andrés Márquez Chávez

Aprobación de dirección y coordinación del programa

Nosotros, Alejandro Cortés López y Karla Estefanía Mora Cajas declaramos que: Richard Xavier Jácome Guanoluisa, Pablo Alexander Molineros Negrete, Cristian Andrés Márquez Chávez son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López

Director/a de la

Maestría en Inteligencia Artificial Aplicada



Karla Estefanía Mora Cajas

Coordinador/a de la

Maestría en Inteligencia Artificial Aplicada

DEDICATORIA

Dedico este trabajo a mis padres y hermanos, por su apoyo, comprensión y confianza incondicional, quienes siempre me brindan la fortaleza necesaria para afrontar cada reto y alcanzar todo lo que tengo hoy en día. De manera muy especial, a mis abuelos, por criarme y tratarme con el mismo amor y dedicación con el que se ama a un hijo. Su presencia siempre ha sido una de las bendiciones más grandes de mi vida.

Richard Xavier Jácome Guanoluiza

Dedico este proyecto a mis padres, y a mi hermana, quienes son mi principal motivación para seguir adelante, mi apoyo incondicional y mi razón de ser. Este logro es reflejo de su amor y entrega. Sin su compañía y respaldo, este camino habría sido mucho más difícil de recorrer.

Pablo Alexander Molineros Negrete

Dedico este trabajo a mi familia, por ser el fundamento de cada uno de mis logros y el motor que me impulsó a perseverar a lo largo de este camino. Su amor incondicional, apoyo constante, comprensión y confianza me brindaron la fortaleza necesaria para afrontar cada desafío y alcanzar esta meta académica. Este logro no solo representa el resultado de mi esfuerzo, sino también el reflejo de los valores, el sacrificio y la inspiración que siempre he recibido de ustedes. Con profundo cariño y gratitud, este trabajo les pertenece.

Cristian Andrés Márquez Chávez

AGRADECIMIENTOS

Agradezco profundamente a mi familia por su amor, paciencia y apoyo en cada momento de mi vida. Asimismo, a mis compañeros, Pablito y Cris por el gran trabajo que hicimos juntos, las largas horas de investigación y su valioso respaldo en este proceso de hacer ciencia.

Richard Xavier Jácome Guanoluisa

Quiero agradecer a mi familia por inspirarme siempre a seguir aprendiendo, salir adelante y luchar cada día por ser mejor. Extiendo también mi gratitud a mis compañeros de tesis, Richard y Cristian, por su dedicación y esfuerzo en la realización de este trabajo.

Pablo Alexander Molineros Negrete

Agradecemos en a la Universidad Internacional del Ecuador (UIDE) y a la Maestría en Inteligencia Artificial Aplicada, por brindarnos las herramientas académicas y el acompañamiento necesarios para desarrollar este trabajo de investigación.

Cristian Andrés Márquez Chávez

RESUMEN

El presente trabajo de investigación evalúa distintas arquitecturas de aprendizaje profundo para el pronóstico de la temperatura en el Distrito Metropolitano de Quito, utilizando 18 años de datos históricos multivariantes (2008–2026) recopilados por la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ). Tras un riguroso preprocesamiento e imputación bayesiana, se compararon tres paradigmas bajo horizontes de corto (3 h), medio (48 h) y largo plazo (72 h): modelos estadísticos clásicos (familia ARIMA), Redes Neuronales de Grafos Espaciotemporales (STGNN: A3T-GCN y DCRNN) y Modelos Fundacionales adaptados mediante QLoRA (LLaMA, Mistral) contrastados con el modelo atmosférico ClimaX, apoyado por Recuperación Aumentada por Flujo (RAF). Los resultados validados estadísticamente con la prueba de Diebold-Mariano, demostraron que la arquitectura DCRNN alcanzó la mayor capacidad predictiva general, capturando eficazmente la dependencia espacial y reduciendo el error absoluto medio (MAE) a 0.658 °C en el corto plazo y a 1.102 °C en el largo plazo. Asimismo, el modelo SARIMA mostró una consistencia robusta, superando el rendimiento de los grandes modelos de lenguaje, los cuales resultaron propensos al sobreajuste. Finalmente, los tres modelos seleccionados fueron empaquetados en Docker como un módulo de inferencia y simulación, comprobando su portabilidad con el conjunto de prueba, lo que sienta las bases funcionales para futuras herramientas orientadas a la gestión del riesgo climático local.

Palabras clave: pronóstico de temperatura, redes neuronales de grafos, modelos fundacionales, series temporales multivariadas

ABSTRACT

This research evaluates different deep learning architectures for temperature forecasting in the Metropolitan District of Quito, using 18 years of historical multivariate data (2008–2026) collected by the Quito Metropolitan Atmospheric Monitoring Network (REMMAQ). Following rigorous preprocessing and Bayesian imputation, three paradigms were compared under short- (3 h), medium- (48 h), and long-term (72 h) horizons: classical statistical models (ARIMA family), Spatiotemporal Graph Neural Networks (STGNN: A3T-GCN and DCRNN), and Foundation Models adapted via QLoRA (LLaMA, Mistral) contrasted with the ClimaX atmospheric model, supported by Retrieval-Augmented Flow (RAF). Results statistically validated using the Diebold-Mariano test demonstrate that the DCRNN architecture achieved the highest overall predictive capacity, effectively capturing spatial dependence and reducing the Mean Absolute Error (MAE) to 0.658 °C in the short term and 1.102 °C in the long term. Additionally, the SARIMA model demonstrated robust consistency, outperforming large language models, which proved prone to overfitting. Finally, the three selected models were packaged in Docker as an inference and simulation module, verifying their portability with the test set, thereby laying the functional foundations for future tools oriented toward local climate risk management.

Keywords: temperature forecasting, graph neural networks, foundation models, multivariate time series

TABLA DE CONTENIDOS

CAPÍTULO I	1
Introducción	1
Alcance	2
Justificación e Importancia del Trabajo de Investigación	2
Objetivos	4
<i>Objetivo General</i>	4
<i>Objetivos Específicos</i>	4
Hipótesis	4
CAPÍTULO II	6
Estado del Arte	6
Algoritmos de Aprendizaje Automático Tradicional	10
Modelos Espaciotemporales Basados en Grafos (STGNN)	12
<i>Modelo A3T-GCN</i>	14
<i>Modelo DCRNN</i>	16
Modelos Fundacionales Aplicados en Series Temporales	19
<i>Arquitectura Nativa Meteorológica: ClimaX y Vision Transformers (ViT)</i>	20
<i>Arquitectura de Lenguaje LLama (Meta)</i>	22
<i>Arquitectura de Lenguaje Mistral</i>	23
<i>Ajuste Fino (Fine-Tuning)</i>	23
<i>Adaptación de Bajo Rango Cuantizada (QLoRA)</i>	24
<i>Retrieval-Augmented Forecasting (RAF)</i>	25
Métricas de Validación de Resultados	26
CAPÍTULO III	29
Descripción del Conjunto de Datos y Variables	29
Análisis Exploratorio de Datos (EDA)	31
<i>Análisis Multivariante</i>	33
<i>Análisis de Dinámicas Temporales</i>	35
Resumen de Decisiones Sobre los Datos	37

Preprocesamiento de Datos	38
<i>Alineación Temporal y Consistencia Estructural</i>	<i>38</i>
<i>Control de Calidad y Filtros Físicos (QA/QC)</i>	<i>40</i>
<i>Estrategia de Particionamiento de Datos</i>	<i>42</i>
<i>Tratamiento de Datos Faltantes</i>	<i>43</i>
Ingeniería de Características (Feature Engineering).....	47
Implementación de Algoritmos Tradicionales	48
<i>Construcción del Conjunto Experimental de Datos</i>	<i>52</i>
<i>Entrenamiento de Algoritmos Tradicionales</i>	<i>53</i>
<i>Construcción del Conjunto Experimental de Datos</i>	<i>56</i>
<i>Entrenamiento de Algoritmos Tradicionales</i>	<i>57</i>
<i>Optimización de Hiperparámetros.....</i>	<i>59</i>
<i>Validación Mediante Walk-Forward</i>	<i>61</i>
Integración de Modelos STGNN.....	66
<i>Formalización del Grafo Espacial</i>	<i>66</i>
<i>Formulación del Problema de Pronóstico</i>	<i>68</i>
<i>Arquitectura y Configuración Experimental de los Modelos.....</i>	<i>70</i>
<i>Evaluación de los Modelos STGNN</i>	<i>74</i>
<i>Protocolo de Comparación y Significancia Estadística.....</i>	<i>75</i>
<i>Flujo de Entrenamiento de Modelos STGNN.....</i>	<i>76</i>
Aplicación de Arquitecturas Fundacionales	77
<i>Fase 1: Preprocesamiento Diferenciado para Modelos Fundacionales</i>	<i>79</i>
<i>Fase 2: Benchmark Inicial y Selección de la Arquitectura Base.....</i>	<i>82</i>
<i>Fase 3: Optimización y Adaptación Multiescala.....</i>	<i>94</i>
CAPÍTULO IV	114
Comparación de Métricas entre Modelos Implementados.....	114
Evaluación Estadística de la Precisión	115
Evaluación de Predicciones	116
Evaluación de Predicciones a Nivel de Estación.....	118
Discusión de Resultados	121

Implementación del Sistema de Simulación e Inferencia	123
CAPÍTULO V	130
Conclusiones	130
Recomendaciones	131
BIBLIOGRAFÍA.....	133
APÉNDICE A.....	139
Repositorio de Código Fuente.....	139
APÉNDICE B	139
Repositorio de la Base de Datos y Artefactos Generados	139

LISTA DE TABLAS

Tabla 1	18
Tabla 2	28
Tabla 3	29
Tabla 4	37
Tabla 5	40
Tabla 6	43
Tabla 7	46
Tabla 8	48
Tabla 9	52
Tabla 10	56
Tabla 11	58
Tabla 12	60
Tabla 13	63
Tabla 14	65
Tabla 15	69
Tabla 16	71
Tabla 17	71
Tabla 18	74
Tabla 19	74
Tabla 20	76
Tabla 21	82
Tabla 22	88
Tabla 23	92
Tabla 24	93
Tabla 25	101
Tabla 26	111
Tabla 27	112
Tabla 28	114
Tabla 29	115
Tabla 30	116

LISTA DE FIGURAS

Figura 1	16
Figura 2	17
Figura 3	22
Figura 4	31
Figura 5	32
Figura 6	33
Figura 7	34
Figura 8	35
Figura 9	36
Figura 10	51
Figura 11	55
Figura 12	64
Figura 13	67
Figura 14	68
Figura 15	70
Figura 16	73
Figura 17	77
Figura 18	89
Figura 19	89
Figura 20	90
Figura 21	91
Figura 22	96
Figura 23	99
Figura 24	103
Figura 25	109
Figura 26	116
Figura 27	117
Figura 28	119
Figura 29	120
Figura 30	120
Figura 31	125
Figura 32	127
Figura 33	129

CAPÍTULO I

Introducción

El estudio se orienta en el diseño y evaluación de arquitecturas híbridas para el pronóstico de la temperatura en entornos de orografía compleja, tomando como caso de análisis el Distrito Metropolitano de Quito (DMQ). La predicción de la temperatura en esta región enfrenta desafíos críticos debido a su geografía irregular y la presencia de múltiples microclimas, factores que limitan la precisión de los modelos predictivos tradicionales y lineales. El comportamiento atmosférico local está condicionado por la interacción entre variables meteorológicas y factores topográficos, los cuales influyen directamente en la variabilidad de la temperatura y constituyen información relevante para su modelado predictivo (Secretaría de Ambiente del Distrito Metropolitano de Quito, 2024). Además, el diseño de este estudio se rige por un protocolo de selección jerárquica estructurado en un torneo de dos etapas, fundamentado en el marco de evaluación propuesto por Cerqueira et al. (2021) y en el uso riguroso del estadístico de Diebold (2015). En la primera etapa (Competición Intra-Arquitectura), se identifica la configuración óptima dentro de cada familia de modelos para, posteriormente, enfrentar a los ganadores en una competición final (Inter-Arquitectura) que determine la solución definitiva. Este proceso integra la significancia estadística mediante la prueba de Diebold-Mariano y criterios de desempate técnico como la robustez ante errores extremos de Chai y Draxler (2014), la coherencia física contra el suavizado excesivo descrito por Zeng et al. (2022) y el principio de parsimonia o Green AI de Schwartz et al. (2020).

Alcance

El diseño de la solución metodológica contempla la predicción de la temperatura a partir de las estaciones ubicadas estratégicamente en el Distrito Metropolitano de Quito, las cuales funcionarán como puntos de observación. Asimismo, se realizará un análisis cuantitativo para determinar qué técnicas propuestas alcanzan la mayor precisión en la tarea de predicción. Como parte de la validación, se desarrollará un entorno de ejecución local basado en contenedores Docker, que permitirá simular el funcionamiento del sistema sobre el conjunto de prueba y visualizar los pronósticos generados por la solución más efectiva. Cabe señalar que el alcance del estudio no incluye la extensión a otras ciudades ni la generación de predicciones en tiempo real.

Adicionalmente, se definen tres escalas de pronóstico en función de la ventana temporal de entrada (datos históricos considerados) y del horizonte de salida (periodo a predecir). Esta diferenciación permite evaluar la capacidad de los modelos en distintos escenarios, desde la identificación de tendencias de largo plazo hasta la detección de variaciones rápidas en el corto plazo.

Justificación e Importancia del Trabajo de Investigación

El impacto socioeconómico de los eventos climáticos extremos exige una investigación profunda orientada a mejorar la precisión de los pronósticos meteorológicos, insumo clave para la planificación y la toma de decisiones en ámbitos como la gestión ambiental, la movilidad, la agricultura y la reducción de riesgos. Para responder a esta problemática, se plantea un marco metodológico basado en la integración de Redes Neuronales de Grafos Espaciotemporales (STGNN), que permiten capturar de manera efectiva las dependencias espaciotemporales

complejas y no lineales. Los modelos paramétricos tradicionales, como ARIMA, asumen relaciones lineales y presentan dificultades críticas para gestionar la incertidumbre y los cambios bruscos en las series de tiempo, limitando su capacidad predictiva (Zhao et al., 2019).

En esta misma línea, y considerando el desafío de procesar conjuntos masivos de datos con múltiples variables interrelacionadas, las arquitecturas propuestas superan las restricciones de los modelos ARIMA al procesar simultáneamente variables exógenas y captar interacciones espaciales no euclidianas mediante la estructura de grafos (Li et al., 2018). Este enfoque se potencia con la transferencia de conocimiento desde Modelos Fundacionales preentrenados (Large Language Models, LLMs), técnica emergente que aprovecha patrones climáticos globales para mejorar predicciones locales con datos limitados (Nguyen et al., 2023). El resultado esperado es un sistema robusto que incorpore la distribución espacial de las estaciones, transformando la topografía en un factor predictivo activo y no solo en un simple ruido (Wu et al., 2020).

La viabilidad técnica del uso de LLMs se sustenta en la implementación de la técnica de Adaptación de Bajo Rango Cuantizada (QLoRA, 4-bit), que reduce el consumo de memoria de video (VRAM) de aproximadamente 28 GB a 8 GB en modelos con 7 mil millones de parámetros (Dettmers et al., 2024). Esta optimización permite realizar ajustes finos en Unidades de Procesamiento Gráfico (GPU) disponibles en entornos de acceso libre, como la NVIDIA T4 (16 GB) y la NVIDIA A100 en plataformas como Google Colab, eliminando la necesidad de infraestructura industrial o clústeres de alto rendimiento. Además, el bajo peso del dataset

consolidado facilita la aplicación de técnicas como RAF, garantizando que las experimentaciones puedan desarrollarse en entornos computacionales accesibles.

Objetivos

Objetivo General

Comparar modelos de predicción de temperatura basados en técnicas estadísticas, arquitecturas de redes de grafos espaciotemporales y modelos fundacionales, con el propósito de evaluar su desempeño en el pronóstico meteorológico del Distrito Metropolitano de Quito.

Objetivos Específicos

Recopilar datos de series temporales multivariantes y metadatos geográficos obtenidos de las estaciones de monitoreo del Distrito Metropolitano de Quito

Establecer un marco comparativo de rendimiento mediante la validación de la capacidad de predicción de la temperatura en el Distrito Metropolitano de Quito utilizando métricas de error (RMSE, MAE, MAPE y R^2) en distintos horizontes temporales.

Desarrollar un entorno de ejecución local mediante contenedores Docker que permita simular el proceso de inferencia y facilitar la visualización de las predicciones obtenidas por el modelo seleccionado.

Hipótesis

Basado en un protocolo experimental estandarizado y riguroso, este estudio plantea que la integración de la distribución espacial de las estaciones de monitoreo, mediante redes de grafos

espaciotemporales (STGNN), mejorará la predicción de la temperatura en el Distrito Metropolitano de Quito. Se espera que, a corto plazo, estos modelos superen de manera consistente a los métodos estadísticos clásicos y fundacionales que ignoran dicha estructura espacial.

Sin embargo, esta ventaja predictiva variará según el horizonte de pronóstico: a medida que este se extienda, la señal espacial perderá relevancia frente al componente temporal, un patrón ya documentado en la literatura reciente (Ma et al., 2023; Nguyen et al., 2025).

CAPÍTULO II

Estado del Arte

La base fundamental de la predicción de series temporales ha sido históricamente dominada por modelos estadísticos lineales, siendo ARIMA el referente de mayor aplicabilidad debido a su sencillez y robustez matemática. Como señalan Box et al. (2015), estos modelos son esenciales para establecer una línea base, ya que permiten identificar la estacionalidad y las tendencias básicas sin la opacidad de los modelos de caja negra. Sin embargo, su principal limitación radica en la incapacidad de procesar dependencias no lineales y la falta de sensibilidad ante variables exógenas complejas, lo que resulta insuficiente para entornos de orografía irregular donde el clima no sigue un patrón puramente cíclico.

Para superar estas limitaciones, las investigaciones recientes han destacado el potencial de las Redes Neuronales de Grafos Espaciotemporales (STGNN). A diferencia de los modelos convencionales, que consideran cada estación meteorológica como un punto aislado, las STGNN representan la topología como un grafo en el que los nodos corresponden a estaciones y las aristas reflejan relaciones geográficas y físicas. Según Wu et al. (2020), esta capacidad de “conectar los puntos” permite capturar dependencias espaciotemporales críticas. La principal ventaja frente al análisis de series independientes es la incorporación de información georreferenciada al integrar la distancia real, el modelo comprende por qué una tormenta en un nodo afecta de manera diferenciada a sus vecinos según la topografía, reduciendo el error de

predicción en más del 60% respecto de enfoques que omiten el contexto geográfico (J. Jiang et al., 2026).

En esta misma línea, Ma et al. (2023) demuestran que es posible modelar simultáneamente las dependencias locales entre variables meteorológicas y las interacciones regionales entre estaciones, capturando interdependencias no lineales del sistema atmosférico en redes urbanas. Por otra parte, Ali et al. (2025) introducen la inferencia causal mediante la prueba de Granger, lo que permite identificar direccionalidad y latencia en la propagación de fenómenos climáticos, mostrando cómo las alteraciones en un nodo repercuten dinámicamente en sus vecinos a lo largo del tiempo.

De manera análoga, los marcos híbridos propuestos por Nurcahyo y Putranto (2026), evidencian la viabilidad de estas tecnologías en redes de estaciones con distribución geográfica irregular en regiones tropicales, aportando una base metodológica sólida que se alinea con el caso de estudio del Distrito Metropolitano de Quito, donde resulta esencial captar la inercia térmica y las variaciones topográficas locales.

Este avance en la comprensión espacial se ve actualmente potenciado por el surgimiento de los Modelos Fundacionales en meteorología. Al igual que los modelos de lenguaje a gran escala, herramientas como ClimaX (Nguyen et al., 2023) y GraphCast (Lam et al., 2023) han sido preentrenados con décadas de datos climáticos globales, permitiendo una transferencia de conocimiento sin precedentes. La ventaja técnica reside en que estos modelos ya "entienden" la física atmosférica global; por lo tanto, al aplicarlos a una región específica, la predicción es mucho más exacta incluso si los datos locales son escasos. Esta capacidad de ajuste fino a partir

de un conocimiento previo de climas similares permite que el sistema identifique patrones latentes que un modelo entrenado desde cero no detectaría.

El análisis de Modelos de Circulación General Neuronales (NeuralGCM) aborda las limitaciones de estabilidad en modelos híbridos. Este tipo de soluciones integra un núcleo dinámico diferenciable capaz de resolver ecuaciones relacionadas a la atmósfera a gran escala utilizando módulos de física entrenados con redes neuronales. Gracias a esta combinación, se evitan las predicciones inestables a largo plazo que suelen caracterizar a los modelos basados únicamente en aprendizaje profundo. Esta arquitectura permite ahorrar hasta cinco veces más en términos de eficiencia computacional comparado con modelos de circulación general (GCM). En este sentido, la integración híbrida propuesta resulta prometedora para contextos urbanos como el Distrito Metropolitano de Quito, ya que las predicciones no solo alcanzan un alto grado de precisión, sino que también se sustentan en la física atmosférica propia de la región (Kochkov, 2024).

En este sentido, la integración de física y aprendizaje profundo no es la única alternativa para mejorar la estabilidad de los pronósticos; también han surgido enfoques complementarios que buscan aprovechar el patrón histórico de los datos, como el marco Retrieval-Augmented Forecasting (RAF), que integran fragmentos históricos con patrones similares al contexto actual, evitando así la pérdida de datos relevantes.

La técnica RAF se basa en los sistemas de Retrieval-Augmented Generation (RAG) que revolucionaron el procesamiento del lenguaje natural mediante el uso de bases de conocimiento para aliviar el problema de las alucinaciones y las respuestas estén acorte al contexto. De manera

similar a cómo RAG recupera datos importantes de texto para responder una pregunta, o a cómo los sistemas de visión por computador extraen imágenes para mejorar la clasificación, RAF aplica esta lógica a las series de tiempo, considerando que se trabaja con datos estructurados y secuenciales (Li, 2025). La aplicación de RAF tiene un efecto importante en datos de entornos climáticos, ya que permiten identificar y aprovechar eventos pasados, como las crisis meteorológicas, para evitar la volatilidad inherente a estos estudios y mejorar la precisión de pronósticos. Además, el uso de mecanismos de hashing dinámico de esquemas facilita el manejo de la rigidez de los datos alineando automáticamente las variables meteorológicas incluso cuando cambian los esquemas de etiquetado. Esto garantiza que el modelo reconozca el contexto de los datos a pesar de las variaciones en los estándares de nombrado (Li, 2025).

Si bien los modelos fundacionales tienen un gran potencial en el pronóstico del clima, su implementación exige un alto consumo recursos computacionales para el ajuste fino. Frente a esta limitación surge la Adaptación de Bajo Rango Cuantizada (QLoRA), que optimiza grandes modelos mediante parámetros adaptativos y mantiene la precisión en 4 bits, reduciendo significativamente el uso de memoria GPU y permitiendo entrenar arquitecturas complejas en entornos de acceso libre (Dettmers, 2024).

Finalmente, la integración de estas tecnologías culmina en marcos metodológicos híbridos que reprograman el conocimiento existente para tareas específicas. Propuestas como TIME-LLM (Jin et al., 2024) demuestran que es posible aprovechar la capacidad de razonamiento de modelos avanzados para interpretar series temporales complejas.

En conclusión, el estado del arte sugiere que una predicción meteorológica moderna para regiones de alta complejidad debe necesariamente combinar la solidez de los métodos tradicionales como punto de comparación, la precisión espacial de las redes de grafos para entender el territorio y la potencia de los modelos fundacionales para garantizar exactitud basada en el conocimiento climático global.

Algoritmos de Aprendizaje Automático Tradicional

El modelo autorregresivo integrado de media móvil (ARIMA, por sus siglas en inglés) es considerado uno de los pilares en la predicción de series temporales dentro del aprendizaje automático tradicional. Su relevancia radica en la capacidad de representar procesos no estacionarios mediante la combinación de tres componentes: autorregresivo, integrado y de media móvil. Según Morera Campos (2025), este modelo se emplea como técnica estadística clásica en la predicción de series meteorológicas, ajustando parámetros a través de funciones de autocorrelación y autocorrelación parcial, lo que permite capturar patrones de variabilidad en fenómenos como la temperatura.

La lógica del modelo ARIMA se fundamenta en que los valores futuros de una serie pueden explicarse a partir de sus valores pasados y de los errores previos. Esta estructura lo convierte en una herramienta flexible, capaz de adaptarse a diferentes tipos de datos mediante la selección adecuada. En el caso del Distrito Metropolitano de Quito, donde la variabilidad climática es marcada, esta flexibilidad resulta esencial para modelar fluctuaciones irregulares y tendencias de corto plazo.

La aplicación de ARIMA está condicionado por supuestos estadísticos que garantizan su validez. La estacionariedad es uno de los más importantes, ya que la serie debe mantener media y varianzas constantes en el tiempo, lo cual se logra mediante diferenciación (Quispe Galindo, 2024). También se asume linealidad en las relaciones entre valores pasados y presentes, y se exige que los residuos se comporten como ruido blanco, es decir, independientes y con distribución normal de media cero (Naquintero Garavito, 2025). Estos supuestos, aunque estrictos, son los que permiten que el modelo obtenga estimaciones confiables; sin embargo, en fenómenos naturales como el clima pueden resultar difíciles de cumplir en su totalidad.

Las ventajas de esta metodología se relacionan con su simplicidad matemática y su capacidad de interpretación directa. González González et al. (2018) señala que los modelos de predicción basados en ARIMA han demostrado eficiencia en la estimación de variables meteorológicas de corto plazo, siempre que se cumplan los supuestos de estacionariedad y ruido blanco. A ello se suma la metodología Box-Jenkins, que proporciona un procedimiento sistemático de identificación, estimación y diagnóstico, garantizando reproducibilidad y rigor en la construcción del modelo (Muñoz Mandujano, 2024).

No obstante, el modelo presenta limitaciones que deben ser consideradas. Su incapacidad para capturar relaciones no lineales lo hace menos adecuado en fenómenos atmosféricos complejos, y su sensibilidad a la estacionalidad suele requerir extensiones como SARIMA. Además, la necesidad de diferenciar la serie puede implicar pérdida de información estructural en datos con tendencias prolongadas (Morera Campos, 2025). Estas restricciones son

especialmente relevantes en el Distrito Metropolitano de Quito, donde la interacción de factores geográficos y microclimáticos genera patrones difíciles de representar con un modelo lineal.

La validación de estos modelos se realiza mediante análisis de residuos, pruebas de significancia estadística y métricas de error como RMSE, MAE o MAPE. Morera Campos (2025) señala que, en estudios aplicados, los residuos del modelo se ajustan adecuadamente a los datos, al estar centrados en cero y con distribución normal.

En conclusión, ARIMA se presenta como un algoritmo tradicional robusto y ampliamente validado. Su función será establecer una línea base sólida sobre la cual se evaluará la capacidad predictiva de técnicas más sofisticadas, asegurando un análisis académico riguroso y estructurado.

Modelos Espaciotemporales Basados en Grafos (STGNN)

Las metodologías tradicionales de predicción de variables temporales presentan grandes limitantes al tener que validar ciertos supuestos de estacionariedad o linealidad en los datos, que no reflejan el comportamiento natural de las propiedades climatológicas. Métodos como ARIMA o la regresión lineal suelen tener problemas para capturar condiciones ambientales, ya que no logran representar la estructura topológica ni las dependencias a largo plazo, lo que las hace sensibles a eventos extremos o a datos faltantes en diversos periodos de tiempo (Tang et al., 2025).

Por su parte, metodologías basadas en Deep Learning, como las RNN y sus variantes LSTM, han sido ampliamente utilizadas para tareas de predicción de series de tiempo gracias a

su capacidad de capturar dependencias temporales; sin embargo, no modelan explícitamente la estructura espacial y relacional entre diferentes puntos de observación. Además, asumen secuencias de datos completas y de longitud fija, lo que también las hace sensibles a fallos en los sensores o en las lecturas satelitales. Estos desafíos evidencian la necesidad de un marco de trabajo más flexible y consistente con dicha estructura (Tang et al., 2025).

Frente a esta necesidad de un marco más flexible ante la topología de los datos, han surgido diversas arquitecturas capaces de capturar tanto el comportamiento temporal como el espacial. Estas se engloban dentro de los modelos de Redes Neuronales de Grafos Espaciotemporales (STGNN), que permiten modelar relaciones no euclidianas e irregulares entre distintos puntos de observación. Un ejemplo destacado es el modelo HiSTGNN, que utiliza un enfoque jerárquico para capturar las interacciones entre variables meteorológicas en una estación local, así como la transferencia de información entre diferentes regiones a nivel global (Ma et al., 2023).

Bajo este enfoque, el caso del Distrito Metropolitano de Quito resulta particularmente relevante: integrar su topología y características geográficas en los modelos de predicción ofrece una herramienta poderosa para entender la dispersión de variables como la temperatura y el clima. La ciudad, ubicada a 2850 metros sobre el nivel del mar, experimenta fenómenos como inversiones térmicas inducidas topográficamente, por lo que considerar la morfología urbana y el relieve permite identificar patrones y variaciones climáticas que los modelos generales omiten (Álvarez et al., 2025).

En línea con el marco espaciotemporal, resulta necesario formalizar el problema de predicción de temperatura en términos de grafos, dado que las arquitecturas STGNN se construyen sobre esta representación matemática. Así, cada estación meteorológica del Distrito Metropolitano de Quito se modela como un nodo dentro de un grafo $G = (V, E)$, donde las aristas representan las relaciones espaciales derivadas de la proximidad geográfica, la topografía o la correlación climática entre estaciones. La temperatura registrada en cada nodo se representa mediante una matriz de características X , que contiene la evolución temporal de las observaciones. De esta manera, la predicción climática se formula como el aprendizaje de una función capaz de captar simultáneamente las dependencias espaciales y temporales (Zhu et al., 2021).

A partir de esta representación, arquitecturas como A3T-GCN y DCRNN se convierten en soluciones idóneas para abordar el problema. A3T-GCN introduce mecanismos de atención que ponderan la importancia de las conexiones espaciales, logrando mayor robustez y precisión, así como la capacidad de capturar variaciones globales y relaciones de largo alcance (Zhu et al., 2021). DCRNN, por su parte, combina convoluciones de grafos con redes recurrentes, lo que le permite capturar tanto la propagación espacial de la información como las dependencias espaciales a largo plazo (Li et al., 2018).

Modelo A3T-GCN

La arquitectura A3T-GCN (Attention Temporal Graph Convolutional Network) es un modelo de predicción diseñado para capturar simultáneamente las correlaciones espaciales y las dinámicas temporales globales en datos estructurados como grafos. Este modelo representa una

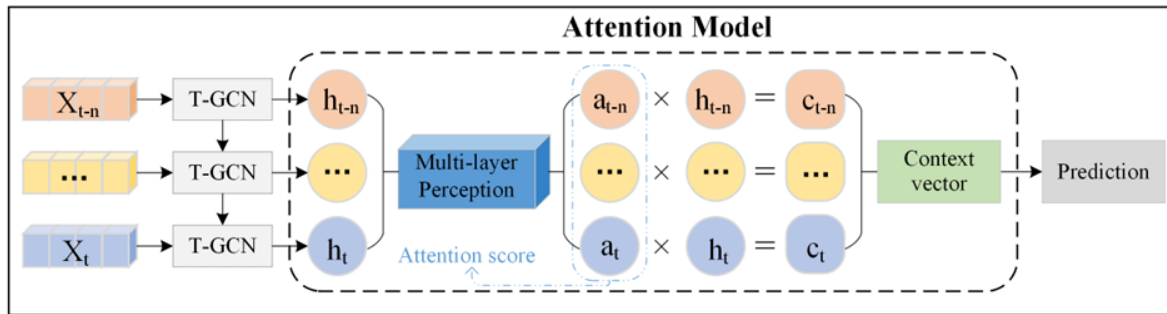
mejora significativa sobre el modelo base T-GCN (Zhao et al., 2019), al introducir un mecanismo de atención que permite reajustar la influencia de los estados históricos. Su objetivo es aprender una función que, a partir de las secuencias de datos pasados y la topología de la red, infiera los estados futuros de manera precisa (Zhu et al., 2021).

El marco de trabajo del modelo A3T-GCN integra tres componentes fundamentales que funcionan de manera conjunta:

Red Convolutiva de Grafos (GCN): se utiliza para aprender la dependencia espacial a partir de la morfología de la red, codificando simultáneamente las estructuras topológicas y los atributos de los nodos mediante operaciones de convolución.

Unidad Recurrente de Compuerta (GRU): captura las tendencias de corto plazo y la dinámica temporal de la serie de datos. Mediante mecanismos de compuertas (puertas de reinicio y actualización), retiene la información relevante a largo plazo y descarta el ruido irrelevante, facilitando el aprendizaje de las variaciones temporales locales.

Mecanismo de Atención: se implementa un modelo de atención suave (soft-attention) para ensamblar la información temporal global. Este mecanismo calcula una puntuación para cada estado oculto generado por la combinación de la GCN y la GRU, lo que permite al modelo identificar qué momentos del pasado son más relevantes para la predicción actual, independientemente de su proximidad en el tiempo (Zhu et al., 2021). La Figura 1 ilustra los componentes de la arquitectura.

Figura 1*A3T-GCN Framework*

Nota. Tomado de "A3T-GCN: Attention Temporal Graph Convolutional Network for Traffic Forecasting", por J. Zhu, Y. Song, L. Zhao y H. Li, 2021, ISPRS International Journal of Geo-Information, 10(7), 485, p. 6.

Modelo DCRNN

El modelo DCRNN (Diffusion Convolutional Recurrent Neural Network) es un framework de aprendizaje profundo diseñado para la predicción espaciotemporal que integra dependencias geográficas y dinámicas temporales. A diferencia de otros modelos, el DCRNN formula el flujo de información como un proceso de difusión estocástico sobre un grafo dirigido, lo que permite capturar la naturaleza compleja de los datos en redes de sensores (Li et al., 2018).

El marco de trabajo del modelo DCRNN está compuesto por tres pilares importantes:

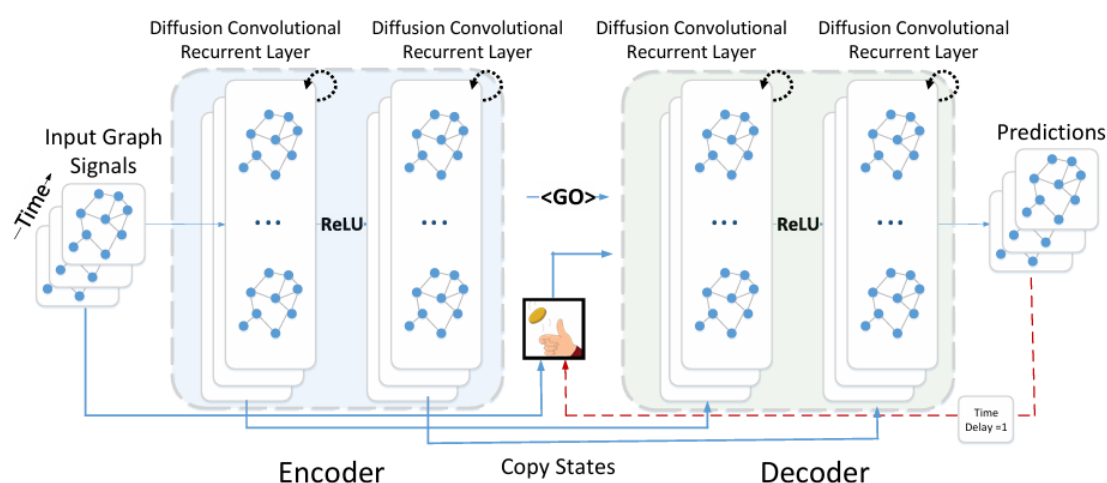
Convolución por Difusión: Esta operación reemplaza a las convoluciones estándar para capturar dependencias espaciales mediante caminatas aleatorias bidireccionales en el grafo, permitiendo que el modelo aprenda la influencia de nodos que se conectan bidireccionalmente.

DCGRU (Diffusion Convolutional Gated Recurrent Unit): Es una variante de la unidad recurrente GRU donde las multiplicaciones de matrices tradicionales son sustituidas por la operación de convolución por difusión, permitiendo al modelo aprender dinámicas temporales sin perder la estructura espacial.

Arquitectura Encoder-Decoder y Scheduled Sampling: Para realizar predicciones de múltiples pasos hacia adelante, utiliza un esquema de secuencia a secuencia. Durante el entrenamiento, se implementa el muestreo programado, una técnica que alimenta al modelo con una mezcla de datos reales y sus propias predicciones previas para mitigar la acumulación de errores y el sesgo de exposición durante la fase de prueba (Li et al., 2018). La Figura 2 muestra los componentes de la arquitectura.

Figura 2

DCRNN Framework



Nota. Tomado de "Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting", por Y. Li, R. Yu, C. Shahabi y Y. Liu, 2018, p. 7.

En conjunto, tanto A3T-GCN como DCRNN ofrecen aproximaciones complementarias al problema de la predicción de la temperatura. La Tabla 1 presenta un resumen comparativo que permite identificar las fortalezas y limitaciones de cada enfoque.

Tabla 1

Comparativa de los Modelos A3T-GCN y DCRNN

Característica	A3T-GCN	DCRNN
Modelado Espacial	Utiliza Redes Convolucionales de Grafos (GCN) basadas en la topología de la red.	Utiliza convolución por difusión mediante caminatas aleatorias bidireccionales.
Modelado Temporal	Emplea Unidades Recurrentes de Compuerta (GRU) para aprender tendencias de corto plazo y periodicidad.	Utiliza una variante DCGRU (Diffusion Convolutional GRU) que integra la difusión directamente en las compuertas de la unidad.
Tipo de Grafo	Opera sobre grafos que codifican estructuras topológicas fijas o basadas en proximidad.	Opera sobre grafos dirigidos, lo que permite modelar de manera asimétrica los flujos de datos.
Mecanismo	Mecanismo de Atención (Soft Attention) para ajustar la importancia de diferentes puntos en el tiempo y capturar dinámicas globales.	Muestreo Programado (Scheduled Sampling) dentro de una arquitectura encoder-decoder para mitigar la propagación de errores.
Objetivo Principal	Capturar simultáneamente correlaciones espaciales y dinámicas temporales globales ponderando estados históricos.	Modelar la dinámica espaciotemporal como un proceso de difusión estocástico en una red.
Fortaleza	Es robusto frente al ruido y posee una alta capacidad para identificar la relevancia de los datos históricos.	Es preciso en la captura de asimetrías espaciales y de cambios abruptos, gracias a su naturaleza de difusión.

Limitación	Las estructuras recurrentes tienden a otorgar mayor importancia a la información reciente, lo que puede requerir un entrenamiento más extenso.	Es sensible a la calidad del grafo; si la definición de los pesos es deficiente, la capacidad de difusión del modelo pierde precisión.
------------	--	--

Nota. Elaboración propia a partir de “A3T-GCN: Attention Temporal Graph Convolutional Network for Traffic Forecasting”, por J. Zhu, Y. Song, L. Zhao y H. Li, 2021, ISPRS International Journal of Geo-Information, 10(7), 485; y de “Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting”, por Y. Li, R. Yu, C. Shahabi y Y. Liu, 2018, pp. 1-11.

Modelos Fundacionales Aplicados en Series Temporales

El surgimiento de los Modelos Fundacionales aplicados a la meteorología plantea un fortalecimiento significativo en el pronóstico de variables climáticas en entornos complejos. Este potencial se fundamenta en los siguientes principios:

Los modelos preentrenados como ClimaX (Nguyen et al., 2023) y GraphCast (Lam et al., 2023), son métodos cuyo entrenamiento se basa en datos históricos climáticos globales, lo cual permite transferir el conocimiento precedente aprovechando los patrones climáticos para mejorar predicciones locales.

Los modelos preentrenados ya tienen el conocimiento de la física atmosférica global, siendo esta su principal ventaja porque pueden identificar patrones al aplicarlos sobre una zona específica, lo cual no se podría directamente entrenando un modelo desde cero.

Propuestas como TIME-LLM (Jin et al., 2024) sustentan que al usar la capacidad de modelos preentrenados se pueden interpretar series temporales complejas de manera efectiva, ya que se puede adaptar el conocimiento existente para realizar tareas en el dominio numérico.

Arquitectura Nativa Meteorológica: ClimaX y Vision Transformers (ViT)

Para superar las limitaciones de la reprogramación de modelos de lenguaje en dominios numéricos, surgieron arquitecturas nativas para el modelado meteorológico como ClimaX (Nguyen et al., 2023). Este modelo adapta la arquitectura Vision Transformer (ViT) al procesamiento de variables atmosféricas, fundamentándose en tres componentes principales:

Incrustación de parches espaciales (Patch Embedding): Este proceso se basa en Vision Transformer (ViT) propuesto por Dosovitskiy et al. (2021), ClimaX reemplaza el procesamiento local mediante convoluciones por una representación basada en parches espaciales. Los campos meteorológicos de entrada se dividen en parches de tamaño fijo, los cuales son proyectados a un espacio latente de alta dimensión ($d = 1,024$). Esta representación permite que el mecanismo de autoatención modele relaciones entre cualquier región del dominio espacial, capturando dependencias atmosféricas de corto y largo alcance (Zeng et al., 2022).

Proyección continua de variables meteorológicas (Variable Tokenization y Variable Aggregation): A diferencia de los modelos de lenguaje (LLMs), que convierten los valores numéricos en cadenas de texto para posteriormente dividirlos en tokens pertenecientes a un vocabulario discreto, ClimaX mantiene toda la información en el dominio numérico continuo. En primer lugar, el módulo Variable Tokenization proyecta de forma independiente cada variable meteorológica de entrada (temperatura, humedad relativa, presión barométrica, radiación solar,

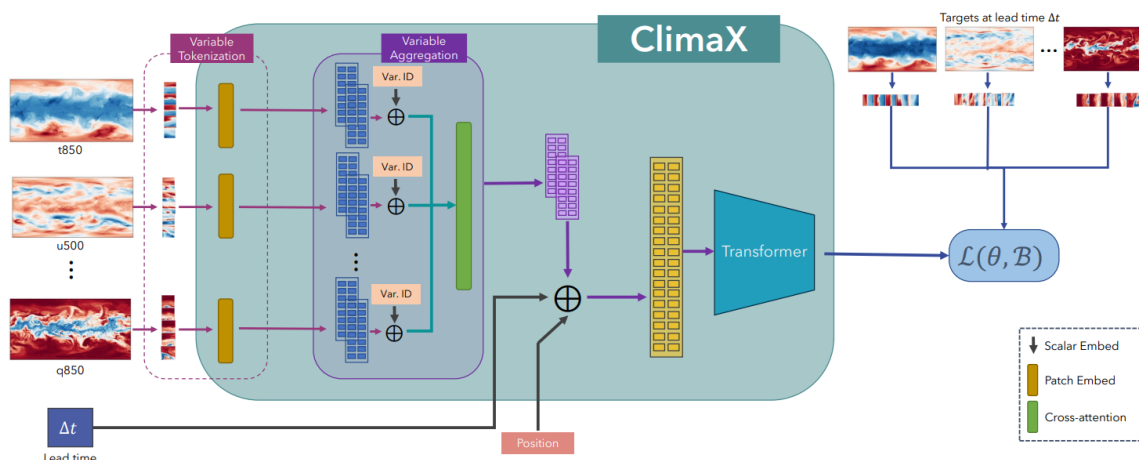
viento) hacia un vector en $\mathbb{R}^{1024}$ mediante transformaciones lineales (Nguyen et al., 2023).

Posteriormente, el módulo Variable Aggregation integra estas representaciones incorporando un identificador de variable (Var. ID), generando una representación unificada para cada parche que luego es procesada por el backbone Transformer (Nguyen et al., 2023). Este mecanismo preserva la información física original, evita las pérdidas de precisión asociadas al procesamiento textual y reduce el costo computacional durante la inferencia.

Diseño Arquitectónico Modular (Global Backbone y Local Prediction Head):

ClimaX está compuesta por un backbone basado en Transformer, el cual es preentrenado de forma autosupervisada con grandes conjuntos de datos climáticos, como ERA5 y CMIP6, incorporando además información espacial, temporal y del horizonte de predicción (lead time). Sobre este backbone es posible incorporar componentes de salida específicos (Local Prediction Heads) para tareas de regresión, permitiendo adaptar el conocimiento climático global a diferentes problemas meteorológicos sin modificar la estructura principal del modelo (Nguyen et al., 2023).

El flujo de procesamiento de ClimaX, tiene las variables meteorológicas proyectadas de forma independiente al espacio latente mediante vectores numéricos (Variable Tokenization), luego se integran con identificadores (Variable Aggregation) y se combinan con codificaciones espaciales, temporales y del horizonte de predicción antes de ingresar al backbone basado en Vision Transformer, tal como se observa en el diagrama de arquitectura de la Figura 3.

Figura 3*Arquitectura de Preentrenamiento del Modelo Fundacional ClimaX*

Nota. Tomado de ClimaX: A Foundation Model for Weather and Climate (p. 3), por T. Nguyen, J. Brandstetter, A. Kapoor, J. K. Gupta y A. Grover, 2023.

Arquitectura de Lenguaje LLama (Meta)

De la familia de modelos fundacionales abiertos que se basan en la arquitectura de transformador de un solo codificador (decoder-only transformer) proviene LLaMA 3 (Large Language Model Meta AI).

En la estructura de este modelo encontramos el mecanismo Grouped-Query Attention (GQA) que agrupa las representaciones clave y valor, lo cual reduce significativamente la carga computacional y el consumo de memoria en la inferencia a diferencia del mecanismo de atención multi-head tradicional (Meta, 2024).

Una característica importante de este modelo es Rotary Positional Embedding (RoPE), ya que permite que al modelo codificar la distancia temporal entre eventos, preservando la secuencialidad subyacente de manera más estable que los sistemas tradicionales (Meta, 2024).

Arquitectura de Lenguaje Mistral

Este modelo fundacional abierto, destaca por su optimización de recursos computacionales comparado con modelos que usan una cantidad mayor de parámetros para razonamiento secuencial (Jin et al., 2024).

Para el análisis secuencial utiliza el mecanismo Sliding Window Attention (SWA) que en lugar de calcular la atención global donde cada token atiende a toda la historia, SWA restringe la atención usando ventanas temporales superpuestas de tamaño fijo (A. Q. Jiang et al., 2023), este enfoque es similar a la ventana deslizante usada en pronósticos a corto plazo, esto permite capturar variaciones inmediatas sin perder estabilidad.

Este modelo logra manejar datos con un historial largo con un bajo impacto en la memoria caché, debido a que combina la atención de ventana deslizante con mecanismo de agrupamiento, esto permite la integración con técnicas como Retrieval-Augmented Forecasting (RAF), ya que facilita el análisis y recuperación de eventos climáticos pasados que se relacionen con la predicción actual.

Ajuste Fino (Fine-Tuning)

Este proceso consiste en tomar un Modelo Fundacional entrenado con una gran cantidad de datos y adaptar su conocimiento al entrenarlo con un conjunto de datos específico. Orientando

esto al contexto meteorológico, el ajuste fino nos permite identificar patrones que un modelo entrenado de cero no podría detectar, este proceso ayuda a adaptar el conocimiento global a las particularidades de climas locales.

Al requerir un nuevo entrenamiento sobre un LLM como LLaMA3 o Mistral se tiene una barrera de recursos computacionales, ya que el ajuste fino completo actualiza todos los pesos sinápticos, cuyo proceso necesita por encima de 100 GB de memoria de video (VRAM), lo cual supera las capacidades de entornos independientes, siendo esta una barrera técnica importante (Hu et al., 2022).

Adaptación de Bajo Rango Cuantizada (QLoRA)

Debido a que los modelos fundacionales demandan un alto consumo de recursos computacionales, para resolver este problema del ajuste fino tradicional existen técnicas de Ajuste Fino Eficiente en Parámetros (PEFT), entre ellas tenemos la Adaptación de Bajo Rango Cuantizada, la cual permite reducir el uso de memoria GPU mediante parámetros adaptativos, manteniendo la precisión del modelo a 4 bits (Dettmers et al., 2024). QLoRA se basa en dos mecanismos interdependientes:

Cuantización extrema (NF4): Esta abstracción comprime los pesos originales del modelo fundacional de un formato de coma flotante hacia un formato de 4 bits (NormalFloat 4), lo cual es teóricamente óptimo para pesos de distribución normal, ya que no se degrada la capacidad de razonamiento del modelo (Dettmers et al., 2024).

Adaptadores de Bajo Rango (LoRA): Durante la optimización, los parámetros originales cuantizados se mantienen sin modificación (Hu et al., 2022), en su lugar se agregan matrices de descomposición de bajo rango a los bloques de atención de la arquitectura, en este proceso el sistema solo actualiza los gradientes de los adaptadores auxiliares, lo cual reducen el número de parámetros a entrenar en más del 99% (Hu et al., 2022).

Retrieval-Augmented Forecasting (RAF)

Existe un desafío en los modelos de aprendizaje profundo avanzado, ya que en sistemas predictivos e investigaciones recientes se demuestra que pierden patrones en periodos de tiempo pasados porque se enfocan en periodos recientes, entonces para mitigar el problema de olvidar los datos históricos se tiene la técnica Retrieval-Augmented Forecasting (RAF), la cual permite buscar patrones similares en la historia que se asemejen al contexto de objetivo (Li, 2025).

El método RAF se basa en los sistemas de Retrieval-Augmented Generation (RAG) usados principalmente en procesamiento del lenguaje natural para mitigar las alucinaciones y asegurar que las respuestas estén acordes al contexto, mediante el uso de bases de conocimiento. Entonces de forma similar, RAF utiliza esta lógica y las aplica sobre datos estructurales y secuenciales como las series de tiempo (Li, 2025).

Esta técnica es usada principalmente en la mitigación de volatilidad climática porque ayuda a identificar y usar eventos pasados, como sequías, olas de calor, para evitar la volatilidad de este tipo de datos para obtener modelos más precisos. Otra aplicación es a nivel algorítmico, mediante mecanismos de hashing dinámico que alinean las variables meteorológicas sin

dependen del etiquetado, esto ayuda al modelo a interpretar el contexto de los datos a pesar de que el nombrado de las variables cambien (Li, 2025).

Métricas de Validación de Resultados

Para evaluar el desempeño de los modelos de predicción espaciotemporal propuestos en este estudio, se seleccionaron métricas estadísticas que permiten cuantificar tanto la magnitud del error como la capacidad de ajuste de los modelos.

El **Error Absoluto Medio (MAE)** mide el promedio de las diferencias absolutas entre los valores observados y las estimaciones del modelo. Su utilidad radica en que se interpreta directamente en las mismas unidades de la variable climática analizada, ofreciendo una visión clara de la precisión lineal sin amplificar el impacto de valores atípicos (Tang et al., 2025).

La **Raíz del Error Cuadrático Medio (RMSE)** constituye otra medida fundamental de precisión, ya que penaliza con mayor rigor las desviaciones de gran magnitud al elevar los errores al cuadrado antes de promediarlos. En variables meteorológicas como temperatura o precipitación, el RMSE permite identificar errores significativos que podrían comprometer la utilidad de los pronósticos en la toma de decisiones críticas (Ma et al., 2023).

El **Coefficiente de Determinación (R^2)** se utiliza para medir la capacidad del modelo de explicar la varianza de los datos reales, indicando qué tan bien las predicciones reflejan la tendencia de la serie histórica y el ajuste global del sistema (Ma et al., 2023; Zhu et al., 2021).

En el modelado de variables atmosféricas complejas, esta métrica resulta importante para determinar la proporción de la varianza observada que el sistema logra explicar a partir de las

dependencias espaciotemporales aprendidas (Nurcahyo & Putranto, 2026; Tang et al., 2025). En contextos de pronóstico regional, se considera que valores de R^2 superiores a 0.80 reflejan una alta capacidad predictiva, mientras que rangos entre 0.60 y 0.80 se catalogan como niveles de ajuste moderados o aceptables (Nurcahyo & Putranto, 2026).

El **Error Porcentual Absoluto Medio (MAPE)** evalúa la exactitud del modelo en términos relativos, expresando el error como porcentaje del valor real observado. Es especialmente útil para comparar el rendimiento en distintas escalas temporales o espaciales, aunque debe aplicarse con cautela en variables cercanas a cero, donde pequeñas desviaciones pueden generar errores porcentuales desproporcionados (Ali et al., 2025).

La prueba estadística de **Diebold-Mariano (DM)** se aplica para validar formalmente si la superioridad observada en el desempeño de un modelo sobre otro es estadísticamente significativa. A diferencia de la simple comparación de métricas agregadas, esta prueba contrasta la hipótesis nula de que ambos sistemas de pronóstico poseen la misma exactitud predictiva (Campbell & Diebold, 2005).

Todas las métricas se calculan sobre valores desescalados a grados Celsius, tanto a nivel global como por estación, como se muestra en la Tabla 2.

Tabla 2*Métricas de Evaluación*

Métrica	Fórmula	Nota
MAE	$\frac{1}{n} \sum \hat{y} - y $	Error medio en °C.
RMSE	$\sqrt{\frac{1}{n} \sum (\hat{y} - y)^2}$	Penaliza errores grandes.
R^2	$1 - \frac{\sum (\hat{y} - y)^2}{\sum (y - \bar{y})^2}$	Varianza explicada.
MAPE	$\frac{100}{n} \sum \frac{ \hat{y} - y }{\max(y , \varepsilon)}$	$\varepsilon = 10^{-2}$; cauteloso cerca de 0°C.

Nota. Elaboración propia basada en las fórmulas de Ali et al. (2025), Ma et al. (2023) y Tang et al. (2025). Se ajustó la fórmula de MAPE con un factor de estabilidad ε para el manejo de valores térmicos cercanos al punto de congelación.

CAPÍTULO III

Descripción del Conjunto de Datos y Variables

Los datos de este estudio son públicos y oficiales, obtenidos del portal de datos abiertos de la Red de Monitoreo de Calidad del Aire del Distrito Metropolitano de Quito (REMMAQ). La elección de este conjunto se justifica por su alta confiabilidad y por la disponibilidad de registros diarios y horarios durante aproximadamente 22 años. Esta amplitud permite aplicar metodologías de aprendizaje profundo, garantizando que el modelo capture la variabilidad existente, ya que los datos han sido validados oficialmente para el estudio de microclimas en entornos complejos (Secretaría de Ambiente del Distrito Metropolitano de Quito, 2026).

El dataset contiene mediciones meteorológicas y de contaminantes ambientales, organizadas en series temporales horarias con valores numéricos. Las variables disponibles en los datos abiertos se describen en la Tabla 3.

Tabla 3

Variables del Dataset Meteorológico y de Calidad del Aire de la REMMAQ

Categoría	Variable	Identificador	Unidad de Medida
Meteorológica	Temperatura media	TMP	Grados Celsius (°C)
Meteorológica	Precipitaciones	LLU	Milímetros (mm)
Meteorológica	Humedad relativa	HUM	Porcentaje (%)
Meteorológica	Presión barométrica	PRE	Millibares (mb)
Meteorológica	Radiación solar	RS	Watts por metro cuadrado (W/m ²)

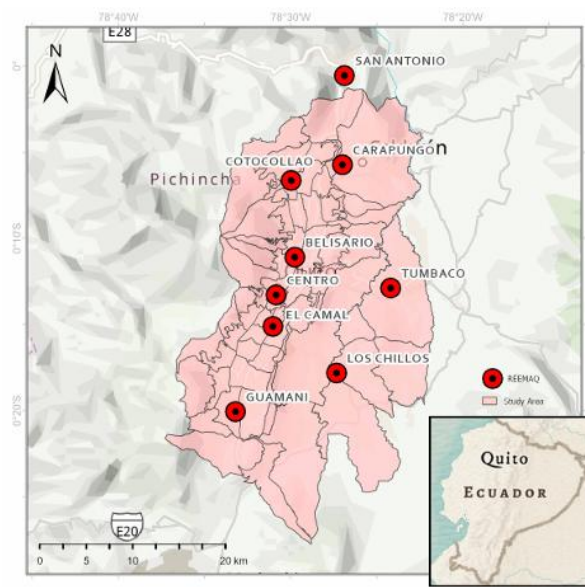
Meteorológica	Radiación ultravioleta	UV	Índice UV Solar Mundial (IUV)
Meteorológica	Velocidad del viento	VEL	Metros por segundo (m/s)
Meteorológica	Dirección del viento	DIR	Grados (°)
Calidad del Aire	Monóxido de carbono	CO	Miligramos por metro cúbico (mg/m ³)
Calidad del Aire	Dióxido de nitrógeno	NO2	Microgramos por metro cúbico (µg/m ³)
Calidad del Aire	Ozono	O3	Microgramos por metro cúbico (µg/m ³)
Calidad del Aire	Dióxido de azufre	SO2	Microgramos por metro cúbico (µg/m ³)
Calidad del Aire	Material particulado suspendido menor a 2.5 µg	PM2.5	Microgramos por metro cúbico (µg/m ³)
Calidad del Aire	Material particulado suspendido menor a 10 µg	PM10	Microgramos por metro cúbico (µg/m ³)

Nota. Elaboración propia a partir de los datos abiertos de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ, 2024).

La base incluye once estaciones de monitoreo en el Distrito Metropolitano de Quito: Belisario, Carapungo, Centro, Condado, Cotocollao, El Camal, Guamaní, Jipijapa, Los Chillos, San Antonio y Tumbaco. El conjunto reúne 1.643.527 observaciones registradas entre 2004 y 2026. La distribución espacial de las estaciones a lo largo de la ciudad se aprecia en la Figura 4.

Figura 4

Ubicación del Área de Estudio en la Zona Urbana de Quito, Ecuador



Nota. Tomado de “Machine Learning for Urban Air Quality Prediction Using Google AlphaEarth Foundations Satellite Embeddings: A Case Study of Quito, Ecuador”, por C. I. Alvarez, C. A. Ulloa Vaca y N. A. Echeverria Llumipanta, 2025, Remote Sensing, 17(20), 3472, p. 4 (<https://doi.org/10.3390/rs17203472>).

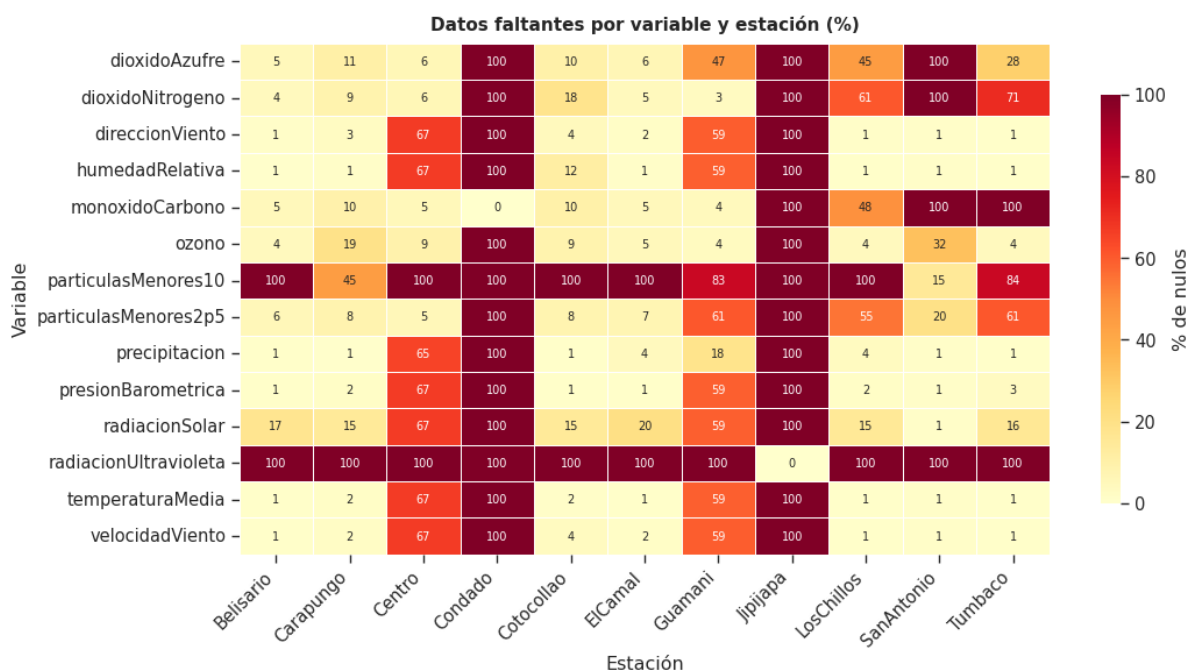
Análisis Exploratorio de Datos (EDA)

La fase inicial del análisis exploratorio se centra en garantizar la integridad estructural y la coherencia física de las series temporales. Siguiendo los protocolos de aseguramiento de calidad, se prioriza la eliminación de ruidos instrumentales y la corrección de anomalías que podrían afectar la capacidad predictiva de los algoritmos (Secretaría de Ambiente del Distrito Metropolitano de Quito, 2024).

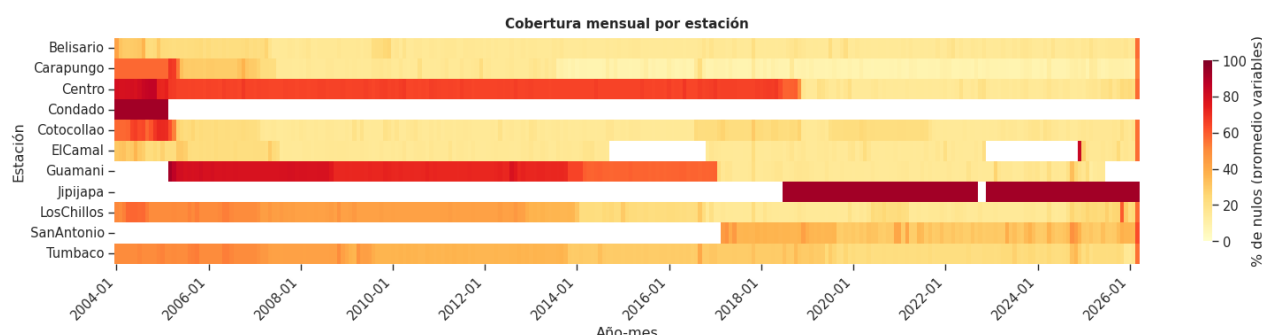
La Figura 5 muestra un 100% de valores faltantes en las estaciones del Condado y Jipijapa, salvo el monóxido de carbono en el primero y la radiación ultravioleta en el segundo. En las estaciones Centro y Guamaní, la ausencia supera el 59% en la mayoría de las mediciones. Además, variables como la radiación ultravioleta y las partículas suspendidas menores a 10 microgramos presentan limitaciones de cobertura. De manera similar, el dióxido de azufre, dióxido de nitrógeno y monóxido de carbono muestran más del 45% de datos faltantes en Los Chillos, San Antonio y Tumbaco. La Figura 6 complementa este análisis al mostrar los períodos con ausencia de datos por estación, destacando una alta concentración antes de 2008.

Figura 5

Datos Faltantes por Variable y por Estación



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Figura 6*Períodos con Datos Faltantes por Estación*

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Con base en estos resultados y siguiendo los lineamientos de la U.S. Environmental Protection Agency (2017), se adopta un umbral mínimo del 75% de captura de datos válidos para incluir estaciones o variables en el análisis final. Este criterio coincide con investigaciones recientes sobre la red REMMAQ (Mancheno & Jorquera, 2025). Aunque existen métodos avanzados de imputación, la literatura advierte que el análisis estadístico puede sesgarse significativamente cuando la ausencia supera el 10% (Bennett, 2001). Por ello, para evitar ruido excesivo en el pipeline de aprendizaje, se descartan dimensiones con más del 40-50% de datos faltantes.

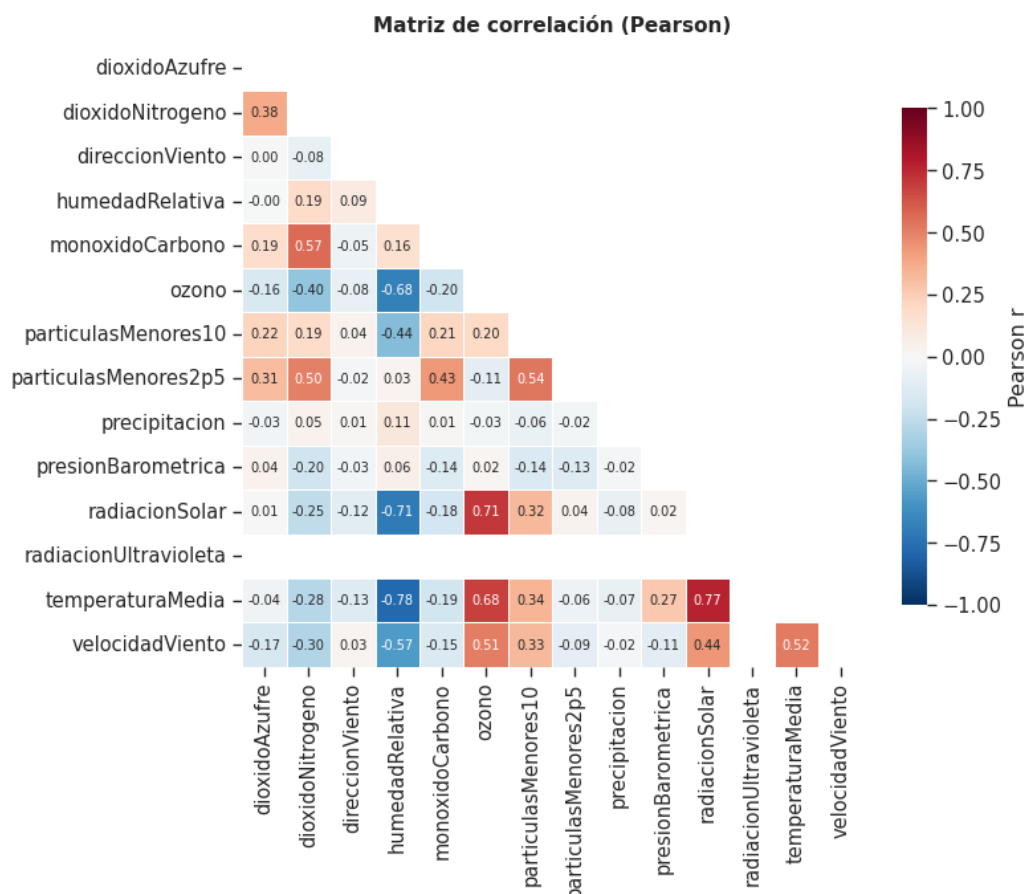
Análisis Multivariante

El análisis multivariado permite tener una visión más profunda del sistema atmosférico, donde las variables no actúan de forma aislada, sino que presentan interdependencias físicas y dinámicas (Tang et al., 2025). La Figura 7 muestra la matriz de coeficientes de Pearson, que cuantifica la fuerza y dirección de las relaciones lineales entre las condiciones meteorológicas.

En este contexto, la temperatura media presenta una alta correlación positiva con la radiación solar ($r=0.77$) y el ozono ($r=0.68$), mientras que mantiene una correlación negativa con la humedad relativa ($r=-0.78$). En contraste, su relación con las partículas suspendidas menores a 2.5 y con el dióxido de azufre es baja, lo que evidencia una menor dependencia con estos contaminantes.

Figura 7

Matriz de Correlación de Variables



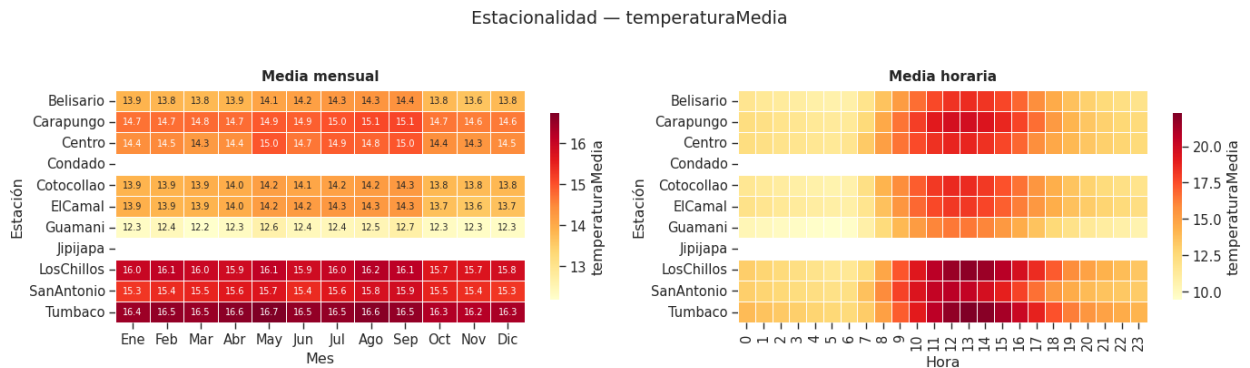
Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Análisis de Dinámicas Temporales

El análisis de las dinámicas temporales permite identificar patrones de recurrencia y la persistencia de los datos meteorológicos, lo cual resulta fundamental para configurar el horizonte histórico de entrada de los modelos.

Estacionalidad y ciclo diurno de la temperatura. En el Distrito Metropolitano de Quito, la temperatura media presenta una estacionalidad atípica en comparación con latitudes templadas (Álvarez et al., 2025). Mientras la variación mensual es casi plana debido a su ubicación ecuatorial, el hallazgo de un ciclo marcado, como se observa en la Figura 8, confirma que la variabilidad térmica es principalmente diurna. Esta característica climática justifica que los modelos no requieran capturar tendencias estacionales de largo plazo (meses o años), sino que deben priorizar el aprendizaje de patrones dentro de una ventana de 24 horas (Ma et al., 2023).

Figura 8
Estacionalidad Mensual y Diaria



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

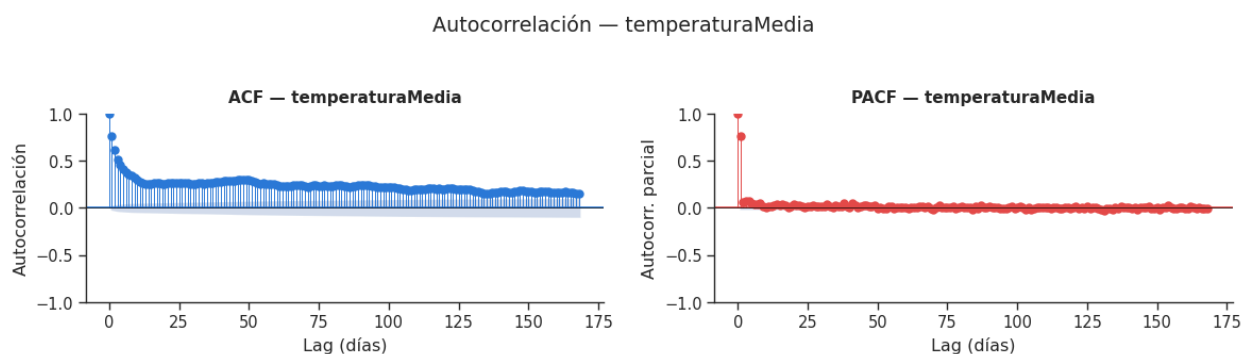
Análisis de autocorrelación y diseño de ventanas. El estudio de las funciones de autocorrelación permite establecer la profundidad de la memoria histórica que deben procesar las redes neuronales (Guerrero Muñoz & Armijos Toro, 2025).

Persistencia de la señal (ACF): La Figura 9, correspondiente a la temperatura media, muestra una correlación que inicia en 1.0 y decae gradualmente, manteniéndose positiva durante un horizonte prolongado. Este comportamiento confirma que la temperatura no es un proceso aleatorio, pues las condiciones térmicas de horas o días previos influyen en los estados actuales.

Dependencia de largo plazo (PACF): De manera complementaria, el gráfico de PACF en la Figura 9 evidencia un pico en los dos primeros rezagos, seguido de un decaimiento abrupto hacia valores cercanos a cero. Esto indica que los valores inmediatamente anteriores explican la mayor parte de la varianza del tiempo actual.

Figura 9

Funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) para la Temperatura Media



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Resumen de Decisiones Sobre los Datos

Con el objetivo de garantizar la calidad y consistencia de la información utilizada en el estudio, se definieron criterios específicos para la depuración de la base de datos. Estos criterios resumen las principales decisiones adoptadas durante el proceso de preparación. A continuación, se presenta la Tabla 4 con el detalle de dichas decisiones.

Tabla 4

Resumen de Decisiones Técnicas y Configuración Final del Dataset

Fase del EDA	Decisión Aplicada	Justificación
Integridad Estructural	Exclusión de estaciones Condado, Jipijapa, Centro, Guamaní y San Antonio.	Incumplimiento del umbral de datos faltantes.
Cobertura Temporal	Selección del rango cronológico [2008-2026].	Asegura que todos los nodos disponibles mantengan uniformidad en los datos.
Análisis Multivariante	Exclusión de las variables de radiación ultravioleta, dióxido de azufre, partículas menores a 2.5 y 10 micrómetros, dióxido de nitrógeno y monóxido de carbono.	Presentan baja relación física con la temperatura media y un alto nivel de datos faltantes en la estación de Los Chillos. Sin embargo, esta estación resulta necesaria en el grafo para capturar la influencia térmica en los valles.
Diseño de ventana para corto plazo	Fijación de la longitud de la serie temporal histórica a 24 horas.	El análisis de PACF y el ciclo diurno marcado justifican el horizonte de un día completo.
Diseño de ventana para medio plazo	Fijación de la longitud de la serie temporal histórica a 96 horas.	Captura la persistencia de largo alcance identificada en la ACF.

Diseño de ventana para largo plazo	Fijación de la longitud de la serie temporal histórica a 168 horas.	Captura el ciclo completo de 7 días, permitiendo al mecanismo de atención global ponderar estados históricos similares (Li et al., 2018).
------------------------------------	---	---

Nota. Elaboración propia a partir de los criterios de calidad de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ, 2024) y de la U.S. Environmental Protection Agency (2017).

Preprocesamiento de Datos

Para el análisis y pronóstico de series temporales mediante soluciones como ARIMA, STGNN y Modelos Fundacionales, se requiere que las observaciones de entrada mantengan intervalos regulares en el tiempo (Fawaz et al., 2019; Goodfellow et al., 2016). Al trabajar con datos de los sensores de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ), se deben considerar los fallos y cortes de energía, ya que pueden ocasionar series con valores faltantes o anómalos (Van Buuren, 2018). Estos problemas de asimetría y ruptura inducen a que el modelo asuma que dos horas separadas estén consecutivas tras un apagón; por lo tanto, es necesario aplicar técnicas que permitan reconstruir la sincronía horaria y asegurar la consistencia de las series temporales en este estudio.

Alineación Temporal y Consistencia Estructural

Como primer paso para abordar los problemas de continuidad y consistencia en las series temporales ocasionados por fallos en los sensores, se completa el esquema temporal con el fin de garantizar la integridad horaria de cada serie. Este procedimiento se desarrolla mediante los siguientes pasos:

1. Generación del vector temporal: A partir del EDA se identificó que las series temporales con frecuencia horaria inician en enero de 2008 y se extienden hasta el 31 de marzo de 2026, lo que asegura la inclusión de variables críticas como la radiación solar.
2. Producto cartesiano: Para evitar vacíos en las series temporales, se construyó el cruce entre el vector de tiempo (fechas y horas) y las estaciones mediante producto cartesiano, garantizando que cada estación mantenga la misma cronología.
3. Acoplamiento de series temporales: El esqueleto temporal se combinó con los datos reales de los sensores. De esta manera, las horas donde las estaciones no registraron información, por fallos de energía o transmisión, aparecen en la matriz como valores nulos, lo que permite cuantificar los vacíos y analizar técnicas de imputación.
4. Optimización computacional: Para mejorar la eficiencia del conjunto de datos, se redujo la precisión numérica de 64 bits a 32 (float32) y las variables tipo texto, como los nombres de estaciones, se transformaron en categóricas. Esta adaptación permite estandarizar los datos de entrada en tensores multidimensionales y favorecer su procesamiento numérico durante el entrenamiento en unidades de procesamiento gráfico (GPU) (Micikevicius et al., 2018).

Una vez completados los pasos descritos, el rango temporal considerado para el estudio abarca desde el 1 de enero de 2008 a las 00:00:00 hasta el 31 de marzo de 2026 a las 23:00:00. El tratamiento aplicado genera una base optimizada y consistente, tal como se presenta en la Tabla

Tabla 5*Proceso de Alineación Temporal e Integridad Estructural*

Parámetro / Métrica Estructural	Matriz Cruda (Antes)	Matriz Densa (Después)	Diferencia / Impacto Técnico
Registros temporales (Filas)	1,125,813	959,760	−166,053 filas (poda por fecha de corte y regularización)
Continuidad cronológica ($\Delta t = 1$ h)	Asimétrica / Con saltos	Equidistante estricta	100% de cobertura teórica por producto cartesiano
Valores nulos en Temperatura (NaN)	13,492	51,430	+37,938 fallos eléctricos expuestos

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Control de Calidad y Filtros Físicos (QA/QC)

Con las series temporales completas se asegura su calidad, limitando el tratamiento de valores atípicos únicamente a aquellos por fallos de hardware que afectaron la obtención de los datos en los sensores, los cuáles no cumplen con las leyes de la termodinámica y deben ser tratados antes de cualquier procesamiento posterior. (Secretaría de Ambiente, 2025). Por otro lado, los valores atípicos producto de eventos climáticos anormales no se tratan en este punto debido a que los modelos de aprendizaje profundo pueden detectar esos patrones para mejorar su capacidad de predicción.

Para evitar la contaminación de las variables se aplicaron filtros basados en principios de la física atmosférica, alineados con las directrices de la Organización Meteorológica Mundial

(WMO-GAW) y adaptados al contexto geográfico del Distrito Metropolitano de Quito. El procedimiento incluyó los siguientes pasos:

1. Imputación histórica: La variable de ozono, con más del 80% de valores faltantes, fue eliminada. Como se evidenció en el EDA, no existen registros históricos suficientes, y su imputación por más de diez años generaría covarianzas ficticias frente a las demás variables.
2. Filtro barométrico: Dado que el Distrito Metropolitano de Quito se ubica a 2800 m.s.n.m., la presión atmosférica varía entre 715 y 740 hPa (Secretaría de Ambiente, 2025). Lecturas fuera de este rango corresponde a errores de sensor y se transformaron en valores nulos, obteniéndose 93 registros depurados.
3. Corrección de ceros diurnos: Considerando la latitud 0° de la ciudad y sus altos niveles de radiación (Secretaría de Ambiente, 2024), valores cercanos a cero entre las 09:00 y las 15:00 horas son físicamente imposibles. Estos registros se reemplazaron por nulos, afectando a 1,180 observaciones.
4. Rango de irradiancia máxima: El rango de radiación solar en el Distrito Metropolitano de Quito es de $1,400 \text{ W/m}^2$ (Secretaría de Ambiente, 2025). Aunque no se detectaron valores que superen este umbral, el criterio se mantuvo como parte de la lógica de control físico.

Estos criterios garantizan que el conjunto de datos resultante esté libre de valores derivados de fallos instrumentales que contradigan leyes físicas, manteniendo así la variabilidad atmosférica real de la ciudad (Secretaría de Ambiente, 2025).

Estrategia de Particionamiento de Datos

La fase de particionamiento de los datos es parte fundamental para evaluar la capacidad de generalización de los modelos. Debido a la dependencia temporal de las series meteorológicas, no se aplicó una partición aleatoria, ya que esto introduciría fuga de información (*data leakage*) al agregar registros futuros durante el entrenamiento. En su lugar, se implementó un particionamiento cronológico (*Out-of-Time Validation*), manteniendo el orden temporal de las observaciones (Hyndman & Athanasopoulos, 2021).

Los puntos de corte se definieron considerando dos criterios. En primer lugar, se garantizó que los conjuntos de validación y prueba tuvieran ciclos completos, para representar adecuadamente el comportamiento climático de Quito. En segundo lugar, se buscó mantener una distribución aproximada del 80 % para entrenamiento, 10 % para validación y 10 % para prueba sobre toda la serie temporal.

Bajo este esquema, el conjunto de entrenamiento comprende los registros entre el 1 de enero de 2008 y el 31 de diciembre de 2021, utilizados para el aprendizaje de los modelos. El conjunto de validación abarca del 1 de enero de 2022 al 31 de diciembre de 2023, y se empleó de manera interna para monitorear el desempeño durante la fase de entrenamiento, optimizar hiperparámetros y aplicar criterios de parada temprana (Early Stopping). Finalmente, el conjunto de prueba comprende del 1 de enero de 2024 al 31 de marzo de 2026, permaneciendo completamente aislado durante todas las fases previas de modelado. Este conjunto se utilizó exclusivamente en la etapa final para realizar la evaluación objetiva sobre datos no observados y

seleccionar las arquitecturas con mejor capacidad de generalización, tal como se aprecia en la Tabla 6.

Tabla 6

Estructura Cronológica del Particionamiento de Datos

Partición	Rango temporal	Proporción temporal	Registros
Entrenamiento	01/01/2008 – 31/12/2021	76.71 %	736,224
Validación	01/01/2022 – 31/12/2023	10.96 %	105,120
Prueba	01/01/2024 – 31/03/2026	12.33 %	118,416
Total	01/01/2008 – 31/03/2026	100.00 %	959,760

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

El particionamiento cronológico se realizó antes de cualquier transformación estadística, incluyendo la imputación y el escalado de las variables. De esta forma, los parámetros de cada procedimiento se ajustaron únicamente con el conjunto de entrenamiento, evitando la fuga de información (*data leakage*) hacia los conjuntos de validación y prueba.

Tratamiento de Datos Faltantes

Al lidiar los datos erróneos generados por fallos de hardware, estos se transformaron en valores nulos que debían ser tratados. Se descartaron métodos de imputación univariante, como la media aritmética o el algoritmo de K Vecinos Cercanos (KNN), ya que no permiten conservar las distribuciones reales en los datos y tienden a aplanar la varianza, además de ignorar la topología de la región analizada (Van Buuren, 2018).

En su lugar, se implementó un método de imputación fundamentado en la termodinámica de las variables meteorológicas, estructurado en cuatro etapas:

1. Imputación determinista: Se aplicó a la variable de radiación solar, cuya dinámica responde más a la mecánica orbital de la Tierra que a una distribución estocástica (WMO, 2014). Para su imputación se utilizó la mediana histórica agrupada por estación, mes y hora, diferenciando entre horas diurnas y nocturnas, lo que evita errores en los valores de irradiancia.
2. Descomposición vectorial del viento: Se abordó la paradoja de la medianoche geográfica (Grange, 2014), dado que la dirección del viento se registra en escala polar. Una oscilación entre 358° y 2° podría interpretarse erróneamente como 180° (dirección sur), generando una inversión hacia el norte (WMO, 2011). Para evitar esta limitación, la variable se transforma en componentes cartesianas ortogonales, permitiendo el uso de regresores espaciales lineales. Bajo este esquema, U indica la fuerza Este-Oeste y V la fuerza Norte-Sur siguiendo las siguientes fórmulas:

$$U = velocidad \cdot \cos\left(direccion \cdot \frac{\pi}{180}\right) \quad (1)$$

$$V = velocidad \cdot \sin\left(direccion \cdot \frac{\pi}{180}\right) \quad (2)$$

Donde:

- U: componente zonal del viento (Este–Oeste)
- V: componente meridional del viento (Norte–Sur)
- velocidad: velocidad del viento (m/s)

- dirección: dirección del viento (grados)
3. Imputación topológica: Se aplicó sobre las variables de temperatura, humedad relativa, presión barométrica y precipitación mediante un pivoteo donde las estaciones meteorológicas pasaron a representar las columnas de la matriz. Posteriormente, se utilizó un estimador bayesiano regularizado (*IterativeImputer con Bayesian Ridge*) para imputar los valores faltantes a partir de la covarianza lineal existente entre las estaciones en un mismo instante temporal (Pedregosa et al., 2011; Van Buuren, 2018). Dado que el particionamiento cronológico se realizó antes, el modelo de imputación se ajustó únicamente con el conjunto de entrenamiento y posteriormente se aplicó a los conjuntos de validación y prueba, evitando la fuga de información (*Data Leakage*). Para garantizar la coherencia, se tomaron en cuenta las normas atmosféricas locales, donde se indica que la humedad debe estar en el rango [0%, 100%] y la precipitación se limita a valores no negativos (Secretaría de Ambiente, 2025).
 4. Interpolación lineal: Finalmente, los valores nulos restantes se trataron mediante interpolación lineal considerando intervalos de tres horas contiguas.

Como resultado, se obtuvo una base de datos sin valores nulos. La mayor incidencia correspondió a la imputación topológica, aplicada antes de la interpolación lineal. Debido a que el estimador bayesiano se ajustó exclusivamente con el conjunto de entrenamiento para evitar fuga de información, en los periodos de validación y prueba únicamente se aplicó la transformación previamente aprendida. Cuando se presentaron valores faltantes de forma

simultánea en todas las estaciones, estos no pudieron ser tratados mediante el imputador y pasaron a la etapa de interpolación lineal, con un máximo de 252 registros por variable. No obstante, el total imputado mediante esta técnica representó menos del 0.16 % de la muestra, por lo que su incidencia sobre la distribución general de los datos fue mínima, como se muestra en la Tabla 7.

Tabla 7

Proceso para la Imputación de Valores Faltantes

Etapas	Valores reconstruidos
Imputación determinista	Radiación solar: 57,396 valores nulos imputados mediante astronomía.
Imputación topológica	Temperatura media: 51,190 valores. Humedad relativa: 71,337 valores. Presión barométrica: 55,383 valores. Precipitación: 52,037 valores. Viento u: 56,324 valores. Viento v: 56,324 valores.
Interpolación lineal	Humedad relativa: 240 valores. Temperatura media: 240 valores. Precipitación: 216 valores. Presión barométrica: 246 valores. Viento u: 252 valores. Viento v: 252 valores.

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Ingeniería de Características (Feature Engineering)

En esta fase se proporciona a los modelos un espacio de características que optimiza la extracción de patrones y facilita la convergencia de las soluciones (Goodfellow et al., 2016). Los procesos aplicados se detallan a continuación:

1. Codificación del tiempo: Dado que las variables temporales presentan un comportamiento cíclico, por ejemplo, pasar de las 23:00 a las 00:00 matemáticamente implica una reducción de 23, cuando en realidad es el fin e inicio de un ciclo. Por este motivo, se trataron las variables temporales de hora, mes y día de la semana, las cuales se proyectaron sobre un círculo unitario mediante funciones seno y coseno (Fawaz et al., 2019).
2. Homologación de variables físicas: Tras el análisis exploratorio, se decidió no aplicar transformaciones matemáticas adicionales a las variables climáticas principales, como la transformación logarítmica de la precipitación o la generación de variables derivadas y retardos temporales. Las variables se mantuvieron en su representación original para homologar el espacio de entrada utilizado por los modelos fundacionales y evitar modificaciones diferenciadas durante el preprocesamiento. De esta manera, las diferencias observadas en el desempeño predictivo responden principalmente a las capacidades de cada arquitectura y no a un diseño de características específico.

Como resultado del tratamiento aplicado en las cuatro fases, el esquema final que recibieron los modelos corresponde a una Matriz de Datos de Trabajo (MDT) con 959,760 filas y 20 columnas. De este total, 13 corresponden a características predictivas (*features*) por estación,

conformadas por 7 variables climáticas puras y 6 variables cíclicas. Las 7 columnas restantes corresponden a metadatos utilizados para la identificación temporal y espacial de las observaciones. La estructura de la MDT se detalla en la Tabla 8.

Tabla 8

Conjunto de Variables en la MDT

Variable	Tipo de Dato	Variable	Tipo de Dato
fechaHora	datetime64[ns]	anio	int32
estacion	category	dia_semana	int32
fecha	object	viento_u	float32
hora	int32	viento_v	float32
humedadRelativa	float32	hora_sin	float32
precipitacion	float32	hora_cos	float32
presionBarometrica	float32	mes_sin	float32
radiacionSolar	float32	mes_cos	float32
temperaturaMedia	float32	dia_semana_sin	float32
mes	int32	dia_semana_cos	float32

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Implementación de Algoritmos Tradicionales

Con el objetivo de proporcionar una referencia estadística que permitiera una evaluación objetiva del rendimiento de las arquitecturas competidoras propuestas en este estudio, se implementó un conjunto de modelos pertenecientes a la familia ARIMA. Estos modelos se destacan como una de las metodologías más ampliamente utilizadas en el análisis y pronóstico de

series temporales, debido a su capacidad para representar la dependencia temporal de una variable a través de componentes autorregresivos, de diferenciación, de media móvil, estacionales y, en sus variantes extendidas, a través de la integración de variables exógenas. En el marco de este estudio, los modelos ARIMA se interpretaron como la línea base sobre la cual se comparó posteriormente el rendimiento de las arquitecturas competidoras, facilitando así la formulación de una referencia objetiva respecto a las mejoras alcanzadas por los modelos propuestos.

La aplicación de estos modelos no se limitó al entrenamiento de una configuración específica, sino que empleó un proceso metodológico estructurado que permitiera una evaluación objetiva del comportamiento de las diferentes variantes bajo condiciones experimentales similares. Para ello, se diseñó un protocolo que comprende diversas etapas: elaboración del conjunto experimental, optimización de hiperparámetros, validación mediante un esquema de predicción secuencial y análisis comparativo de los resultados obtenidos. Esta metodología asegura que todas las configuraciones sean evaluadas siguiendo un protocolo experimental uniforme, lo que permite atribuir las diferencias detectadas exclusivamente a las características inherentes de cada modelo, evitando cualquier variación en el procedimiento de entrenamiento o evaluación.

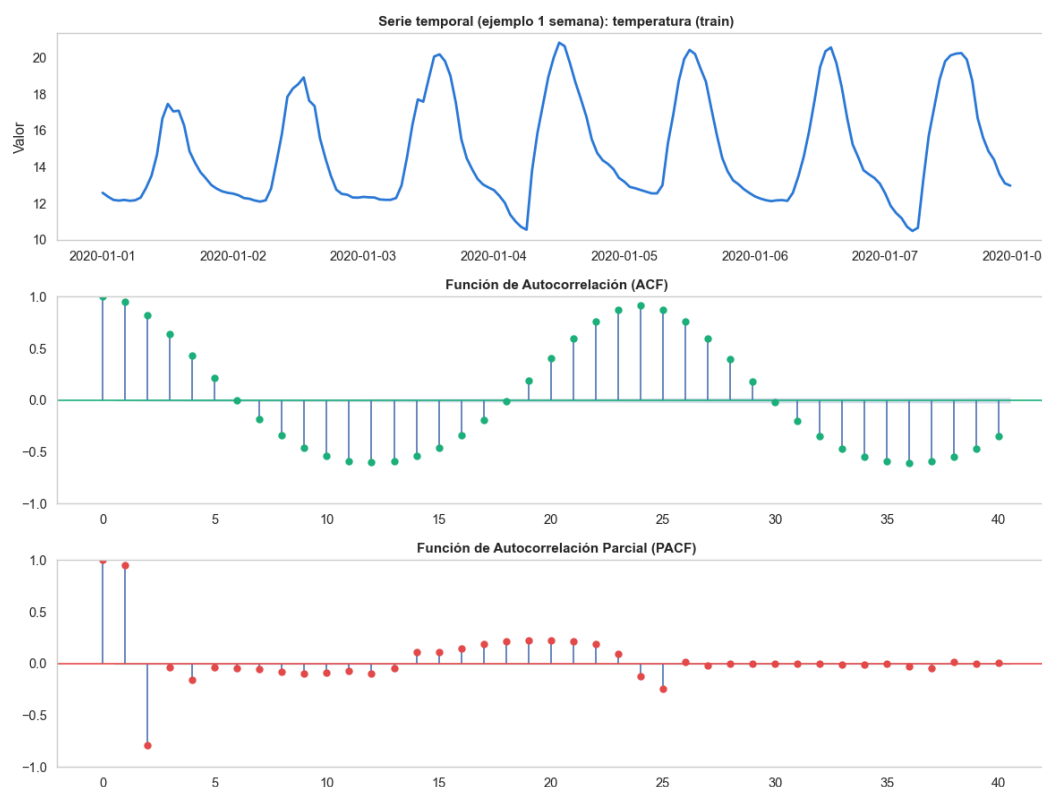
Debido a que la finalidad de este trabajo es la predicción de la temperatura para horizontes de 3, 48 y 72 horas, la implementación de modelos con variables exógenas se realizó de manera independiente para cada horizonte. Se elaboraron conjuntos específicos de variables explicativas, alineados temporalmente con cada horizonte, garantizando que durante el proceso

de entrenamiento y la generación de predicciones únicamente se dispusiera de la información necesaria en cada escenario.

Previo al proceso de entrenamiento, se verificó uno de los supuestos fundamentales de los modelos ARIMA: la estacionariedad de la serie temporal. Para ello, se implementó la prueba de Dickey-Fuller Aumentada (ADF), cuya hipótesis nula postula la presencia de una raíz unitaria y, por ende, la no estacionariedad de la serie en cuestión. Adicionalmente, se realizaron análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), las cuales facilitaron la identificación de la persistencia temporal de la serie y ofrecieron una aproximación inicial sobre el comportamiento de los componentes autorregresivos y las medias móviles. Los resultados de estas pruebas se ilustran en la Figura 10.

Figura 10

Funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) para la Temperatura Media



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Esta metodología implicó la implementación de cuatro familias de modelos estadísticos, compuestas por las variantes ARIMA, SARIMA, ARIMAX y SARIMAX. Los modelos ARIMA y SARIMA fueron entrenados a partir de la serie histórica de temperatura, mientras que las variantes ARIMAX y SARIMAX integraron las variables exógenas establecidas durante el proceso de ingeniería de características para cada horizonte de predicción.

Con el objetivo de examinar el impacto de la optimización en el rendimiento predictivo, cada una de estas configuraciones fue evaluada inicialmente a través de un modelo de referencia y posteriormente mediante una versión optimizada derivada de un proceso sistemático de identificación de hiperparámetros. La Tabla 9 sintetiza las configuraciones adoptadas en esta fase de la investigación.

Tabla 9

Modelos implementados durante el proceso experimental

Familia	Modelo	Modelo optimizado
ARIMA	ARIMA	ARIMA (Tuning)
SARIMA	SARIMA	SARIMA (Tuning)
ARIMAX	H+3, H+48 y H+72	H+3, H+48 y H+72
SARIMAX	H+3, H+48 y H+72	H+3, H+48 y H+72

Nota. Elaboración propia.

Construcción del Conjunto Experimental de Datos

Una vez establecido el protocolo metodológico para la implementación de los modelos, se procedió a la construcción del conjunto experimental que serviría como base para el entrenamiento y evaluación de todas las configuraciones propuestas. La finalidad de esta fase fue la creación de una estructura de datos homogénea que asegurara la comparabilidad entre los diversos modelos, preservando las mismas segmentaciones temporales y los mismos criterios de evaluación a lo largo de todo el proceso experimental.

La creación de conjuntos independientes para cada horizonte representa una de las decisiones metodológicas más significativas de este estudio, dado que disminuye de manera

significativa el riesgo de fuga de datos durante el proceso de entrenamiento. En lugar de utilizar un conjunto único de variables exógenas para todas las instancias de predicción, se diseñaron estructuras específicas para todos los horizontes, asegurando que ninguna variable incorporara información derivada del futuro en relación con el instante de predicción.

Esta consideración adquiere especial relevancia en contextos meteorológicos, en los que mínimas discrepancias temporales pueden generar estimaciones excesivamente optimistas durante la evaluación y comprometer la capacidad de generalización del modelo.

Entrenamiento de Algoritmos Tradicionales

Con la intención de contar con una referencia estadística que permitiera evaluar de manera objetiva el desempeño de las arquitecturas competidoras propuestas en este estudio, se implementó un conjunto de modelos pertenecientes a la familia ARIMA. Son unos de los esquemas de trabajo más utilizados para analizar y predecir series temporales porque permiten captar la dependencia temporal de una variable mediante componentes autorregresivos, de diferenciación, de media móvil, estacionales y, en sus variantes más extendidas, la integración de variables exógenas. Dentro de este estudio, los modelos ARIMA se interpretaron como la línea base sobre la cual se comparó posteriormente el rendimiento de las arquitecturas competidoras, lo cual facilitó la formulación de una referencia objetiva respecto a las mejoras alcanzadas por los modelos propuestos.

Estos modelos no se aplicaron únicamente para entrenar una configuración específica, sino que se utilizó un proceso metodológico estructurado que permitiera evaluar de manera objetiva el comportamiento de las diferentes variantes en condiciones experimentales similares.

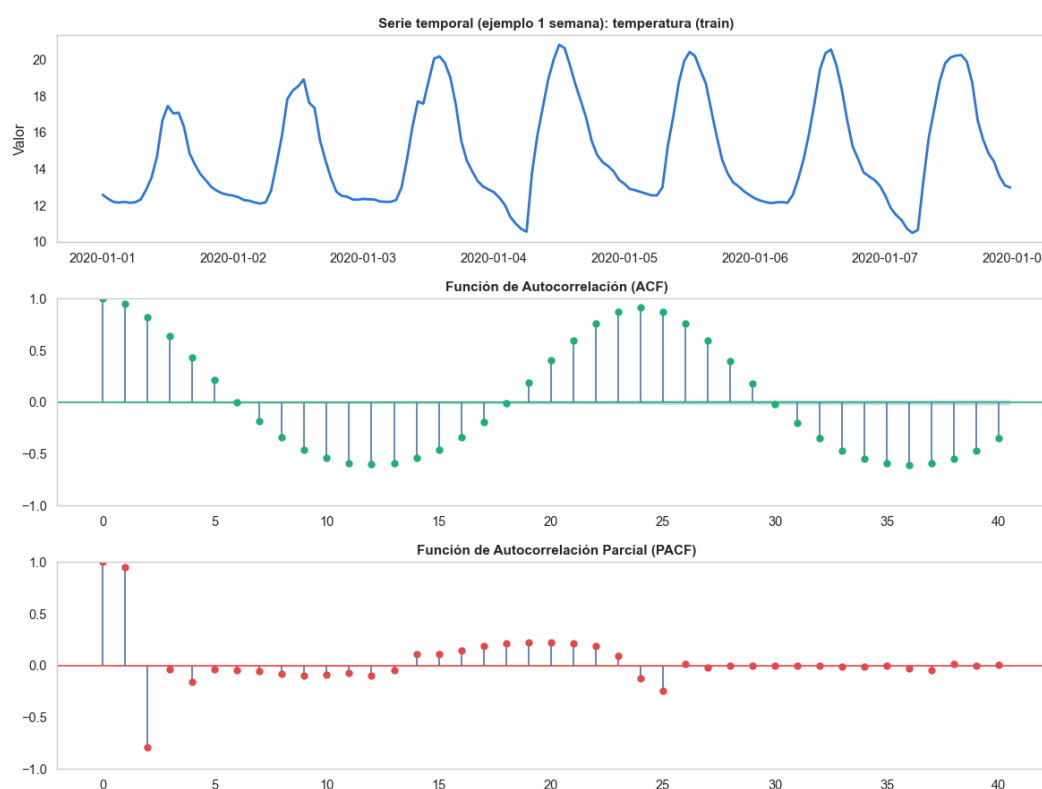
Se diseñó un protocolo para ello, el cual comprende varias etapas: la elaboración del conjunto experimental, la optimización de hiperparámetros, la validación por un esquema de predicción secuencial y el análisis comparativo de los resultados obtenidos. Esta metodología garantiza que todas las configuraciones sean probadas siguiendo un mismo protocolo experimental, por lo que las diferencias observadas pueden atribuirse únicamente a las características de los distintos modelos, descartando cualquier variabilidad en el procedimiento de entrenamiento o evaluación.

Dado que la predicción de temperatura para horizontes de 3, 48 y 72 horas es el objetivo de este trabajo, la implementación de modelos con variables exógenas se efectuó de manera independiente para cada horizonte. Se crearon conjuntos específicos de variables explicativas, temporales con cada horizonte, asegurando que durante el proceso de entrenamiento y la generación de predicciones sólo se tuviera la información necesaria en cada escenario.

Antes del proceso de entrenamiento se comprobó una de las hipótesis fundamentales de los modelos ARIMA, la estacionariedad de la serie temporal. Para tal fin se utilizó la prueba de Dickey-Fuller aumentada (ADF) cuya hipótesis nula supone la existencia de raíz unitaria y, por tanto, la no estacionariedad de la serie analizada. Además, se hicieron análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), que permitieron identificar la persistencia temporal de la serie y dieron una idea inicial del comportamiento de los componentes autorregresivos y medias móviles. Los resultados de estas pruebas se muestran en la Figura 11.

Figura 11

Funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) para la Temperatura Media



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Esta metodología implicó la implementación de cuatro familias de modelos estadísticos, integradas por las variantes ARIMA, SARIMA, ARIMAX y SARIMAX. Se entrenaron los modelos ARIMA y SARIMA con la serie histórica de temperatura, mientras que los ARIMAX y SARIMAX incorporaron las variables exógenas establecidas en el proceso de ingeniería de características para cada horizonte de predicción.

Para examinar el efecto de la optimización en el rendimiento predictivo, cada configuración se probó primero con un modelo de referencia y después con una versión optimizada, obtenida mediante un proceso sistemático de identificación de hiperparámetros. En la Tabla 10 se resumen las configuraciones utilizadas en esta etapa de la investigación.

Tabla 10

Modelos implementados durante el proceso experimental

Familia	Modelo	Modelo optimizado
ARIMA	ARIMA	ARIMA (Tuning)
SARIMA	SARIMA	SARIMA (Tuning)
ARIMAX	H+3, H+48 y H+72	H+3, H+48 y H+72
SARIMAX	H+3, H+48 y H+72	H+3, H+48 y H+72

Nota. Elaboración propia.

Construcción del Conjunto Experimental de Datos

Con el fin de proceder a la implementación de los modelos, se estableció un protocolo metodológico, construyéndose un conjunto experimental que se utilizaría como base para el entrenamiento y evaluación de todas las configuraciones propuestas. El objetivo de esta fase fue construir una estructura de datos homogénea que garantizara la comparabilidad entre los diferentes modelos, manteniendo las mismas segmentaciones temporales y los mismos criterios de evaluación a lo largo de todo el proceso experimental.

Una de las decisiones metodológicas más importantes adoptadas en este estudio ha sido la creación de conjuntos independientes para cada horizonte, ya que esto permite reducir en gran medida el riesgo de fuga de datos durante el entrenamiento. En vez de utilizar un conjunto único

de variables exógenas para todas las instancias de predicción, se diseñaron estructuras específicas para todos los horizontes, de manera que ninguna variable incorporase información del futuro respecto al instante de predicción.

Este aspecto resulta de especial importancia en contextos meteorológicos, en los que pequeñas diferencias temporales pueden dar lugar a estimaciones demasiado optimistas en la fase de evaluación, poniendo en riesgo la capacidad de generalización del modelo.

Entrenamiento de Algoritmos Tradicionales

Una vez finalizada la construcción del conjunto experimental, se procedió a entrenar los modelos tradicionales para establecer una referencia inicial del desempeño de cada una de las variantes de la familia ARIMA, previa a la implementación de la optimización de hiperparámetros. Esta fase permitió obtener un punto de comparación homogéneo para evaluar, posteriormente, el efecto de la identificación de hiperparámetros y de la integración de componentes estacionales. Todos los modelos se entrenaron con la misma partición de datos y bajo condiciones experimentales uniformes, asegurando que las variaciones observadas en la evaluación fueran exclusivamente atribuibles a la estructura de cada modelo.

El entrenamiento contempló un total de ocho configuraciones iniciales, repartidas entre los modelos univariantes ARIMA y SARIMA, y las variantes multivariadas ARIMAX y SARIMAX para los horizontes de predicción de 3, 48 y 72 horas respectivamente. Los modelos ARIMA y SARIMA únicamente emplearon la información histórica de la variable objetivo para la estimación de los valores futuros, mientras que los modelos ARIMAX y SARIMAX incorporaron las variables exógenas creadas específicamente para cada horizonte de predicción.

Esta distinción facilitó que cada configuración pudiera capturar las relaciones temporales propias del horizonte que debía estimar, evitando así la reutilización de estructuras de datos entre problemas de predicción con características distintas.

En esta fase inicial no se aplicó la optimización de hiperparámetros. Todos los modelos se entrenaron con una configuración inicial predefinida con la finalidad de establecer una línea representativa que sirviera de referencia en las etapas posteriores del experimento. En la Tabla 11 se resumen las configuraciones iniciales empleadas.

Tabla 11

Resultados de Entrenamiento en Enfoques Tradicionales

Modelo	Horizonte de predicción	Parámetros	RMSE (°C)
ARIMA	N/A	ARIMA(1, 0, 1)	3.16
SARIMA	N/A	SARIMAX(1,0,1)(1,0,1,24)	2.41
ARIMAX H+3	3 horas	ARIMA(1,0,1) + exógenas	2.62
ARIMAX H+48	48 horas	ARIMA(1,0,1) + exógenas	1.86
ARIMAX H+72	72 horas	ARIMA(1,0,1) + exógenas	1.93
SARIMAX H+3	3 horas	SARIMAX(1,0,1)(1,0,1,24) + exógenas	2.28
SARIMAX H+48	48 horas	SARIMAX(1,0,1)(1,0,1,24) + exógenas	2.76
SARIMAX H+72	72 horas	SARIMAX(1,0,1)(1,0,1,24) + exógenas	2.79

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

N/A: modelo univariado sin variables exógenas por horizonte; se evalúa con un único pronóstico continuo sobre todo el conjunto de validación.

Optimización de Hiperparámetros

Una vez finalizado el entrenamiento de los modelos de referencia, se implementó un procedimiento de optimización de hiperparámetros con la intención de mejorar la capacidad predictiva de cada una de las configuraciones implementadas. Si bien los modelos iniciales proporcionan una referencia adecuada para la evaluación del comportamiento global de la familia ARIMA, su desempeño depende de la elección de los parámetros que configuran su estructura. Si se mantiene una configuración constante para todos los escenarios de predicción, podría limitar la representación dinámica de la serie temporal, especialmente cuando hay patrones estacionales o dependencias temporales de diferentes escalas.

Para mitigar este efecto, se implementó un proceso sistemático de identificación de hiperparámetros orientado a encontrar la configuración que diera mejor desempeño en el conjunto de validación. Para cada modelo se emplearon diferentes combinaciones de parámetros autorregresivos, de diferenciación y de medias móviles, y en los modelos estacionales se incluyeron también los parámetros de la componente estacional, con el fin de capturar la periodicidad existente en la serie. La selección final se hizo teniendo en cuenta el desempeño en validación, procurando un buen equilibrio entre precisión y capacidad de generalización.

Debido al número de configuraciones evaluadas y el costo computacional asociado al entrenamiento de cada modelo, la búsqueda se realizó en paralelo, logrando una reducción importante del tiempo requerido para completar la exploración del espacio de hiperparámetros. Esta estrategia resultó especialmente importante para las variantes SARIMA y SARIMAX, que tienen mayor tiempo de implementación.

Los parámetros optimizados para cada una de las familias de modelos utilizados en el proceso de ajuste de los modelos fueron fijados en orden el máximo de autorregresión (p) = 3 y el orden máximo de medias móviles (q) = 3. Se estableció el parámetro de diferenciación d en cero ($d = 0$) ya que las evaluaciones de estacionalidad realizadas evidenciaron que la serie presentaba estacionariedad en nivel.

Este procedimiento dio lugar a configuraciones extra para cada modelo, que reemplazaron a las versiones originales en fases posteriores de evaluación. Al conservar ambas versiones, se pudieron realizar comparaciones directas y cuantificar el impacto de la optimización de hiperparámetros en el desempeño predictivo de cada familia de modelos.

La fase de optimización finalizó con ocho modelos optimizados, cada uno ajustado con sus respectivos hiperparámetros. Estos modelos fueron guardados por separado para conservar su configuración y permitir su comparación en análisis posteriores. En la Tabla 12 se resumen las configuraciones entrenadas durante esta etapa de ajuste.

Tabla 12

Modelos Ajustados Mediante Optimización de Hiperparámetros

Modelo	Horizonte de predicción	Parámetros	RMSE (°C)
ARIMA	N/A	ARIMA (2, 0, 2)	3.16
SARIMA	N/A	SARIMAX(2,0,1)(1,0,1,24)	2.41
ARIMAX H+3	3 horas	ARIMA (2,0,2) + exógenas	1.89
ARIMAX H+48	48 horas	ARIMA (2,0,1) + exógenas	1.63
ARIMAX H+72	72 horas	ARIMA (2,0,1) + exógenas	1.69
SARIMAX H+3	3 horas	SARIMAX(2,0,2)(1,0,1,24) + exógenas	1,82
SARIMAX H+48	48 horas	SARIMAX(2,0,2)(1,0,1,24) + exógenas	2.76

SARIMAX H+72	72 horas	SARIMAX(2,0,1)(1,0,1,24) + exógenas	2.79
--------------	----------	-------------------------------------	------

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

N/A: modelo univariado sin variables exógenas por horizonte; se evalúa con un único pronóstico continuo sobre todo el conjunto de validación.

Validación Mediante Walk-Forward

Una vez identificadas las configuraciones óptimas para cada uno de los modelos de la familia ARIMA, se realizó la evaluación final mediante una estrategia de validación Walk-Forward, con el fin de analizar su desempeño en condiciones reales. A diferencia de la evaluación tradicional basada en un conjunto estático de pruebas, este método hace las predicciones de forma secuencial, integrando de manera progresiva la información observada a medida que se va haciendo disponible. Esta técnica es especialmente útil para casos climáticos, ya que los modelos deben dar predicciones continuas a medida que se añaden nuevos datos a la serie temporal.

La estrategia se ejecutó con las configuraciones obtenidas de la optimización de hiperparámetros, ya que representan la mejor predicción para cada modelo respecto al conjunto de validación. Los modelos se inicializaron con los datos del conjunto de entrenamiento y se evaluaron de manera secuencial a lo largo del periodo de prueba, para asegurar la uniformidad del procedimiento. En cada iteración, las predicciones se elaboraron con la información disponible hasta el instante de evaluación, preservando la coherencia temporal y evitando la incorporación de observaciones futuras que pudieran comprometer la validez de los resultados.

En los casos ARIMAX y SARIMAX, en la predicción se consideraron las variables exógenas específicas de cada horizonte, de manera coherente con la estrategia definida para la construcción del conjunto experimental. De esta manera cada horizonte fue tratado como un problema independiente, asegurando que las variables explicativas estuvieran alineadas en el tiempo y evitando que surgiera data leakage.

En la evaluación, los modelos no se reentrenaron por completo después de cada predicción, sino que su estado interno se actualizó gradualmente con la información nueva. Con esta táctica las estimaciones posteriores pudieron reflejar la evolución más reciente de la serie temporal, sin tener que volver a realizar todo el proceso de estimación de parámetros. Esto permitió disminuir de manera importante el costo computacional, sin perder la capacidad de adaptación de los modelos ante la incorporación continua de nuevas observaciones.

La aplicación de la metodología permitió obtener una serie de resultados que evidencian el comportamiento de las distintas variantes de la familia ARIMA bajo un mismo protocolo experimental. Dado que todos los modelos se entrenaron, optimizaron y evaluaron con las mismas particiones temporales y el mismo esquema de validación, las variaciones en el desempeño se pueden atribuir directamente a la estructura de cada modelo y la información que se utilizó en el proceso de predicción.

Finalmente, para facilitar la interpretación de los resultados, el análisis se realizó de forma secuencial. Primero se comparó el desempeño de los modelos referenciales contra las versiones optimizadas, lo que permitió cuantificar el impacto de la búsqueda de hiperparámetros sobre la capacidad predictiva. Después, se estudiaron las diferencias entre los modelos

univariados y los que incluyen variables exógenas, para comprobar el efecto de la información meteorológica extra en los distintos horizontes de predicción. Por último, se evaluaron todas las configuraciones y se identificó el modelo que durante el periodo de prueba alcanzó un mejor punto medio entre precisión, estabilidad y poder de generalización. Los resultados de la implementación y el diagrama de análisis secuencial se muestran en la Tabla 13 y en la Figura 12.

Tabla 13

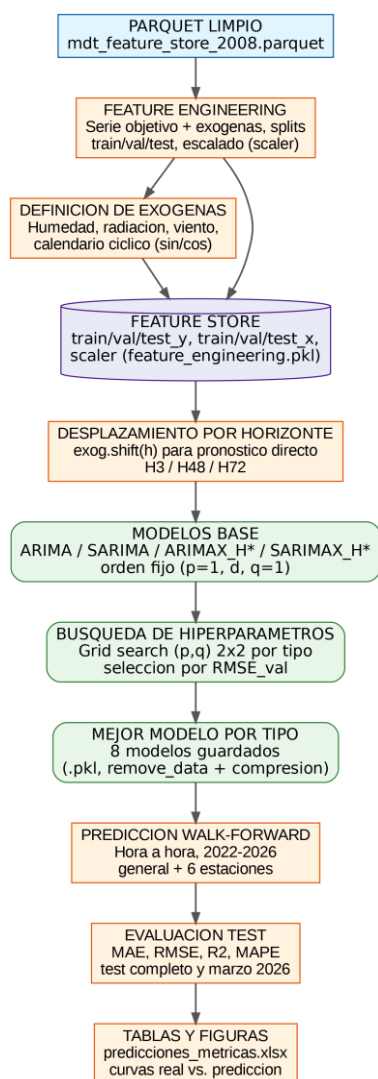
Evaluación de Modelos con Ajuste de Hiperparámetros sobre el Conjunto de Prueba

Modelo	Horizonte	RMSE (°C)	MAE (°C)	R²	MAPE (%)
ARIMA	3	3.25	2.74	-0.01	17.76
ARIMA	48	3.25	2.74	-0.01	17.76
ARIMA	72	3.25	2.74	-0.01	17.76
ARIMAX	3	1.98	1.49	0.62	9.64
ARIMAX	48	1.75	1.33	0.71	8.64
ARIMAX	72	1.80	1.37	0.69	8.86
SARIMA	3	0.95	0.67	0.91	4.35
SARIMA	48	1.31	0.98	0.84	6.39
SARIMA	72	1.33	0.99	0.83	6.49
SARIMAX	3	0.99	0.74	0.91	4.97
SARIMAX	48	1.30	0.97	0.84	6.35
SARIMAX	72	1.33	1.00	0.83	6.52

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Figura 12

Diagrama de Flujo para la Implementación de los Modelos de Aprendizaje Tradicionales



Nota. Elaboración propia.

Los resultados muestran que la incorporación del componente estacional generó una mejora consistente en la capacidad predictiva de los modelos evaluados. El modelo SARIMA, en particular, arrojó las mejores métricas, superando al modelo ARIMA. Este resultado muestra que

la temperatura del Distrito Metropolitano de Quito mantiene patrones periódicos que pueden representarse con mayor precisión mediante componentes estacionales específicos.

De forma análoga, los modelos SARIMAX presentaron un mejor desempeño que los ARIMAX, al considerar la información histórica de la serie con variables exógenas y la estructura estacional del proceso. Por ello, el modelo SARIMA fue elegido como el modelo estadístico de referencia principal de la familia ARIMA, conformando el modelo base sobre el que se comparará el desempeño de las arquitecturas híbridas propuestas en los siguientes capítulos. Esta elección se corroboró con la prueba de Diebold-Mariano, donde el SARIMA fue estadísticamente superior al ARIMA, ARIMAX y SARIMAX en los tres horizontes evaluados, con la mayor significancia frente a los modelos sin componente estacional, como se puede apreciar en la Tabla 14.

Tabla 14

Evaluación de Significancia Estadística del Modelo SARIMA Mediante la Prueba DM

Escenario	Par	MAE SARIMA	MAE rival	DM	p-valor	Conclusión
Corto	vs. ARIMA	0.807	2.851	-90.82	0.000	SARIMA mejor
Corto	vs. ARIMAX	0.807	2.120	-114.39	0.000	SARIMA mejor
Corto	vs. SARIMAX	0.807	0.889	-16.47	0.000	SARIMA mejor
Medio	vs. ARIMA	1.065	2.851	-63.40	0.000	SARIMA mejor
Medio	vs. ARIMAX	1.065	4.533	-172.46	0.000	SARIMA mejor
Medio	vs. SARIMAX	1.065	1.072	-2.23	0.026	SARIMA mejor
Largo	vs. ARIMA	1.084	2.851	-55.82	0.000	SARIMA mejor
Largo	vs. ARIMAX	1.084	4.155	-145.29	0.000	SARIMA mejor
Largo	vs. SARIMAX	1.084	1.095	-2.81	0.005	SARIMA mejor

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

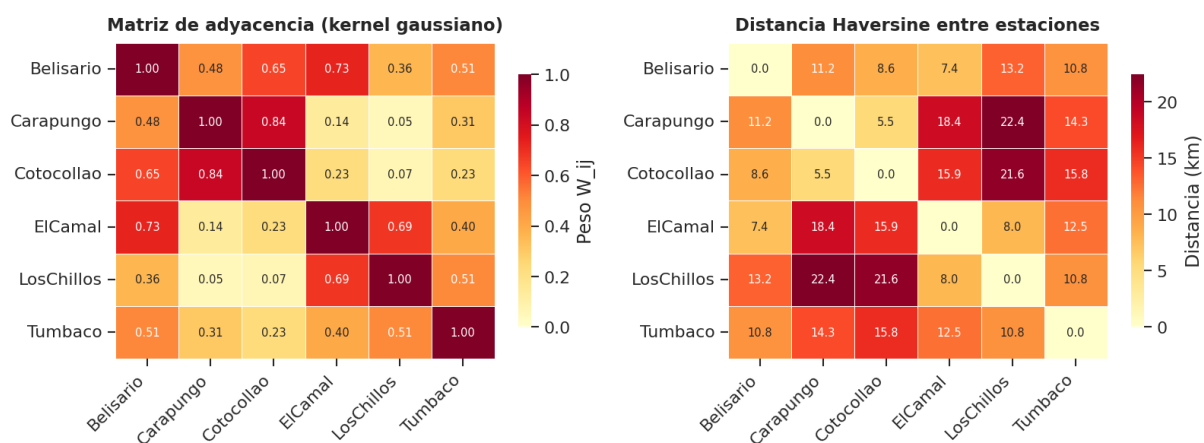
Integración de Modelos STGNN

Formalización del Grafo Espacial

Para modelar las dependencias espaciales se define un grafo $G = (V, E, W)$, donde los nodos (V) representan las seis estaciones meteorológicas. Las aristas (E) indican las conexiones entre estaciones y la matriz de adyacencia (W) asigna un peso W_{ij} que refleja el grado de relación entre las estaciones i y j . A mayor cercanía o influencia, mayor será el valor de W_{ij} .

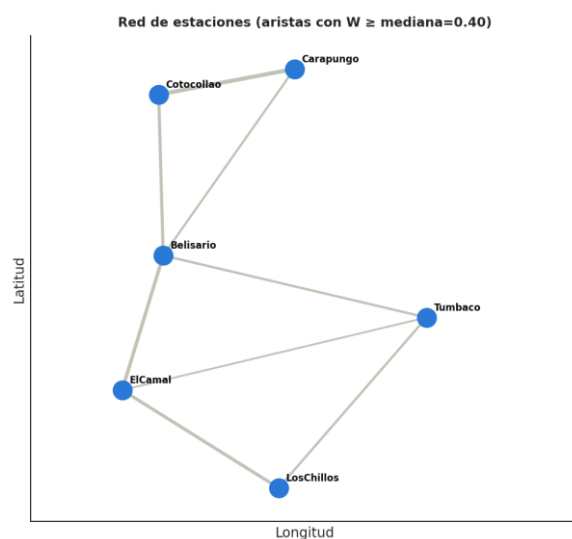
La matriz se construye bajo la hipótesis de que estaciones próximas tienden a compartir dinámicas térmicas similares, mientras que las más alejadas muestran comportamientos más independientes. Para calcular los pesos se utiliza la distancia Haversine (d_{ij}), que mide la separación física entre estaciones, y un kernel gaussiano que transforma esa distancia en un valor de afinidad (Li et al., 2018).

La Figura 13 muestra las distancias y la matriz de pesos. El parámetro σ corresponde a la distancia promedio entre todas las estaciones: valores pequeños hacen que los pesos tiendan a 1, mientras que valores grandes los acercan a 0. Por ejemplo, Cotocollao y Carapungo presentan el mayor peso (0.84) al estar a solo 5.5 km de distancia, mientras que Carapungo y Los Chillos, ubicadas en extremos opuestos de la red (22.4 km), tienen uno de los pesos más bajos (0.05).

Figura 13*Matrices de Adyacencia y Distancia Entre Estaciones*Grafo de 6 estaciones — $\sigma = 13.10$ km

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

En complemento, la Figura 14 ubica las seis estaciones según sus coordenadas reales y dibuja únicamente las aristas cuyo peso supera la mediana de la distancia promedio, con el fin de destacar qué pares de estaciones concentran la mayor parte del componente espacial que el modelo aprenderá.

Figura 14*Diagrama de Conexión de las Seis Estaciones*

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Formulación del Problema de Pronóstico

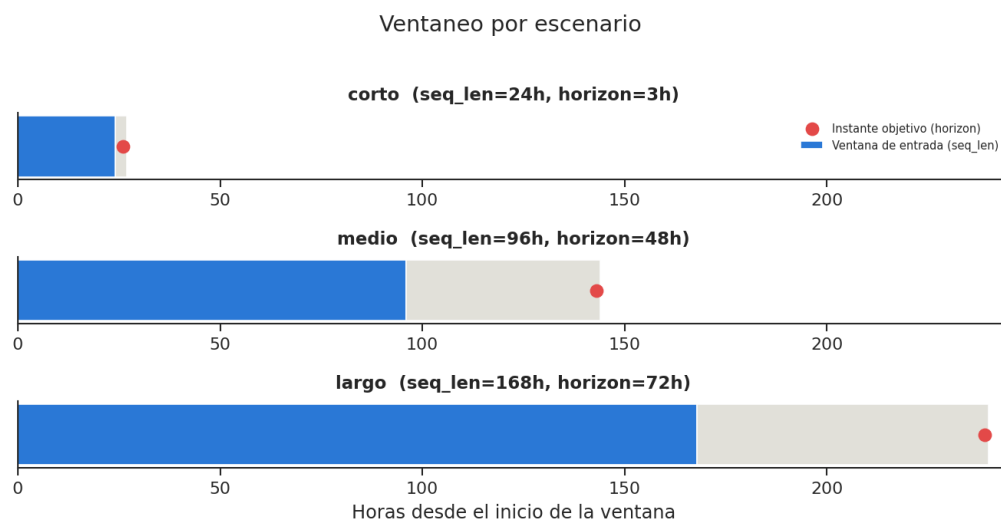
La formulación del problema se plantea bajo una estrategia de predicción directa, utilizando modelos independientes para cada horizonte temporal en lugar de un enfoque recursivo. En cada muestra, el modelo procesa una ventana de entrada $X_{i:i+L}$ de dimensiones $(L, 6, 13)$, que representa seis estaciones y trece variables, para predecir la temperatura exactamente h horas después de la última observación. La longitud de las ventanas históricas (L) se ajusta según el horizonte de predicción que se presenta en la Tabla 15.

Tabla 15*Configuración de las Ventanas Históricas por Escenario*

Escenario	Horizonte (h)	Ventana de entrada (L)	Justificación
Corto	3 h	24 h	Ciclo diurno.
Medio	48 h	96 h (4 días)	Persistencia de varios días.
Largo	72 h	168 h (7 días)	Patrón semanal.

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Las ventanas se construyen dentro de cada split. La Figura 15 ilustra, para cada escenario, el tamaño relativo de la ventana de entrada (barra azul) frente al salto hasta el instante objetivo (punto rojo): en el caso de corto plazo, el objetivo se ubica apenas 3 horas después del final de la ventana; mientras que, en el largo plazo, la ventana completa (168 h) junto con el salto (72 h) suman 10 días de contexto temporal por muestra.

Figura 15*Diagrama de Ventanas por Escenario*

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Arquitectura y Configuración Experimental de los Modelos

Los modelos A3T-GCN y DCRNN utilizan una arquitectura de salida idéntica (ReLU y capa lineal) para proyectar sus resultados, pero difieren en su estrategia para procesar el tiempo.

El A3T-GCN analiza la ventana histórica de forma global y utiliza un mecanismo de atención para identificar y dar mayor peso a los momentos del pasado que más influyen en el futuro. Por el contrario, el DCRNN opera de manera secuencial, procesando los datos paso a paso mediante convoluciones por difusión para actualizar un estado oculto que actúa como una memoria de la "inercia térmica" del sistema. La Tabla 16 resume la arquitectura de ambos modelos.

Tabla 16*Arquitectura de los Modelos STGNN*

Componente	A3T-GCN	DCRNN
Componente espacial	GCN sobre la topología del grafo.	Convolución por difusión de orden $K = 2$.
Componente temporal	GRU + atención sobre los pasos de tiempo.	GRU recurrente.
Entrada por muestra	(6, 13, L), una sola llamada.	(6, 13) por paso, se recorren L pasos.
Estado propagado	Atención pondera todos los pasos.	Estado oculto H arrastrado en la secuencia.
Canales ocultos	64 (configuración estándar).	64 (configuración estándar).

Nota. Elaboración propia basada en las especificaciones técnicas de Li et al. (2018) y Zhu et al. (2021).

De manera similar, ambos modelos se entrenan con la misma configuración, evitando que las diferencias en los resultados puedan atribuirse a un protocolo distinto. La Tabla 17 presenta la configuración experimental y de entrenamiento.

Tabla 17*Configuración Experimental y Entrenamiento*

Parámetro	Valor
Función de pérdida	MSE (en espacio escalado)
Optimizador	Adam
Learning rate	1e-3
Weight decay	1e-5
Recorte de gradiente	Norma máx. 1.0
Épocas máximas	60

Early stopping	Paciencia 8 (sobre MAE de validación en °C)
Semilla	123

Nota. Elaboración propia basada en las especificaciones técnicas de Li et al. (2018) para DCRNN y Zhu et al. (2021) para A3T-GCN, con ajustes de optimización realizados para el conjunto de datos de la REMMAQ.

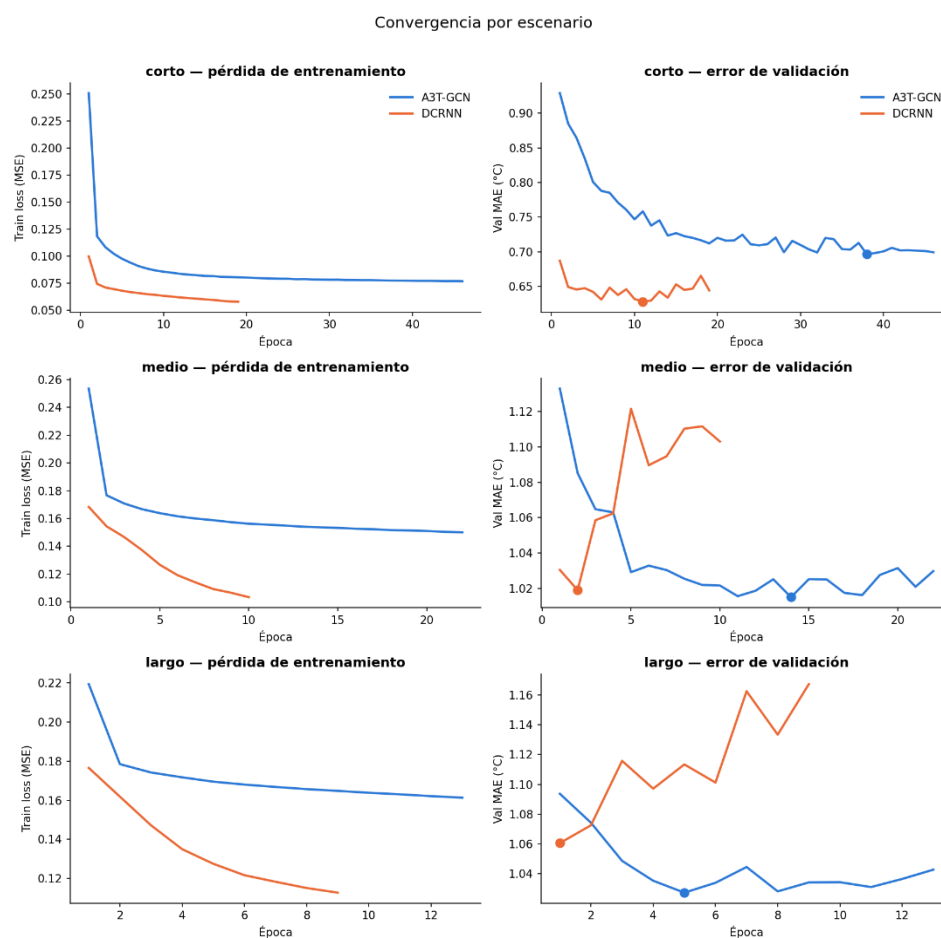
Batching a nivel de grafo. Para optimizar el rendimiento, el entrenamiento utiliza una técnica de batching por grafo disjunto, que procesa el lote completo de ventanas como una sola estructura en lugar de realizar múltiples llamadas independientes. Al desplazar los índices de los nodos y replicar sus conexiones para cada elemento del lote, el sistema resuelve el grupo entero simultáneamente sin que los datos se mezclen entre sí. En la práctica, esto permite que un lote de 3 ventanas (con 6 nodos y 30 aristas cada una) se trate como un solo grafo de 90 aristas, logrando que el tiempo de entrenamiento por época se reduzca de aproximadamente 5 horas a menos de 12 minutos.

Curvas de convergencia. Los resultados del entrenamiento muestran un patrón consistente (Figura 16 y Tabla 18): el modelo DCRNN alcanza una pérdida menor y mayor precisión en el corto plazo, aunque con un costo computacional más alto por época (313.6 s frente a 201.1 s). Esta diferencia se acentúa en el escenario de largo plazo, donde el A3T-GCN resulta más rápido al procesar la ventana temporal de forma global mediante su mecanismo de atención. En contraste, la naturaleza secuencial del DCRNN, que recorre la serie paso a paso, incrementa el tiempo de procesamiento. Asimismo, es importante señalar que el entrenamiento de los escenarios de corto y mediano plazo se realizó de manera simultánea en la misma GPU, mientras que para el largo plazo se ejecutó en una GPU dedicada.

En términos de eficiencia de aprendizaje, el DCRNN converge antes, deteniéndose por early stopping entre las 9 y 19 épocas, mientras que el A3T-GCN requiere entre 13 y 46. Cabe destacar que ambos modelos alcanzaron la convergencia optimizando el Error Cuadrático Medio (MSE) y restaurando el mejor estado antes de llegar al límite de 60 épocas, lo que valida la estabilidad de la configuración experimental en la red REMMAQ.

Figura 16

Curvas de Convergencia por Escenario



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Tabla 18*Tiempo de Entrenamiento por Escenario*

Escenario	Modelo	Épocas	Seg/época (media)	Tiempo total (min)
Corto	A3T-GCN	46	201.1	154.2
	DCRNN	19	313.6	99.3
Medio	A3T-GCN	22	1,230.3	451.1
	DCRNN	10	1,876.3	312.7
Largo	A3T-GCN	13	2,500.2	541.7
	DCRNN	9	4,027.4	604.1

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Evaluación de los Modelos STGNN

En la fase de evaluación, ambos modelos muestran una degradación progresiva de la precisión a medida que el horizonte de predicción se amplía, un comportamiento estándar en series temporales meteorológicas, como se resume en la Tabla 19.

Tabla 19*Métricas de Evaluación sobre el Conjunto de Prueba*

Escenario	Modelo	MAE	RMSE	R²	MAPE
Corto	A3T-GCN	0.720	1.019	0.914	4.680
	DCRNN	0.658	0.934	0.928	4.252
Medio	A3T-GCN	1.047	1.402	0.838	6.919
	DCRNN	1.064	1.425	0.833	6.970

Largo	A3T-GCN	1.096	1.462	0.824	7.188
	DCRNN	1.102	1.464	0.824	7.204

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

El modelo DCRNN supera a A3T-GCN en los escenarios de corto y mediano plazo; sin embargo, esta ventaja competitiva del DCRNN se reduce significativamente en el escenario de largo plazo, donde ambos modelos convergen hacia un desempeño similar (1.096 °C vs. 1.102 °C de MAE), sugiriendo que las diferencias arquitectónicas tienen un menor impacto cuando se proyectan tendencias a mayor escala temporal.

Protocolo de Comparación y Significancia Estadística

Para asegurar una comparación justa, ambos modelos comparten un protocolo experimental idéntico en cuanto a preprocesamiento, ventaneo y métricas, garantizando que cualquier variación en el desempeño se deba exclusivamente a sus arquitecturas. La validez estadística de estas diferencias se determinó mediante la prueba de Diebold-Mariano, el cual confirmó que la superioridad del modelo DCRNN es significativa únicamente en el horizonte de corto plazo (3 h). En cambio, en los escenarios de medio (48 h) y largo plazo (72 h), las métricas no aportan evidencia suficiente para rechazar la hipótesis de que ambos modelos poseen una exactitud predictiva equivalente, como se resume en la Tabla 20.

Tabla 20*Comparación y Significancia Estadística*

Escenario	h	MAE A3T-GCN	MAE DCRNN	Estadístico DM	p-valor	Conclusión
Corto	3	0.720	0.658	18.35	0.00	DCRNN mejor
Medio	48	1.047	1.064	-1.91	0.057	Equivalentes
Largo	72	1.096	1.102	-0.63	0.531	Equivalentes

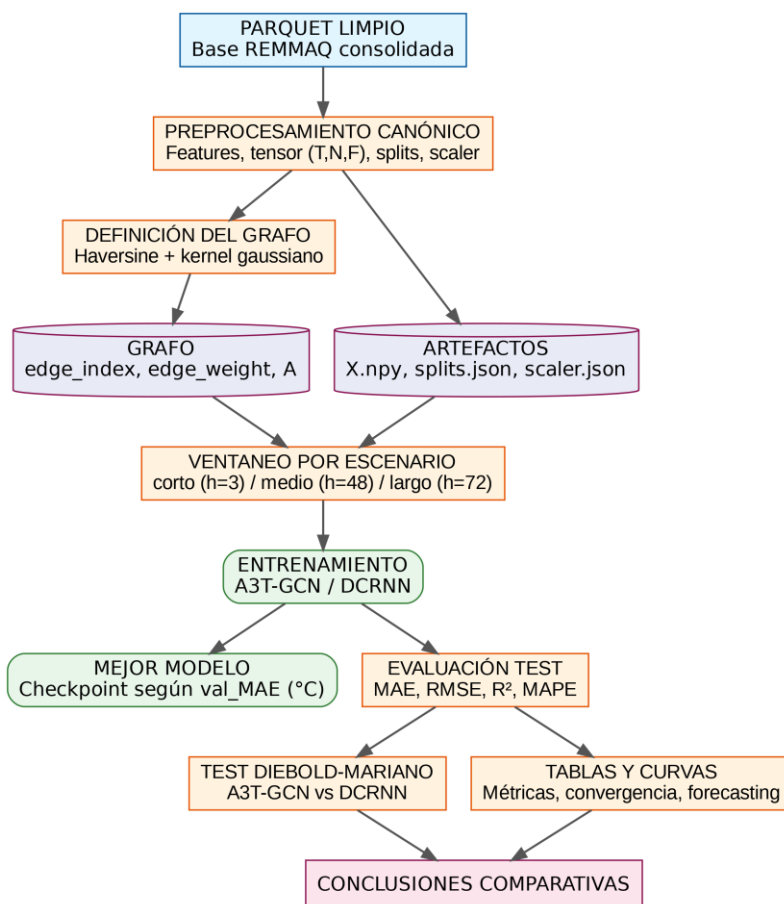
Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Flujo de Entrenamiento de Modelos STGNN

El flujo metodológico transforma los datos base en dos componentes paralelos: el tensor de variables (características climáticas y cíclicas) y la estructura del grafo (adyacencia geográfica). Ambos convergen en la fase de ventaneo, donde los datos se segmentan según los escenarios de corto, medio y largo plazo para alimentar el entrenamiento de los modelos A3T-GCN y DCRNN. Tras el entrenamiento con parada temprana, se restaura el estado óptimo de los modelos para realizar la evaluación sobre el conjunto de prueba. Los resultados de esta fase constituyen insumos para la validación estadística mediante la prueba de Diebold-Mariano, la construcción de las tablas de métricas (MAE, RMSE, R^2 , MAPE) y las curvas de convergencia que documentan el desempeño final. Todo este proceso se resume en el diagrama completo de la Figura 17.

Figura 17

Diagrama de Flujo para la Implementación de los Modelos STGNN



Nota. Elaboración propia.

Aplicación de Arquitecturas Fundacionales

En esta etapa se estudian las arquitecturas de propósito general Meta (Llama) y Mistral, las cuales en los últimos años se han utilizado en pronóstico de series temporales porque pueden identificar patrones cuando los valores numéricos se los representan en tokens (Jin et al., 2024). Sin embargo, debido a que estos modelos generalistas fueron diseñados principalmente para el

procesamiento del lenguaje natural y no para representar relaciones físicas inherentes a datos meteorológicos (Gruver et al., 2024), en este estudio se plantea la exploración de un enfoque alternativo que supere la rigidez de la tokenización textual.

Por este motivo, se evaluó el modelo fundacional ClimaX, desarrollado específicamente para el dominio climático. A diferencia de los modelos de lenguaje de propósito general, ClimaX procesa directamente las variables meteorológicas como tensores multidimensionales (Nguyen et al., 2023), empleando una representación numérica basada en parches espaciales, sin requerir la conversión previa de los datos a texto. Este enfoque preserva la información física original y permite modelar de forma más eficiente las relaciones espaciotemporales entre las variables meteorológicas (Lam et al., 2023).

Para asegurar una comparación consistente entre las arquitecturas evaluadas, se utilizaron las mismas variables meteorológicas de entrada durante las etapas de entrenamiento, validación y prueba. El conjunto de variables quedó integrado por temperatura media (variable objetivo), humedad relativa, presión barométrica, precipitación, radiación solar y las componentes u y v del viento. De esta manera, se mantuvo un espacio de entrada homogéneo, permitiendo que las diferencias observadas en el desempeño respondieran principalmente a la capacidad de modelado de cada arquitectura y no a un tratamiento diferenciado de las variables de entrada.

Finalmente, se compara el desempeño de las tres metodologías para determinar cuál funciona mejor para pronosticar de temperatura en Quito, debido a la complejidad inherente a la topografía y alta variabilidad climática de la ciudad, este estudio permite evaluar si un LLM adaptado puede alcanzar un desempeño similar o superior a un modelo especializado como

ClimaX. De esta forma, se garantiza que el modelo ganador de esta sección no solo se base en la precisión de la predicción, sino que sea consistente con el comportamiento atmosférico.

Fase 1: Preprocesamiento Diferenciado para Modelos Fundacionales

Para iniciar esta fase, se toma como base la MDT, la cual fue procesada mediante dos pipelines independientes. Esta separación se realizó debido a que los modelos evaluados necesitan formatos diferentes: por un lado, los modelos de lenguaje (Meta y Mistral) reciben información en formato de texto, por el otro, ClimaX requiere datos numéricos dispuestos en tensores espaciotemporales.

A pesar de las diferencias en la estructura de los datos, ambos *pipelines* utilizaron la misma estrategia de ventana deslizante, donde cada muestra se construyó con un historial de 72 horas para predecir las siguientes 72 horas, empleando un *stride* de 24 horas. Esta configuración se utilizó para comparar las diferentes arquitecturas bajo las mismas condiciones, tomando como referencia el horizonte de 72 horas por representar el escenario de mayor incertidumbre y complejidad predictiva (Hyndman & Athanasopoulos, 2021). Aunque esta configuración difiere de los escenarios objetivo definidos previamente, se empleó únicamente durante el *benchmark* para seleccionar la arquitectura con mejor desempeño. Posteriormente, el modelo seleccionado se entrenó nuevamente con las configuraciones de secuencia, horizonte y *stride* correspondientes a los escenarios de corto, medio y largo plazo. Definidos estos criterios comunes, a continuación, se describen las características de cada *pipeline*.

Pipeline JSONL: Serialización Conversacional para LLMs (Meta / Mistral). Los modelos Meta y Mistral se ajustaron mediante *Supervised Fine-Tuning* (SFT) con QLoRA,

debido a que estas arquitecturas procesan texto se necesitó transformar las series temporales en ejemplos conversacionales. Para ello, se estructuraron bloques de 72 horas con las variables de temperatura, humedad, presión, precipitación, radiación solar y las componentes u, v del viento, estos datos se los serializó en una tabla utilizando nombres abreviados (temp, hum, pres, precip, rad, u, v).

Cada registro se organizó en un archivo formato jsonl siguiendo el estándar conversacional de los modelos de lenguaje de tres roles:

- *System*: Define la tarea del modelo, se incluye la información de la estación y su ubicación geográfica, y se pide devolver el resultado con separadores según el formato csv para posteriormente leer los resultados en un dataframe.
- *User*: Este rol contiene la historia de las últimas 72 horas de las variables meteorológicas en formato de una tabla de texto y se le solicita la predicción del siguiente bloque.
- *Assistant*: Este rol contiene la respuesta esperada, la cual corresponde a las temperaturas reales de las siguientes 72 horas requeridas (la muestra de validación).

Este primer pipeline se realizó iterativamente para las 6 estaciones, resultando en un archivo de texto plano donde cada línea representa una muestra para el entrenamiento en el *fine-tuning*.

Pipeline ClimaX: Tensores Espaciotemporales (Vision Transformer). A diferencia de los modelos de lenguaje, que requieren transformar los datos numéricos en representaciones textuales, ClimaX opera directamente sobre tensores espaciotemporales. Estos tensores

contienen las variables meteorológicas (temperatura, humedad relativa, presión barométrica, precipitación, radiación solar y las componentes u y v del viento) organizadas espacial y temporalmente, preservando su representación numérica continua durante el proceso de inferencia.

Para la construcción multidimensional, las seis estaciones fueron mapeadas en una cuadrícula regular de dos dimensiones con un tamaño de 4 x4 celdas ($GRID_H = 4$, $GRID_W = 4$), cada estación fue asignada en una coordenada específica (r, c), estas posiciones se completan mediante *zero-padding*. Asimismo, a pesar de que ya se cuenta con continuidad en las series temporales, se aplican métodos de *Forward Fill* y *Backward Fill* para garantizarla en casos futuros. Los tensores resultantes tienen las siguientes dimensiones:

- *Entrada (X)*: Tiene la forma [Batch, 72, 7, 4, 4], la cual representa a las 72 horas de historia para las 7 variables distribuidas en una rejilla espacial de 4 x 4.
- *Salida (Y)*: Tiene la forma [Batch, 72, 1, 4, 4], la cual representa a las 72 horas del horizonte temporal para la variable objetivo (temperatura) sobre la misma rejilla.

Particionamiento Cronológico en los Pipelines. En concordancia con el criterio establecido en la Estrategia de Particionamiento de Datos, se aplicó la división cronológica en ambos pipelines.

Al evaluar las particiones resultantes para cada pipeline, se observa que en el pipeline de los LLMs se obtienen 39,954 muestras independientes, mientras que el pipeline de ClimaX genera 6,559 tensores espaciotemporales compactando las seis estaciones, la relación entre estos

dos procesamientos resultantes es exactamente de seis veces, esto valida la simetría y consistencia de los datos en ambos enfoques, tal como se muestra en la Tabla 21.

Tabla 21

Distribución de Muestras y Tensores por Partición Cronológica

Partición	Ventana Temporal	Muestras LLM (JSONL)	Tensores ClimaX (.pt)
Entrenamiento	Antes de 01-01-2022	30,666	5,111
Validación	Entre 01-01-2022 y 01-01-2024	4,380	730
Prueba	Después de 01-01-2024	4,908	818
Total	Completo	39,954	6,559

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Fase 2: Benchmark Inicial y Selección de la Arquitectura Base

Una vez generadas las particiones de los datos para cada arquitectura, se realizó un *benchmark* para comparar el desempeño de *Meta Llama 3.1-8B-Instruct*, *Mistral 7B-Instruct-v0.3* y *ClimaX* bajo las mismas condiciones. En esta etapa, se identifica que la arquitectura con la mejor capacidad de predicción utilizando la configuración temporal de 72 horas de entrada, 72 horas de predicción y un *stride* de 24 horas. Este proceso, se debe realizar antes de optimizar los hiperparámetros y experimentos para la arquitectura más robusta en los escenarios objetivo (corto, mediano y largo plazo).

Ajuste Fino de Meta y Mistral. Previo a la comparativa, se adaptaron los modelos de lenguaje al dominio climático de Quito utilizando *Supervised Fine-Tuning (SFT)* bajo las mismas

condiciones de entrenamiento. Esta homogeneidad en la metodología permite que la comparación se enfoque en las diferencias de las arquitecturas evitando que los resultados sean sesgados por la hiperparametrización y estilo de entrenamiento.

Para llevar a cabo el ajuste fino de manera eficiente, se realizó mediante el ecosistema *Unsloth*, el cual optimiza el uso de memoria. Asimismo, se usó la librería TRL para administrar los ciclos de entrenamiento. Por otra parte, para reducir el consumo de memoria GPU y que el entrenamiento de los LLMs sea posible se recurrió a la estrategia QLoRA (Dettmers et al., 2024), donde los modelos fueron cargados usando cuantización de 4 bits en el formato NormalFloat4 (NF4). Bajo esta configuración, los adaptadores LoRA que permiten mantener un equilibrio entre el aprendizaje y la prevención del sobreajuste son los siguiente: (Hu et al., 2022):

- **Rango (r)** = 16,
- **LoRA Alpha (α)** = 16.
- **LoRA Dropout** = 0.05.
- **Módulos Objetivo:** q_proj, k_proj, v_proj, o_proj, gate_proj, up_proj y down_proj.

Estos adaptadores se inyectan en las capas lineales de los bloques de atención y del perceptrón multicapa (MLP).

El entrenamiento se limitó a 65 pasos (*steps*) utilizando un batch size de 2 muestras por dispositivo y 8 pasos de acumulación de gradiente, obteniendo un batch efectivo de 16 muestras, esto permite a cada modelo procesar 1,040 ejemplos de entrenamiento (65×16), una cantidad suficiente para adaptar los pesos al dominio meteorológico, un número mayor de muestras de

entrenamiento incrementa el riesgo de sobreajuste y puede afectar el conocimiento adquirido durante el preentrenamiento (Gruver et al., 2024). Además, un batch efectivo de 16 brinda un equilibrio entre estabilidad del entrenamiento y consumo de memoria GPU, siendo una configuración ampliamente utilizada en el ajuste fino eficiente mediante QLoRA (Dettmers et al., 2024; Smith et al., 2018).

Para complementar la configuración y estabilizar gradientes, la optimización se realizó mediante AdamW de 8bits, con una tasa de aprendizaje de 2×10^{-4} , un decaimiento de tipo coseno y 10 pasos de *warmup*, donde el *warmup* reduce cambios bruscos del gradiente al inicio del entrenamiento y el decaimiento coseno permite refinar la convergencia durante las últimas iteraciones (Hu et al., 2022; Loshchilov & Hutter, 2017). Además, AdamW de 8 bits disminuye el consumo de memoria del optimizador sin afectar su desempeño (Dettmers et al., 2022).

Finalmente, como se detalló en secciones anteriores el ambiente de desarrollo para estos modelos fue en Google Colab, donde se utilizó una GPU NVIDIA A100 de 80 GB, por su estabilidad numérica en el entrenamiento y velocidad (Kalamkar et al., 2019). Los modelos se evalúan cada 13 pasos sobre los datos de validación y se configuró la opción `load_best_model_at_end = True`, para recuperar automáticamente el *checkpoint* con el mejor desempeño. Dado que Meta y Mistral operan como modelos autorregresivos de lenguaje, la métrica evaluada para el monitoreo y detención temprana (*early stopping*) es la pérdida de entropía cruzada (Cross-Entropy Loss), la cual cuantifica la precisión en la predicción del siguiente *token* meteorológico.

Adaptación del Modelo Fundacional ClimaX. Debido a que este modelo se basa en *Vision Transformers* (ViTs) y fue preentrenado con datos climáticos globales, no fue necesario entrenarlo desde cero. En su lugar, se aplicó transfer learning manteniendo los codificadores principales del *backbone* congelados (`param.requires_grad = False`), ajustando únicamente la cabeza predictora (*prediction head*) para adaptarla a las condiciones climáticas de Quito.

Para gestionar el proceso de adaptación, se utilizó PyTorch para cargar los tensores generados en el pipeline para ClimaX (archivos con extensión .pt). Con el objetivo de mejorar el rendimiento en el entrenamiento se habilita la opción `pin_memory = True`, la cual acelera la transferencia de datos hacia la memoria de la GPU. Además, la carga de datos se paralelizó mediante múltiples procesos (`num_workers`), reduciendo el tiempo de espera entre batches.

Para el ajuste de los pesos del *prediction head*, la optimización del modelo se realizó mediante `torch.nn.MSELoss()` como función de pérdida y AdamW como optimizador. A diferencia de los modelos de lenguaje, cuyo entrenamiento se basa en la predicción de tokens mediante funciones de pérdida como Cross-Entropy, el desempeño de ClimaX se monitoreó directamente utilizando el Error Cuadrático Medio (MSE) sobre el conjunto de validación, debido a que opera de forma nativa sobre tensores numéricos continuos.

Finalmente, para garantizar una comparación equitativa entre las arquitecturas evaluadas, el entrenamiento de ClimaX se realizó utilizando el mismo número de muestras empleado en los modelos de lenguaje (1.040 muestras de entrenamiento). Aunque ClimaX está diseñado para entrenar mediante iteraciones sobre épocas completas, el proceso se configuró para seleccionar el mejor modelo a partir del desempeño obtenido sobre el conjunto de validación dentro de este

mismo volumen de datos. De esta manera, se evitó que las diferencias de desempeño estuvieran influenciadas por una mayor exposición a los datos de entrenamiento, permitiendo que la comparación reflejara principalmente las capacidades de cada arquitectura.

Diseño de la Evaluación de Control. Una vez ajustadas las tres arquitecturas, se evaluaron las métricas para seleccionar el modelo base que continuaría a la fase de hiperparametrización y adaptación a los escenarios finales.

Para llevar a cabo la comparativa inicial, la evaluación no se realizó sobre todo el conjunto de prueba (enero de 2024 en adelante), sino únicamente sobre marzo de 2026. Esta decisión respondió a tres criterios metodológicos.

- Se seleccionó el mes de marzo por corresponder a uno de los periodos de mayor variabilidad climática en Quito, debido a la transición entre las condiciones de radiación diurna y las precipitaciones características de la región interandina (Secretaría de Ambiente del Distrito Metropolitano de Quito, 2024). Evaluar el desempeño durante este periodo constituye una prueba más exigente para los modelos predictivos, ya que la elevada variabilidad atmosférica permite analizar con mayor precisión sus capacidades de generalización (Hyndman & Athanasopoulos, 2021).
- El objetivo de esta etapa era seleccionar la arquitectura con mayor capacidad de generalización, no obtener aún las métricas finales del estudio.
- El motivo principal fue que la evaluación sobre un único mes redujo considerablemente el costo computacional, especialmente para Meta y Mistral, cuya inferencia se realiza de forma autorregresiva generando un token a la vez. Por lo tanto, extender esta evaluación a

todo el conjunto de prueba habría incrementado significativamente el tiempo de procesamiento sin aportar beneficios para esta fase de selección inicial (Gruver et al., 2024).

Para realizar las predicciones sobre esta ventana de control, en los modelos Meta y Mistral, la inferencia se ejecutó con una temperatura entre 0.01 y 0.1 para obtener respuestas deterministas. Además, se utilizó un inicio de generación forzado en formato CSV y posteriormente se aplicaron expresiones regulares (regex) para convertir las respuestas generadas en matrices numéricas comparables con los valores reales. Por otro lado, el modelo ClimaX recibió los tensores correspondientes a las siete variables meteorológicas de la ventana temporal anterior, generando las predicciones sin procesos intermedios, lo que resultó en una mayor velocidad de inferencia que en los modelos de lenguaje.

Evaluación Cuantitativa y Selección del Modelo Base. Una vez generadas y estandarizadas las predicciones de las tres arquitecturas para marzo de 2026, se procedió a evaluarlas utilizando cuatro métricas: el Coeficiente de Determinación (R^2), el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Error Porcentual Absoluto Medio (MAPE). Estas métricas permitieron analizar el desempeño de cada solución.

Como se observa en la Tabla 22, ClimaX obtuvo el mejor desempeño general al registrar el RMSE (métrica de validación) más bajo (4.77 °C) y el mejor valor de R^2 (-1.04), superando a Meta, que alcanzó un RMSE de 5.11 °C y un R^2 de -1.34, a pesar de que Meta presentó un MAE (3.93 °C) y un MAPE (24.99) ligeramente mejores que los de ClimaX.

Tabla 22

Comparativa Cuantitativa de Métricas de Error en marzo 2026

Modelo	MAE (°C)	RMSE (°C)	R²	MAPE (%)
Meta Llama 3.1	3.93	5.11	-1.34	24.99
Mistral 7B v0.3	6.41	7.39	-3.88	39.56
ClimaX Local	4.15	4.77	-1.04	27.44

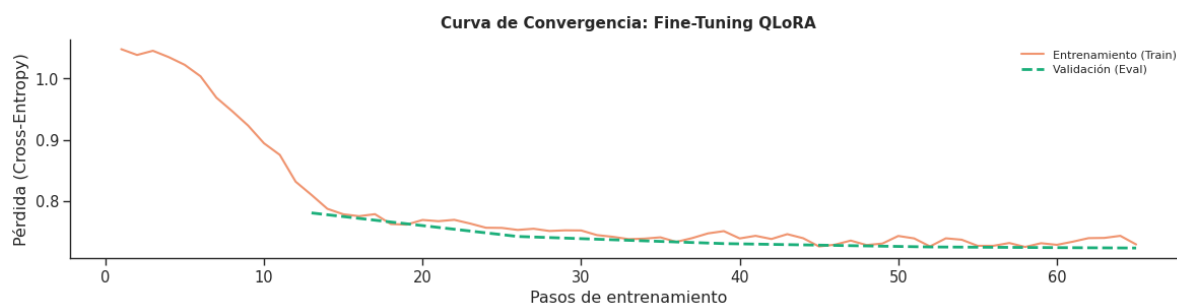
Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Después del análisis cuantitativo presentado en la tabla anterior, es importante evaluar el comportamiento de cada arquitectura durante el entrenamiento, ya que las curvas de pérdida permiten analizar la estabilidad del aprendizaje y ayudan a explicar las diferencias observadas en las métricas de desempeño.

En la evolución de la pérdida de entrenamiento y validación durante los 65 pasos en el ajuste fino de Meta, se observa un descenso progresivo en ambas curvas, alcanzando una convergencia estable a partir del paso 50 con un valor cercano a 0.72 en la pérdida de validación. Este comportamiento indica que el modelo logró adaptarse al dominio meteorológico sin presentar un sobreajuste, tal como se aprecia en la Figura 18.

Figura 18

Curvas de Pérdida de Meta Llama 3.1

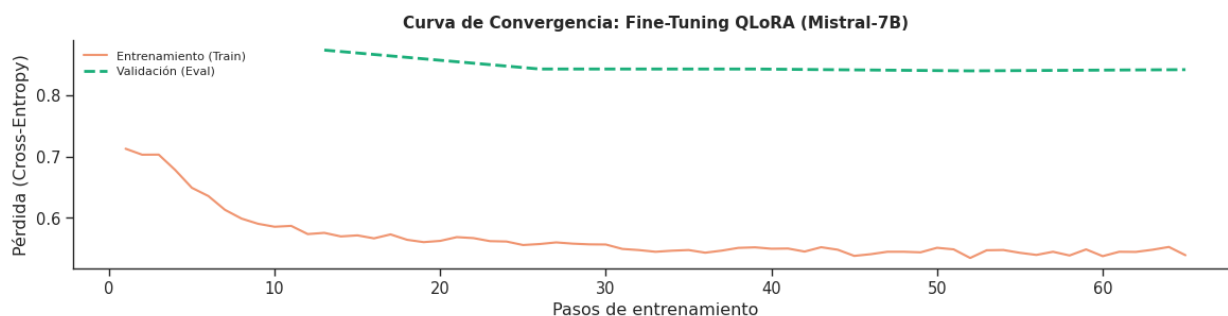


Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

A diferencia de la convergencia observada en Meta, la Figura 19 muestra que en el modelo Mistral mientras la pérdida de entrenamiento disminuye durante las iteraciones, la pérdida de validación comienza a incrementar después de las primeras pruebas. Esta divergencia indica sobreajuste, este comportamiento explica el bajo desempeño obtenido por este modelo, especialmente en las métricas RMSE y R^2 .

Figura 19

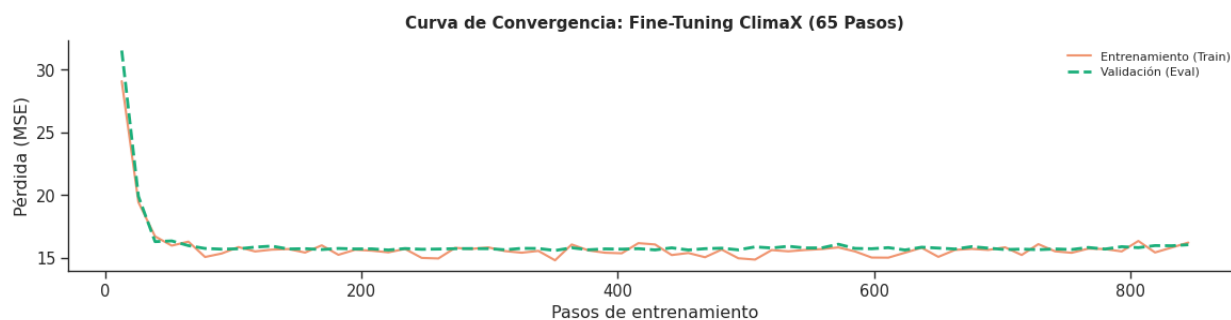
Curvas de Pérdida de Mistral 7B v0.3



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Por otra parte, al analizar las curvas de pérdida del modelo ClimaX, tanto la pérdida de entrenamiento como la de validación disminuyen de forma progresiva y mantienen un comportamiento similar durante todo el proceso, indicando que la estrategia de transfer learning permitió adaptar el modelo a las condiciones de Quito conservando una buena capacidad de generalización, tal como se muestra en la Figura 20.

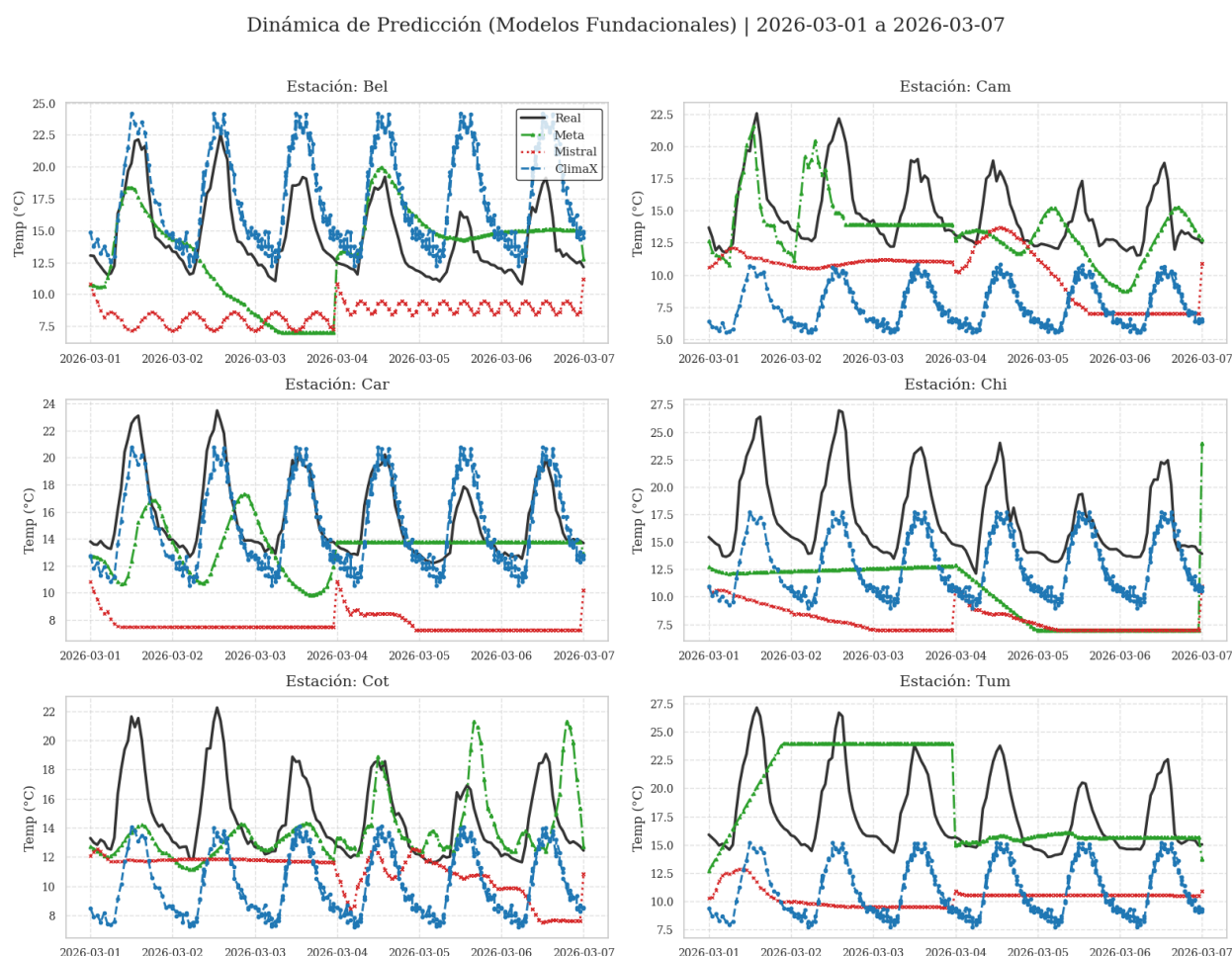
Figura 20
Curvas de Pérdida de ClimaX



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Para complementar el análisis de las curvas de pérdida y evaluar el comportamiento predictivo, en la Figura 21 se muestra una ventana comparativa de 72 horas del mes de marzo de 2026, donde se aprecia que ClimaX se adapta mejor a las variaciones del clima quiteño, especialmente en los incrementos diurnos y descensos nocturno. Por su parte, Meta también logra capturar la tendencia general, aunque tiene diferencias en la amplitud de algunos intervalos de tiempo. Por el contrario, Mistral muestra una desviación notable con respecto a los valores reales, esto es consistente con las métricas obtenidas y el sobreajuste observado en su curva de validación.

Figura 21
Comparación de Predicciones en Modelos Fundacionales



Nota. Elaboración propia a partir de las proyecciones realizadas con los modelos Mistral, Meta y ClimaX adaptados para un horizonte de 72 horas sobre los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

La principal diferencia entre ambas estrategias (modelos de lenguaje y modelo fundacional orientado al clima) es que Meta y Mistral adaptaron las series temporales mediante ajuste fino supervisado con QLoRA, mientras Climax trabajó directamente con tensores mediante *transfer learning*. Esta distinción marca una gran diferencia en la eficiencia

computacional, ya que ClimaX al procesar directamente datos numéricos requirió menor tiempo de entrenamiento e inferencia, mientras que los modelos de lenguaje necesitaron mayor cantidad de recursos y tiempo de procesamiento para una muestra pequeña, tal como se detalla en la Tabla 23.

Tabla 23

Comparación Técnica entre Arquitecturas

Parámetro / Característica	LLMs (Meta / Mistral)	Modelo Climático (ClimaX)
Tipo de Arquitectura	Autorregresiva / Lenguaje (LLM)	Vision Transformer (ViT) Espaciotemporal
Formato de Entrada	Texto estructurado / ChatML (.jsonl)	Tensores continuos tridimensionales (.pt)
Estrategia de Optimización	SFT con QLoRA (4 bits, NF4)	Congelamiento del <i>backbone</i> + Ajuste del cabezal
Adaptadores Inyectados	LoRA ($r=16$, $\alpha=16$, en todo MLP y Atención)	Ninguno (Ajuste lineal directo en el cabezal predictivo)
Métrica de Optimización y Validación	Entropía Cruzada	MSE (Mean Squared Error)
Librerías Clave	Unsloth, TRL, bitsandbytes, torchao	PyTorch, módulos nativos de visión-lenguaje
Procesamiento de Salida	Decodificación textual y formateo por RegEx	Regresión lineal directa por coordenadas temporales
Tiempo de Entrenamiento	~1 hora ($\approx 1,000$ muestras)	~ 2 minutos ($\approx 1,000$ muestras)
Tiempo de Inferencia (Predicción 72h)	~6 minutos	~ 40 segundos

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Los resultados obtenidos muestran que tanto Meta Llama 3.1 como ClimaX superan el desempeño de Mistral 7B. Posteriormente, la selección de la arquitectura final se realizó mediante una evaluación jerárquica. El estadístico de Diebold-Mariano (DM) se utilizó como criterio primario para determinar la existencia de diferencias estadísticamente significativas entre los modelos (Diebold, 2015). Al evaluar el horizonte operativo de 72 horas, la comparación entre ClimaX y Meta Llama 3.1 produjo un estadístico DM de 0.83 y un p-valor de 0.40 tal como se evidencia en la Tabla 24. Dado que el p-valor fue superior al nivel de significancia de 0.05, no se encontraron diferencias estadísticamente significativas entre ambos modelos en términos del Error Absoluto Medio (MAE). Ante esta equivalencia predictiva, el marco metodológico incorporó criterios secundarios de selección para determinar la arquitectura con mejor desempeño operativo (Cerqueira et al., 2021).

Tabla 24

Evaluación de Significancia Estadística (Test Diebold-Mariano) entre Modelos Fundacionales

h	MAE ClimaX	MAE Meta Llama 3.1	Estadístico DM	p-valor	Conclusión
72	4.15	3.93	0.83	0.40	Equivalentes

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Ante la equivalencia estadística demostrada en el MAE, se priorizó la minimización del RMSE, métrica en la que ClimaX obtuvo el menor error frente a los demás modelos fundacionales evaluados. Aunque esta diferencia (en el RMSE) no fue sometida a una prueba de significancia estadística independiente mediante el estadístico de Diebold-Mariano, la selección se fundamentó en el criterio de Chai y Draxler (2014), quienes señalan que el RMSE resulta

apropiado para el modelamiento de fenómenos físicos debido a que penaliza con mayor peso los errores de mayor magnitud. Adicionalmente, el análisis de las predicciones mostró que, mientras los modelos de lenguaje tendían a suavizar las variaciones de la serie, ClimaX preservó en mayor medida la variabilidad asociada al ciclo térmico diurno-nocturno. Finalmente, considerando el principio de parsimonia y eficiencia computacional propuesto por Schwartz et al. (2020), junto con su menor complejidad paramétrica y los tiempos de procesamiento obtenidos (Tabla 22), **ClimaX** fue seleccionado como el modelo fundacional base para la siguiente fase de la investigación, orientada a la optimización de hiperparámetros mediante Optuna, su adaptación a los escenarios de corto, medio y largo plazo, y la incorporación de la técnica RAF para mejorar su desempeño.

Fase 3: Optimización y Adaptación Multiescala

Una vez confirmada la superioridad de ClimaX en el análisis comparativo, se procedió a optimizar sus hiperparámetros y entrenar los modelos definitivos para los escenarios de corto, mediano y largo plazo. Para cumplir este propósito, en esta fase se desarrolló un nuevo preprocesamiento, se ejecutó la optimización bayesiana de hiperparámetros utilizando la librería Optuna y una adaptación independiente para cada horizonte de predicción.

Preprocesamiento Multiescala y Estandarización Diferenciada de Variables. En concordancia con el diseño de las ventanas temporales mostradas en la Tabla 4, se generaron tres conjuntos independientes de tensores, asignando uno para cada escenario de predicción. A diferencia del *benchmark* inicial, donde todos los modelos se compararon utilizando una única

configuración temporal de 72 horas de entrada y 72 horas de salida, en esta etapa cada modelo se construyó y adaptó utilizando la ventana temporal más adecuada para su horizonte de pronóstico.

Asimismo, como parte del preprocesamiento en Climax, se agrega la estandarización de variables mediante `StandardScaler` antes de generar los tensores multidimensionales de entrenamiento. Esta transformación no se aplicó durante el *benchmark* inicial porque esa etapa comparaba los modelos de lenguaje con ClimaX. Sin embargo, una vez seleccionado ClimaX como solución ganadora, la estandarización pasó a ser indispensable para mejorar la estabilidad del entrenamiento, ya que, al ingresar todas las variables en una escala comparable, se facilita la convergencia del optimizador durante la retropropagación.

Optimización Bayesiana (Optuna). Una vez realizado el preprocesamiento, se optimizó la tasa de aprendizaje del modelo mediante Optuna con el algoritmo *Tree-structured Parzen Estimator* (TPE). Este enfoque permitió encontrar automáticamente la configuración con menor error de validación, evitando una búsqueda manual (Akiba et al., 2019).

Esta optimización se realizó únicamente para el escenario de medio plazo (48 horas) y los hiperparámetros obtenidos se reutilizaron posteriormente en los escenarios de corto y largo plazo. Para llevar a cabo este proceso, se utilizó el 20 % de las muestras de entrenamiento y validación de este escenario. Esta estrategia permitió reducir el tiempo de ejecución de cada *trial*, ya que el objetivo de esta no era entrenar el modelo definitivo, sino identificar una configuración óptima de hiperparámetros. El uso de una muestra representativa de datos permite explorar el espacio de búsqueda con un menor costo computacional sin afectar significativamente la calidad de los hiperparámetros encontrados.

Respecto al diseño experimental, el espacio de búsqueda considera la tasa de aprendizaje (*learning rate*) y el tamaño del lote (*batch size*), explorando valores entre 1×10^{-5} y 5×10^{-4} en escala logarítmica para la tasa de aprendizaje y valores discretos de 8, 16 y 32 para el tamaño del lote. Cada configuración fue evaluada utilizando la función *MaskedMSELoss*, esta función calcula el error únicamente sobre las celdas de la cuadrícula que corresponden a estaciones meteorológicas reales e ignora las posiciones generadas por *zero-padding* (coordenadas estaciones genéricas).

Finalmente, el estudio se limitó a 15 trials, un valor suficiente para que el algoritmo TPE identificara la región de convergencia del espacio de búsqueda sin incrementar innecesariamente el tiempo de cómputo. El análisis alcanzó el mejor resultado en el trial 14, donde se identifica a 2.31×10^{-5} como el valor óptimo de *learning rate* y 32 como tamaño de lote ideal. Con esta configuración, el error de validación bajó de 0.47 a 0.18, tal como se muestra en la Figura 22.

Figura 22

Resultado de la Hiperparametrización en el modelo ClimaX

```
[I 2026-08-01 23:16:41,674] A new study created in memory with name: no-name-fae7710e-1842-4334-83b9-f106a572ffff

[INFO] Iniciando tuning ClimaX (medio_plazo)...

[I 2026-08-01 23:16:56,421] Trial 0 finished with value: 0.4704338610172272 and parameters: {'lr': 0.00015250365601434974, 'batch_size': 32}. Best is trial 0 with value: 0.4704338610172272.
[I 2026-08-01 23:16:58,674] Trial 1 finished with value: 0.21138490736454528 and parameters: {'lr': 0.00016686152127171178, 'batch_size': 36}. Best is trial 1 with value: 0.21138490736454528.
[I 2026-08-01 23:16:04,093] Trial 3 finished with value: 0.3356587031748837 and parameters: {'lr': 6.65592741388802e-05, 'batch_size': 32}. Best is trial 1 with value: 0.21138490736454528.
[I 2026-08-01 23:17:06,093] Trial 4 finished with value: 0.22770467541488647 and parameters: {'lr': 2.0419635050899967e-05, 'batch_size': 32}. Best is trial 4 with value: 0.18970483541488647.
[I 2026-08-01 23:17:08,781] Trial 5 finished with value: 0.4313358796596557 and parameters: {'lr': 0.00011962672114570705, 'batch_size': 3}. Best is trial 4 with value: 0.18970483541488647.
[I 2026-08-01 23:17:10,058] Trial 6 finished with value: 0.1835535764694214 and parameters: {'lr': 0.00016881445725305504, 'batch_size': 16}. Best is trial 6 with value: 0.1835535764694214.
[I 2026-08-01 23:17:10,058] Trial 6 finished with value: 0.19043201288114624 and parameters: {'lr': 3.1558871123216274e-05, 'batch_size': 0}. Best is trial 6 with value: 0.1835535764694214.
[I 2026-08-01 23:17:14,499] Trial 8 finished with value: 0.23478431566228483 and parameters: {'lr': 5.3955988474751234e-05, 'batch_size': 8}. Best is trial 6 with value: 0.1835535764694214.
[I 2026-08-01 23:17:10,801] Trial 9 finished with value: 0.19939254224308385 and parameters: {'lr': 5.301003388544435e-05, 'batch_size': 10}. Best is trial 6 with value: 0.1835535764694214.
[I 2026-08-01 23:17:18,721] Trial 10 finished with value: 0.2834915114687129 and parameters: {'lr': 0.00047031542728236, 'batch_size': 16}. Best is trial 6 with value: 0.1835535764694214.
[I 2026-08-01 23:17:29,641] Trial 11 finished with value: 0.2896753567457199 and parameters: {'lr': 1.69956267644484057e-05, 'batch_size': 32}. Best is trial 6 with value: 0.1835535764694214.
[I 2026-08-01 23:17:25,647] Trial 12 finished with value: 0.21884237229634866 and parameters: {'lr': 1.3293032628668217e-05, 'batch_size': 32}. Best is trial 8 with value: 0.1835535764694214.
[I 2026-08-01 23:17:27,550] Trial 13 finished with value: 0.21697016060352325 and parameters: {'lr': 0.00040236364463765856, 'batch_size': 32}. Best is trial 6 with value: 0.1835535764694214.
[I 2026-08-01 23:17:28,599] Trial 14 finished with value: 0.17859435081481934 and parameters: {'lr': 2.3105682322058442e-05, 'batch_size': 32}. Best is trial 14 with value: 0.17859435081481934.

[RESULT] BEST SCORE: 0.17859435081481934
[RESULT] BEST PARAMETERS: {'lr': 2.3105682322058442e-05, 'batch_size': 32}
```

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

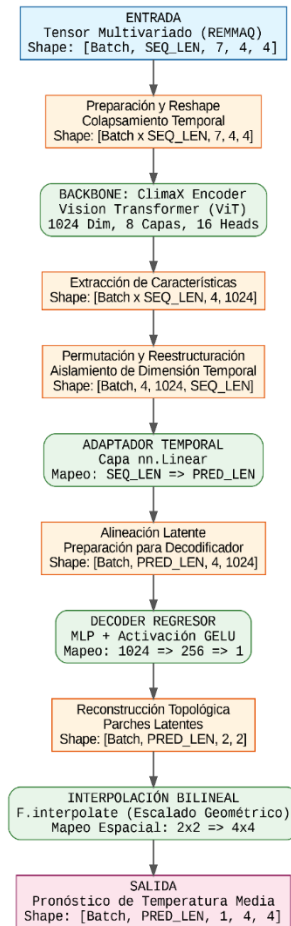
Adaptación y Transfer Learning de ClimaX en Tres Escalas. Con la tasa de aprendizaje óptima de 2.31×10^{-5} , se procedió al entrenamiento de los tres modelos finales de ClimaX para los escenarios de corto, mediano y largo plazo. En todos los casos se usó la clase *ClimaXFineTuningWrapper*, la cual facilitó la adaptación de ClimaX a los datos climáticos de Quito mediante los siguientes tres componentes principales:

- **Backbone pre-entrenado:** Encapsula codificador de ClimaX, conformado por un embedding de 1,024 dimensiones, 8 capas, 16 cabeceras de auto-atención basados en la arquitectura *Vision Transformer* (ViT) (Dosovitskiy et al., 2021). Este componente es responsable de obtener las representaciones espaciotemporales de las variables meteorológicas.
- **Adaptador Temporal:** Implementado con una capa lineal (*nn.Linear(seq_len, pred_len)*), se encargó de transformar la longitud de la secuencia histórica (*seq_len*) hacia el horizonte de predicción objetivo (*pred_len*) requerido en cada escenario.
- **Decodificador y Constructor Espacial:** Basado en un perceptrón multicapa (MLP) con función de activación Unidades Lineales de Error Gaussiano (GELU) (Hendrycks & Gimpel, 2016), proyecta las representaciones latentes hacia la variable objetivo (*temperaturaMedia*) y reconstruye la salida sobre la cuadrícula espacial de 4×4 mediante interpolación bilineal. De esta manera, se preservó la distribución topológica de las estaciones Quito.

En la arquitectura implementada mediante la clase `ClimaXFineTuningWrapper`, el flujo de procesamiento inicia con la ingesta de tensores multivariados con dimensiones `[Batch, SEQ_LEN, 7, 4, 4]`. Estos datos son procesados por el backbone de `ClimaX` para extraer representaciones espaciotemporales. Posteriormente, el adaptador temporal ajusta la longitud de la secuencia al horizonte de predicción definido para cada escenario, mientras que el decodificador reconstruye la temperatura sobre la cuadrícula espacial 4×4 mediante interpolación bilineal, generando finalmente el tensor de salida con dimensiones `[Batch, PRED_LEN, 1, 4, 4]`, tal como se detalla de forma gráfica en la Figura 23.

Figura 23

Arquitectura del módulo `ClimaXFineTuningWrapper` para la adaptación multiescala de `ClimaX`.



Nota. Diagrama del flujo de procesamiento espaciotemporal. Adaptado de `ClimaX: A foundation model for weather and climate`, por Nguyen et al. (2023).

El entrenamiento se realizó mediante el optimizador AdamW con un factor de regularización (*weight decay*) de 1×10^{-4} , valor que permitió controlar el sobreajuste durante el proceso de ajuste fino. Para la evaluación del error durante la retropropagación se utilizó la función de pérdida `nn.MSELoss()`, mientras que el optimizador empleó el valor predeterminado $\epsilon = 1 \times 10^{-8}$ para garantizar estabilidad numérica durante el cálculo de los gradientes.

Aunque la optimización bayesiana identificó un tamaño de lote de 32, durante el entrenamiento definitivo se ajustó el *batch size* de acuerdo con la longitud de las secuencias de cada escenario. Específicamente, se utilizó un lote de 96 para el escenario de corto plazo, debido a la secuencia de entrada 24 horas consume menos memoria VRAM. En cambio, para los escenarios de medio y largo plazo se usó un lote de 64, ya que en estos casos los tensores tienen mayor dimensionalidad. De esta manera, se aprovechó eficientemente la capacidad de la GPU NVIDIA A100 (80GB) sin exceder la memoria disponible en las secuencias de mayor longitud.

Finalmente, el entrenamiento se aceleró mediante Precisión Mixta Automática (AMP) y un GradScaler, reduciendo el consumo de memoria y el tiempo de entrenamiento sin afectar la estabilidad numérica. Además, se implementó un mecanismo de Early Stopping con una paciencia de 10 épocas y un $\text{min_delta} = 0.001$. Esta configuración detuvo automáticamente las iteraciones cuando no se observaban mejoras significativas en el conjunto de validación.

Análisis de Convergencia y Tiempos de Ejecución. Una vez adaptados los modelos a los tres escenarios experimentales, se analizó el costo computacional y comportamiento de convergencia en cada escenario. Las diferencias observadas se deben principalmente a la configuración de las ventanas temporales y al valor del *stride* utilizado para generar las muestras de entrenamiento.

En el escenario de corto plazo, el stride de tres horas produjo un mayor solapamiento entre ventanas, generando 40,904 muestras de entrenamiento distribuidas en 427 lotes por época. Esto resultó en un tiempo promedio de entrenamiento de aproximadamente 36.1 segundos por época. Por el contrario, en el escenario de largo plazo se utilizó un stride de 24 horas, lo cual

reduce las muestras a 5,107 y el lote a 80, permitiendo completar cada época en aproximadamente 27.2 segundos. Por lo tanto, aunque los tensores del escenario de largo plazo contienen secuencias más extensas, el menor número de iteraciones por época reduce el tiempo total en la adaptación del modelo.

Como resultado de las tres adaptaciones de ClimaX, el escenario de corto plazo alcanzó el menor MSE de validación normalizado (0.072), seguido del escenario de medio plazo (0.097) y luego el escenario de largo plazo (0.100). Este incremento del error es esperado, ya que la incertidumbre aumenta de acuerdo con el horizonte de predicción. Además, el mecanismo de *Early Stopping* permitió detener el entrenamiento cuando el modelo dejó de mostrar mejoras en este MSE de validación, evitando el sobreajuste y reduciendo el tiempo de procesamiento, tal como se detalla en la Tabla 25.

Tabla 25

Comparativa computacional y Convergencia de la Adaptación por Escala Temporal

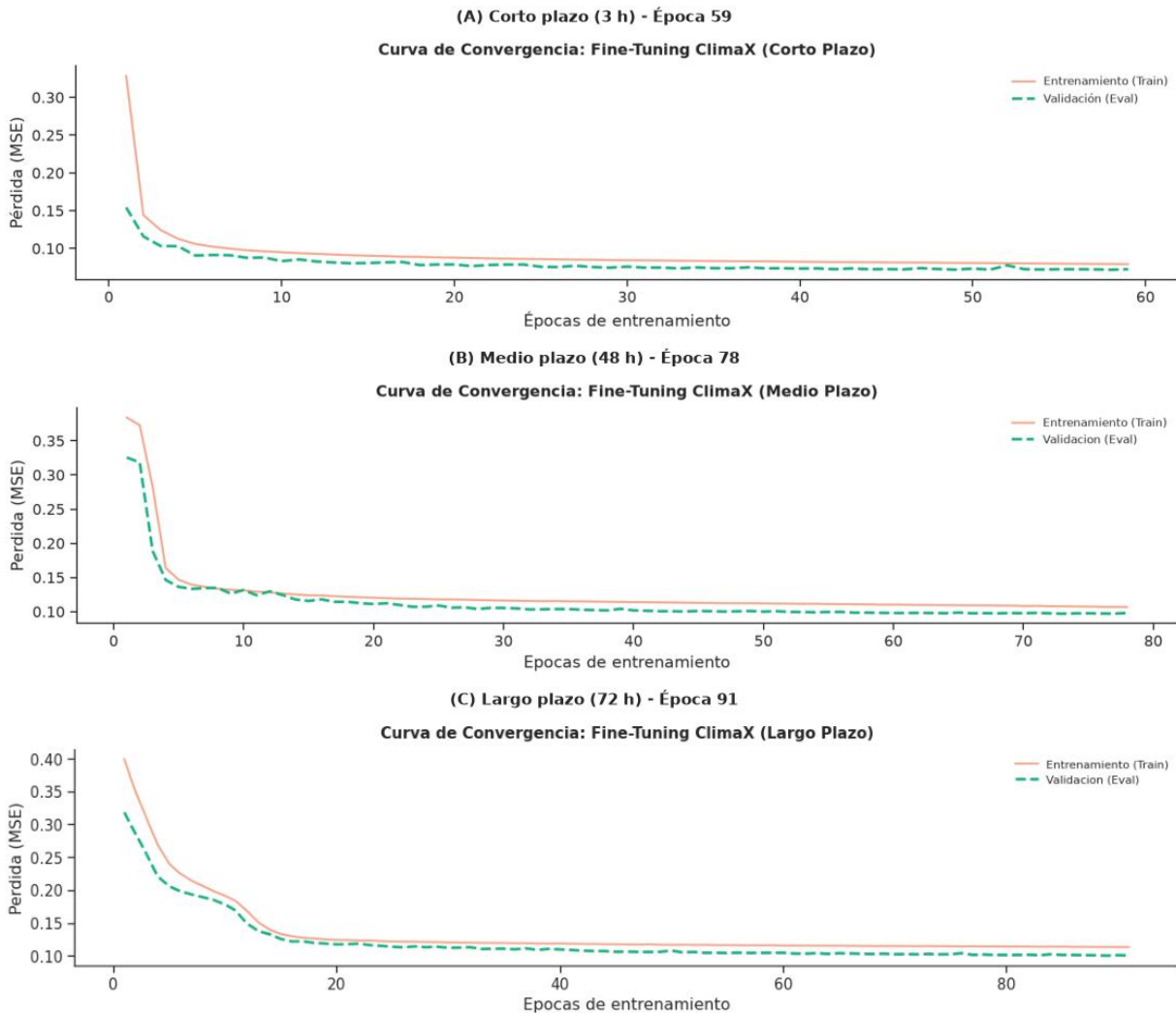
Escenario	Stride (h)	Muestras (Train)	Batches por Época	Batch Size	Tiempo por Época (s)	Época de Parada	MSE Validación (Norm.)
Corto	3	40,904	427	96	~36.1	59/100	0.072
Medio	12	10,220	160	64	~32.3	78/100	0.097
Largo	24	5,107	80	64	~27.2	91/100	0.100

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

En la Figura 24, se presentan las curvas de pérdida de entrenamiento y validación obtenidas durante el ajuste de ClimaX en los tres horizontes de pronóstico. En el escenario de corto plazo (3 h), el MSE de validación normalizado disminuye rápidamente durante las primeras épocas y alcanza su valor mínimo en la época 58 (0.072), evidenciando una buena capacidad de generalización al observar la estrecha cercanía entre las curvas de entrenamiento y validación. Por otra parte, en el escenario de medio plazo (48 h), ambas curvas presentan un descenso progresivo hasta estabilizarse en la época 73, donde se obtiene el menor MSE de validación (0.097), sin observar indicios de sobreajuste. Finalmente, en el escenario de largo plazo (72 h), el mecanismo de *Early Stopping* detiene el entrenamiento en la época 91 y recupera los pesos óptimos de la época 89, donde se registra el menor MSE de validación (0.100). Este comportamiento general demuestra cómo el incremento gradual de la pérdida entre escenarios es matemáticamente consistente con el aumento de la incertidumbre meteorológica conforme se amplía el horizonte temporal.

Figura 24

Curvas de Pérdida y Convergencia del Modelo ClimaX para Pronósticos Multiescala



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Finalmente, los tres modelos entrenados fueron almacenados como **climax_corto_plazo_best.pt**, **climax_mediano_plazo_best.pt** y **climax_largo_plazo_best.pt**. Estos archivos constituyen los modelos definitivos utilizados en la siguiente fase del estudio, la

cual corresponde a la adición de la estrategia **Retrieval-Augmented Forecasting (RAF)** para mejorar el desempeño de las predicciones.

Fase 4: Retrieval-Augmented Forecasting (RAF) e Inferencia Final

En esta última fase del método orientado a modelos fundacionales se enfoca en la reducción del suavizado en los valores extremos presente en los modelos predictivos de series temporales. Para mitigar este efecto, se implementó el mecanismo de RAF sobre los tres modelos ClimaX entrenados previamente con los patrones climáticos históricos de la partición de entrenamiento. De esta manera, se complementaron las predicciones mediante información observada en situaciones similares (Lewis et al., 2020).

Extracción de Embeddings y Construcción de la Base Vectorial. Para construir el motor de búsqueda se reutilizó el *backbone* de la clase *ClimaXFineTuningWrapper*, como extractor de características espacio temporales. Durante la indexación, los tensores históricos de entrada X, con dimensiones [N, seq_len, 7, 4, 4], fueron procesados en modo de inferencia (`torch.no_grad()`), obteniendo para cada ventana un embedding de 1,024 dimensiones que representa el estado atmosférico de ese instante.

Para gestionar eficientemente estos datos, los embeddings se almacenaron utilizando la librería FAISS (Facebook AI Similarity Search), la cual fue seleccionada porque está especializada en búsquedas de similitud sobre espacios vectoriales de alta dimensionalidad (Johnson et al., 2019). Para asegurar la máxima precisión de la recuperación se utilizó el índice

IndexFlatIP, el cual realiza búsquedas mediante producto interno, evitando cualquier proceso de compresión o pérdida de información en los vectores almacenados.

La base de conocimiento se construyó integrando los embeddings generados a partir de los conjuntos de entrenamiento y validación, comprendiendo toda la información histórica disponible hasta el 31 de diciembre de 2023. La adición del conjunto de validación permitió aumentar el espacio vectorial y obtener un mayor número de trayectorias climáticas históricas para la búsqueda de valores análogos. Se debe tener en cuenta que RAF se aplica una vez finalizado el proceso de adaptación de *ClimaX*. Por tanto, los registros del conjunto de validación únicamente se utilizan como base de consulta externa para la recuperación de información y no intervienen en el ajuste de los pesos del modelo.

Arquitectura RAF y recuperación de patrones históricos. Durante la inferencia, cada tensor de consulta es procesado por el *encoder* de *ClimaX* para generar su representación de 1,024 dimensiones. Posteriormente, este embedding se compara con la memoria vectorial utilizando Similitud del Coseno, implementada mediante normalización L2 seguida del producto interno de FAISS:

$$Similitud(u, v) = u \cdot v \quad (3)$$

Donde:

- u : embedding del tensor de consulta actual.
- v : embedding de la trayectoria climática almacenada en memoria vectorial.

La similitud del coseno se seleccionó porque compara la orientación entre vectores independientemente de su magnitud, permitiendo identificar patrones atmosféricos con comportamientos similares a pesar de sus intensidades.

El sistema recupera los cinco vecinos más cercanos (Top-5) proporcionando un equilibrio entre diversidad y precisión, conservando únicamente los valores climáticos más relevantes para cada consulta.

Estos cinco vecinos recuperados aportan sus trayectorias futuras reales (*targets históricos*). Para consolidar esta información, las cinco proyecciones se combinan mediante un promedio aritmético simple, asignando la misma ponderación a cada análogo recuperado y generando una única trayectoria de ensamble (*analog forecast*).

Fusión de Pronósticos dependiente del Horizonte. La predicción generada por ClimaX y la información recuperada por el sistema RAF se combinaron mediante un esquema de fusión, donde el peso de cada componente depende del horizonte de predicción. Esta estrategia se fundamenta en que la incertidumbre intrínseca de los modelos de pronóstico aumenta conforme se incrementa el horizonte temporal (Hyndman & Athanasopoulos, 2021). Por este motivo, la contribución de los patrones históricos recuperados adquiere mayor importancia en predicciones de mayor plazo (Li, 2025). La fusión final se rige por la siguiente ecuación:

$$Y_{final} = \alpha Y_{ClimaX} + (1 - \alpha) Y_{RAF} \quad (4)$$

Donde Y_{ClimaX} corresponde a la predicción generada por el modelo fundacional, Y_{RAF} es la predicción obtenida a partir del ensamble de los análogos históricos y α es el parámetro que controla la contribución de cada componente.

Los valores de α se definieron siguiendo un esquema de decaimiento para cada horizonte, considerando que la incertidumbre aumenta progresivamente conforme se extiende el tiempo de pronóstico reduciendo la capacidad predictiva incluso en arquitecturas especializadas (Lam et al., 2023; Kochkov et al., 2024).:

- **Corto plazo (3 h, $\alpha = 0.95$):** Se asignó un mayor peso a ClimaX debido a que las condiciones atmosféricas recientes contienen la mayor parte de la información necesaria para el pronóstico. En este escenario se toma un margen del 5% para RAF porque actúa como un mecanismo de corrección para las proyecciones, de tal manera que no degradan la generalización capturada por ClimaX, pero ayuda a filtrar valores atípicos.
- **Medio plazo (48 h, $\alpha = 0.85$):** A medida que la incertidumbre del sistema atmosférico comienza a incrementar, se redujo la ponderación del modelo ClimaX en un 15 %, aumentando la contribución de RAF. Este valor permitió incorporar información histórica sin perder la capacidad predictiva del modelo fundacional. Con ello, favorece una mayor estabilidad del pronóstico conforme aumenta el horizonte temporal.
- **Largo plazo (72 h, $\alpha = 0.50$):** Para este horizonte se estableció una distribución equitativa entre ClimaX y RAF. Dado que la incertidumbre es mayor, los patrones históricos similares adquieren un papel más relevante al complementar la predicción del

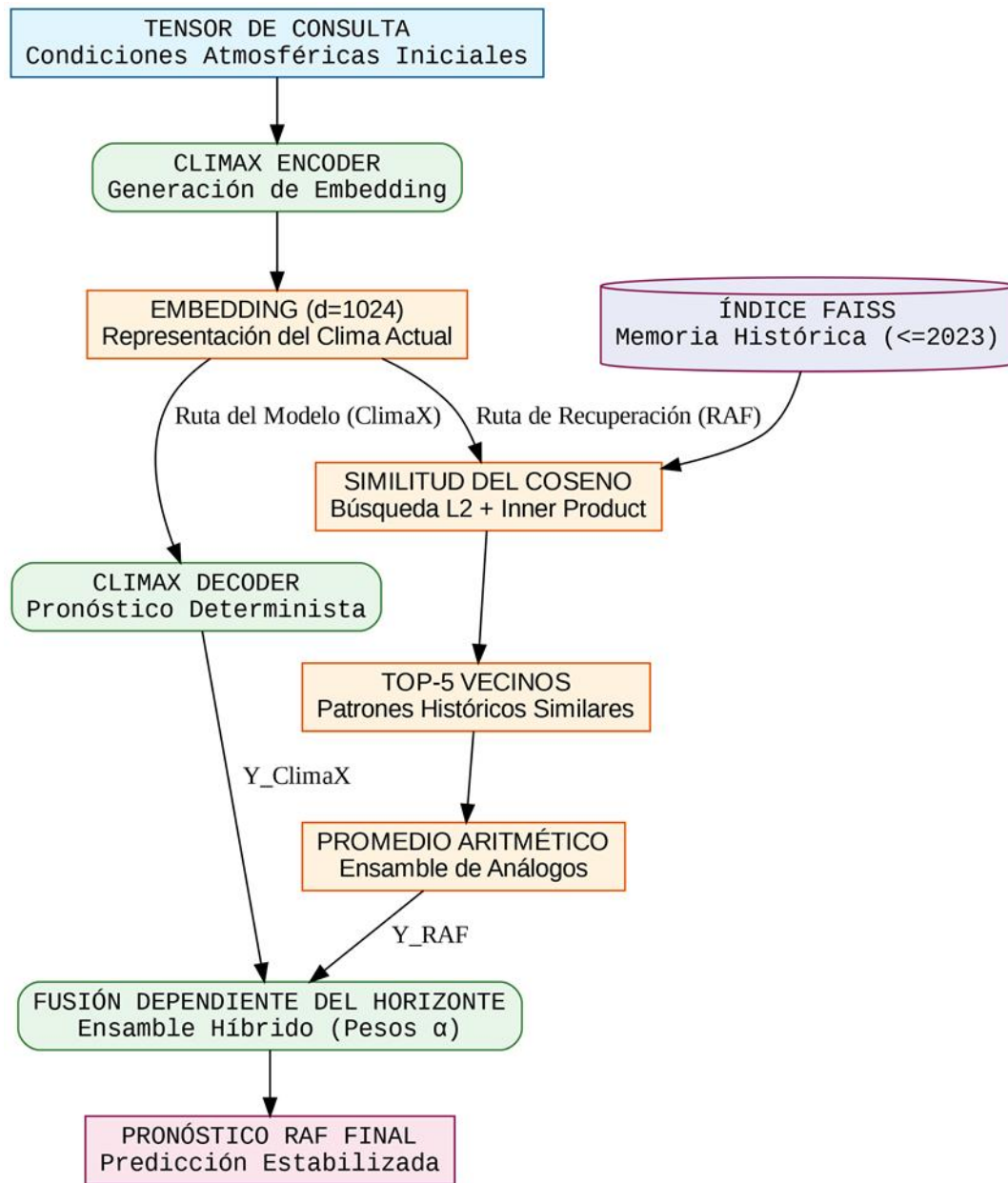
modelo, contribuyendo a reducir la divergencia y a mantener la estabilidad del pronóstico.

Luego, mediante esta integración entre la predicción del modelo ClimaX y la información recuperada desde la memoria histórica, se constituye un ensamble híbrido robusto que genera el pronóstico final utilizado durante la fase de inferencia.

En la Figura 25, se presenta la arquitectura del sistema RAF implementado en este estudio, donde el estado climático actual es procesado por el encoder de ClimaX para generar un embedding de 1024 dimensiones. Posteriormente, este vector consulta la base de conocimiento en FAISS y recupera los cinco patrones históricos más parecidos mediante la similitud del coseno. Finalmente, las trayectorias futuras de estos análogos se promedian aritméticamente y se combinan con la predicción de ClimaX utilizando el esquema de fusión dependiente del horizonte de predicción (α).

Figura 25

Arquitectura Propuesta del Sistema RAF Integrado con ClimaX



Nota. Elaboración propia. Fundamentada en la arquitectura propuesta por Nguyen et al. (2023), a la cual se le incorporó un módulo RAF. La implementación utiliza datos de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Integración, optimización computacional y desnormalización. Debido a que esta fase corresponde únicamente al proceso de inferencia y recuperación de patrones históricos, ya no fue necesario utilizar la GPU NVIDIA A100 empleada durante el entrenamiento de los modelos. La adaptación de ClimaX requirió dicho recurso debido al cálculo de gradientes y la actualización de millones de parámetros durante la retropropagación, mientras que en esta etapa únicamente se realizaron operaciones de inferencia. Por este motivo la generación de *embeddings*, la búsqueda vectorial en FAISS y el ensamble de predicciones se ejecutaron utilizando una GPU NVIDIA T4 (16 GB). Aunque estas tareas también podrían realizarse en una CPU convencional, el uso de la GPU permitió reducir los tiempos de procesamiento sin comprometer el desempeño, ya que no existe actualización de pesos ni optimización del modelo.

A pesar de que todo el proceso de recuperación y ensamble opera sobre datos normalizados mediante *StandardScaler*, la evaluación del desempeño requiere devolver las predicciones en unidades físicas (°C), por este motivo se aplicó la transformación inversa utilizando exclusivamente la media y desviación estándar calculadas sobre el conjunto de entrenamiento durante el proceso de estandarización:

$$Temp (^{\circ}C) = (Y_{norm} \times scale) + mean \quad (5)$$

Donde *Y_{norm}* corresponde a la predicción normalizada generada por el modelo, **scale** representa la desviación estándar y **mean** la media utilizada durante la estandarización.

Una vez obtenidas las predicciones desnormalizadas, se realizó un estudio de ablación con el propósito de evaluar el aporte del mecanismo RAF al desempeño predictivo. Para ello, se

comparó la arquitectura completa (ClimaX-RAF) con el modelo operando sin el módulo de recuperación de información histórica (ClimaX-NoRAF).

Los resultados presentados en la Tabla 26 muestran que, en los horizontes de corto y medio plazo (3 y 48 horas), ClimaX-NoRAF obtuvo un desempeño ligeramente superior, registrando menores errores absolutos y un coeficiente de determinación (R^2) ligeramente más alto (0.669 frente a 0.659 en el corto plazo). Sin embargo, en el horizonte de largo plazo (72 horas), ClimaX-RAF superó al modelo base en todas las métricas de evaluación, reduciendo el MAE a 1.584 °C, el RMSE a 2.140 °C (frente a 2.229 °C del modelo sin RAF) e incrementando el coeficiente de determinación de 0.590 a 0.622. Estos resultados indican que el mecanismo RAF aporta un mayor beneficio conforme aumenta el horizonte de predicción, ya que la recuperación de patrones históricos complementa la predicción del modelo cuando la incertidumbre atmosférica es mayor.

Tabla 26

Comparativa de Métricas de Ablación: ClimaX-RAF vs ClimaX-NoRAF

Escenario	Modelo	MAE (°C)	RMSE (°C)	R^2	MAPE (%)
Corto	ClimaX-NoRAF	1.453	2.001	0.669	8.734
Medio	ClimaX-NoRAF	1.627	2.212	0.596	9.859
Largo	ClimaX-NoRAF	1.649	2.229	0.590	10.028
Corto	ClimaX-RAF	1.467	2.033	0.659	8.798
Medio	ClimaX-RAF	1.692	2.313	0.558	10.219
Largo	ClimaX-RAF	1.584	2.140	0.622	9.664

Nota. Elaboración propia a partir de las predicciones del modelo ClimaX con y sin RAF sobre los datos de prueba de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Para complementar este análisis, se aplicó el estadístico de Diebold-Mariano, cuyos resultados se presentan en la Tabla 27. La prueba confirmó diferencias estadísticamente significativas entre ClimaX-RAF y ClimaX-NoRAF ($p < 0.001$). En los horizontes de corto y medio plazo, ClimaX-NoRAF obtuvo un desempeño ligeramente superior; sin embargo, en el horizonte de 72 horas el estadístico DM de -15.80 indicó una diferencia significativa a favor de ClimaX-RAF. Estos resultados respaldan la incorporación del mecanismo RAF en la arquitectura propuesta, ya que su contribución resulta más evidente conforme aumenta el horizonte de predicción y, con ello, la incertidumbre asociada al pronóstico.

Tabla 27

Comparativa de Métricas de Ablación: ClimaX-RAF vs ClimaX-NoRAF

Escenario	MAE ClimaX- RAF	MAE ClimaX-NoRAF	Estadístico DM	p-valor	Conclusión
Corto	1.467	1.453	29.267	0	ClimaX-NoRAF mejor
Medio	1.692	1.627	19.271	0	ClimaX-NoRAF mejor
Largo	1.584	1.649	-15.803	0	ClimaX-RAF mejor

Nota. Elaboración propia a partir de las predicciones obtenidas por el modelo ClimaX con y sin RAF sobre el conjunto de prueba de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Con base en estos resultados y siguiendo el criterio metodológico establecido durante el *benchmark* inicial, donde la selección se fundamentó en el desempeño alcanzado en el horizonte

de 72 horas por representar el escenario de mayor incertidumbre predictiva, se seleccionó a ClimaX-RAF como el modelo fundacional definitivo del estudio. Aunque en los horizontes de corto y medio plazo la arquitectura ClimaX-NoRAF obtuvo un desempeño ligeramente superior, las diferencias observadas en el MAE (0.014°C y 0.065°C , respectivamente) fueron reducidas desde un punto de vista práctico. En contraste, para el horizonte de largo plazo, ClimaX-RAF alcanzó mejores resultados en todas las métricas de evaluación y la prueba de Diebold-Mariano confirmó que esta diferencia es estadísticamente significativa ($\text{DM} = -15.80; p < 0.001$). En consecuencia, **ClimaX-RAF** fue seleccionado como el representante de los modelos fundacionales para la comparación final frente a las metodologías ARIMA y STGNN.

CAPÍTULO IV

Comparación de Métricas entre Modelos Implementados

La evaluación realizada sobre el conjunto de datos de prueba muestra que el modelo DCRNN obtiene las métricas de error más bajas en el horizonte de corto plazo (3h), con un MAE de 0.658 °C y un R^2 de 0.928, superando tanto a los modelos estadísticos tradicionales (SARIMA, MAE 0.673 °C) como al modelo fundacional (ClimaX-RAF, MAE 1.467 °C). A medida que crece el horizonte, esta ventaja se diluye: en el escenario de 72 horas, SARIMA registra el mejor MAE puntual (0.998 °C), aunque sin diferencia estadísticamente significativa frente a DCRNN (1.102 °C; ver Tabla 28).

Tabla 28

Resultados de Métricas Sobre el Conjunto de Prueba

Escenario	Modelo	MAE (°C)	RMSE (°C)	R^2	MAPE (%)
Corto	ClimaX-RAF	1.467	2.033	0.659	8.80
Corto	DCRNN	0.658	0.934	0.928	4.25
Corto	SARIMA	0.673	0.952	0.913	4.34
Medio	ClimaX-RAF	1.692	2.313	0.558	10.22
Medio	DCRNN	1.064	1.425	0.833	6.97
Medio	SARIMA	0.979	1.308	0.836	6.40
Largo	ClimaX-RAF	1.584	2.140	0.622	9.66
Largo	DCRNN	1.102	1.464	0.824	7.20
Largo	SARIMA	0.998	1.331	0.830	6.52

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Evaluación Estadística de la Precisión

Para determinar si las diferencias observadas responden a la aleatoriedad o a la arquitectura de los modelos, se aplicó la prueba de Diebold-Mariano (DM), cuyos resultados se presentan en la Tabla 29. Este análisis evidencia que el modelo DCRNN es estadísticamente superior a SARIMA y ClimaX en el horizonte de corto plazo. En cambio, en los horizontes de mediano y largo plazo, la prueba DM indica que DCRNN y SARIMA son estadísticamente equivalentes.

Tabla 29

Evaluación de Significancia Estadística Mediante la Prueba DM

Escenario	Par	MAE A	MAE B	DM	p-valor	Conclusión
Corto	DCRNN - SARIMA	0.658	0.807	28.04	0.000	DCRNN mejor
Corto	DCRNN - ClimaX-RAF	0.668	1.536	-66.96	0.000	DCRNN mejor
Corto	ClimaX-RAF - SARIMA	1.536	0.824	54.92	0.000	SARIMA mejor
Medio	DCRNN - SARIMA	1.066	1.066	0.11	0.916	Equivalentes
Medio	DCRNN - ClimaX-RAF	0.982	1.962	-33.60	0.000	DCRNN mejor
Medio	ClimaX-RAF - SARIMA	1.962	0.987	27.34	0.000	SARIMA mejor
Largo	DCRNN - SARIMA	1.085	0.998	-1.27	0.204	Equivalentes
Largo	DCRNN - ClimaX-RAF	0.696	1.062	-7.72	0.000	DCRNN mejor
Largo	ClimaX-RAF - SARIMA	1.062	0.687	7.48	0.000	SARIMA mejor

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Evaluación de Predicciones

Utilizando el mes de marzo de 2026 como criterio para determinar el modelo ganador, se observa que el DCRNN alcanza las mejores métricas en los tres escenarios analizados, tal como se muestra en la Tabla 30 y la Figura 26.

Tabla 30

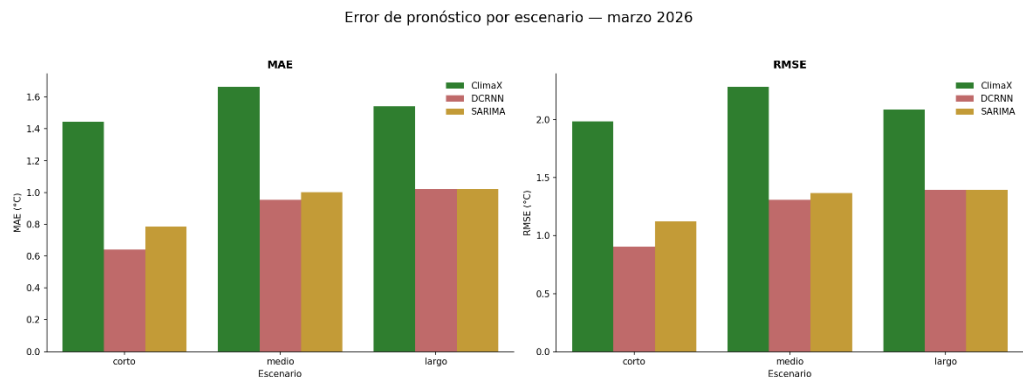
Evaluación del Modelo Ganador en Cada Escenario

Escenario	Ganador	MAE (°C)	RMSE (°C)	R ²	MAPE (%)
Corto	DCRNN	0.641	0.906	0.928	4.10
Medio	DCRNN	0.954	1.308	0.850	6.25
Largo	DCRNN	1.022	1.394	0.830	6.64

Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Figura 26

Evaluación del Error de Pronóstico por Escenario



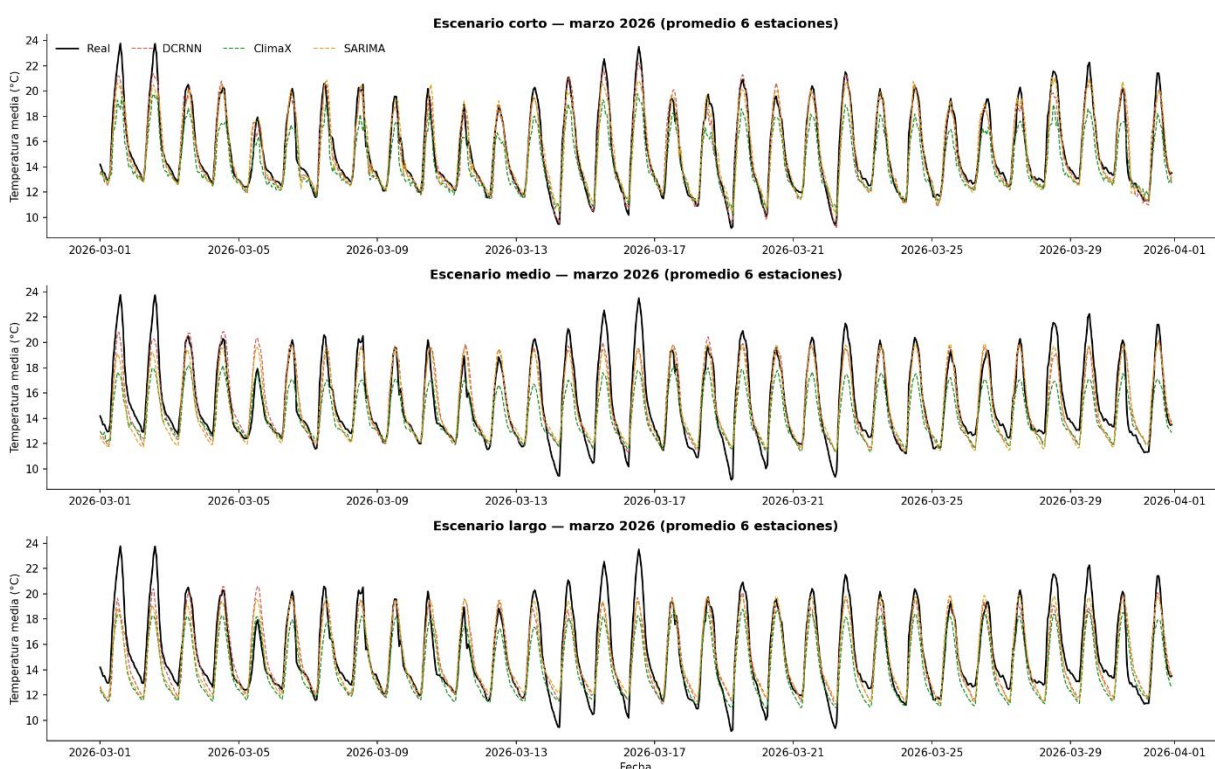
Nota. Elaboración propia a partir del preprocesamiento efectuado sobre los datos de las seis estaciones de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

La Figura 27 presenta un análisis visual del ciclo diario durante el periodo de comparación. El modelo DCRNN (rojo) se ajusta de manera casi exacta a la curva real de

temperatura, capturando con precisión tanto los picos máximos como los valles nocturnos. Por su parte, SARIMA (amarillo) reproduce el ciclo diario, aunque con un ligero desfase en los puntos extremos de temperatura. En contraste, ClimaX (verde), en el horizonte de 72 horas, sigue el ciclo diurno, pero con menor amplitud que los demás modelos, subestimando picos y valles.

Figura 27

Comparación entre Valores Reales y Predichos por Escenario



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

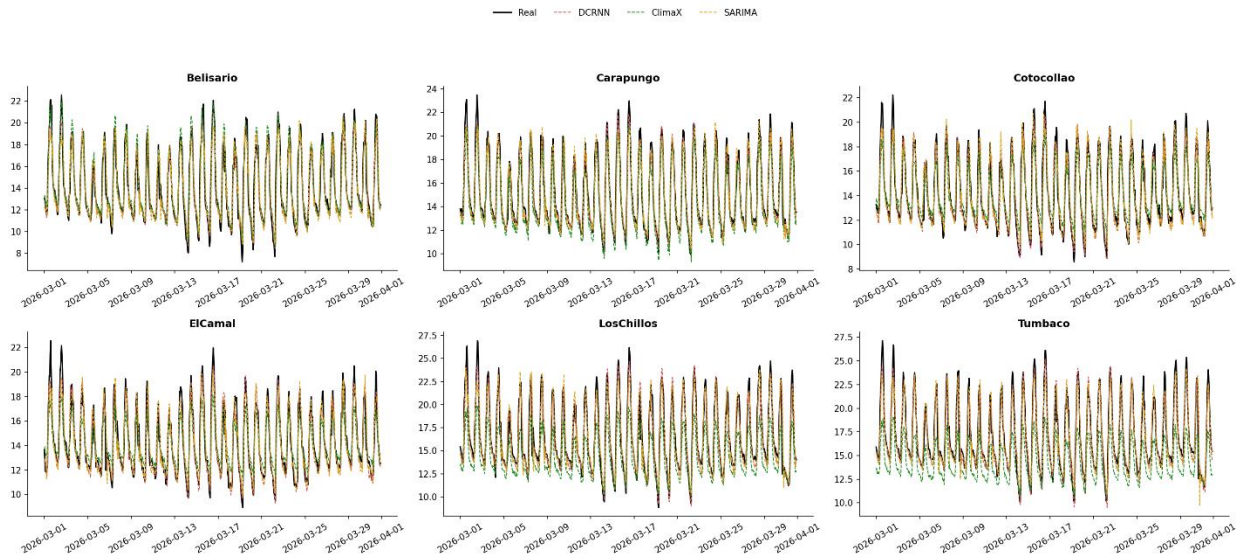
Evaluación de Predicciones a Nivel de Estación

El análisis detallado por estación permite comparar el desempeño global de los modelos, asociando la precisión predictiva con las características climáticas locales de cada sector del Distrito Metropolitano de Quito. En el corto plazo, el modelo vencedor por estación es de manera consistente el DCRNN. Los resultados fueron validados estadísticamente mediante la prueba DM, mostrando un p-valor significativamente menor a 0.001 en todas las estaciones de monitoreo. Esta precisión se explica por la capacidad de la arquitectura para capturar la inercia térmica inmediata y las relaciones espaciales entre estaciones cercanas.

Se evidencia que la precisión de los modelos DCRNN y SARIMA está estrechamente vinculada a la capacidad de representar la variabilidad de temperatura en cada estación. Por ejemplo, las estaciones de Los Chillos y Tumbaco, ubicadas en el valle, presentan una mayor variación térmica durante el día; en consecuencia, estos modelos logran seguir con precisión los picos extremos como se muestra en la Figura 28. En contraste, ClimaX subestima sistemáticamente dichas variaciones, mostrando una limitación al capturar esta dinámica local marcada.

Figura 28

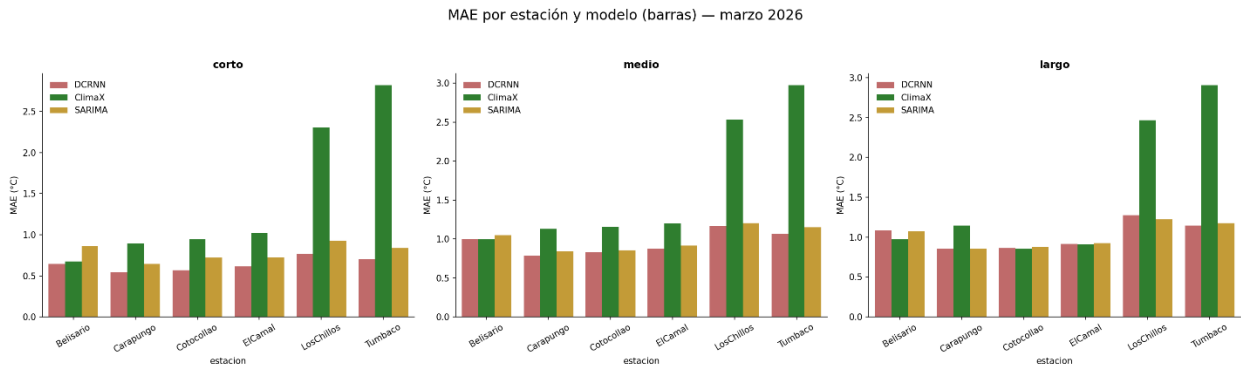
Comparación Entre Valores Reales y Predichos por Estación y Escenario



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

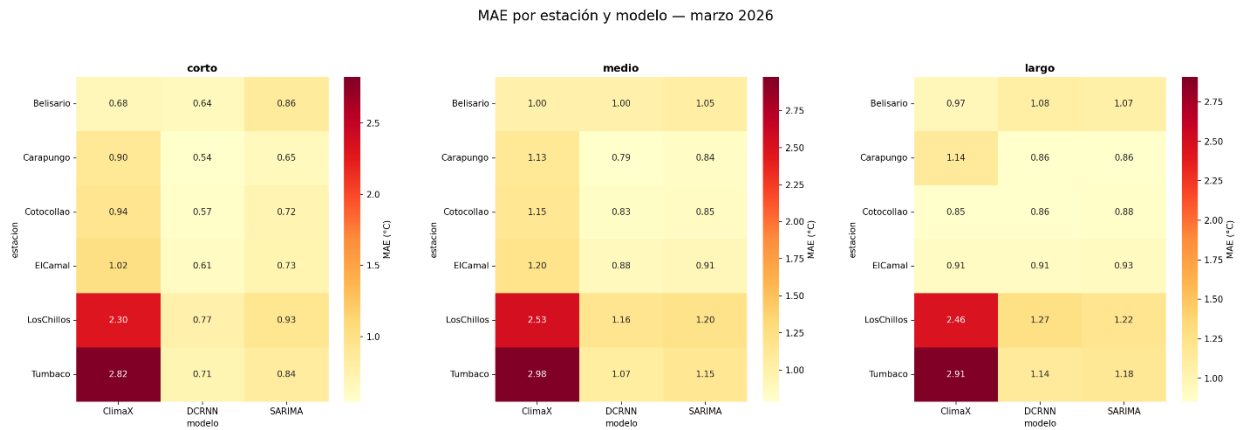
El gráfico de barras de la Figura 29 y el mapa de calor de la Figura 30 complementan este análisis, mostrando la magnitud del MAE por estación en los tres modelos. De manera agregada, se observa que DCRNN y SARIMA tienden a empatar en varias estaciones a partir del mediano y largo plazo, ambos claramente por debajo de ClimaX.

Figura 29
Comparación del MAE por Estación y Escenario



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Figura 30
Mapa de Calor del Error MAE por Estación y Escenario



Nota. Elaboración propia a partir de los datos históricos procesados de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ).

Finalmente, el análisis de predicción permite evidenciar varias tendencias relevantes. Las estaciones de Los Chillos y Tumbaco presentan una variabilidad térmica marcada por picos

diurnos, mientras que Carapungo muestra patrones similares a los observados en los valles, con una alta variación térmica. En contraste, las estaciones de Belisario, El Camal y Cotocollao exponen curvas térmicas más estables, con menor aprovechamiento del grafo temporal.

Discusión de Resultados

El resultado más destacado de este estudio es la dominancia estadística del modelo DCRNN en el horizonte de corto plazo (3h), con un MAE de 0.658 °C y un R^2 de 0.928. Este desempeño se explica por la incorporación de convoluciones por difusión sobre el grafo de la red REMMAQ, lo que permitió capturar con mayor precisión la inercia térmica y superar el rendimiento de los demás modelos evaluados. Este comportamiento coincide con lo reportado por Nurcahyo y Putranto (2026), quienes demostraron que el uso de Redes de Grafos incrementa la precisión en redes de observación irregulares en regiones tropicales. De manera similar, Ali et al. (2025) evidenciaron en su modelo STGNN que la capacidad de representar interdependencias dinámicas y espaciales resulta crítica para superar las limitaciones de los enfoques que consideran cada estación como un punto aislado. En el caso del Distrito Metropolitano de Quito, la construcción del grafo permitió que el modelo aprendiera cómo la topografía y la proximidad geográfica propagan los cambios de temperatura entre nodos vecinos, constituyendo un aporte significativo para el desarrollo del trabajo.

Además, un hallazgo importante y contraintuitivo es la equivalencia estadística entre DCRNN y SARIMA a partir del horizonte de 48 horas en el conjunto de datos de prueba. Mientras que en el corto plazo DCRNN muestra una ventaja significativa, en los horizontes medio y largo esta diferencia es mínima. Este comportamiento sugiere que, en pronósticos

extendidos para el Distrito Metropolitano de Quito, la estacionalidad diaria, capturada por el modelo SARIMA, predomina sobre la correlación espacial que aporta el grafo. Por tal motivo, se plantea una discusión importante de costo-beneficio para la gestión pública: en planificaciones superiores a dos días, un modelo univariado que no requiere GPU ni variables exógenas ofrece una precisión comparable a la de una red neuronal de grafos. Este resultado coincide con lo señalado por Ma et al. (2023) en su estudio HiSTGNN, donde advierten que la optimización multivariante enfrenta desafíos de inestabilidad conforme aumente el horizonte de predicción.

Asimismo, el bajo desempeño de ClimaX-RAF en el horizonte de largo plazo, con un R^2 de 0.622 muestra las dificultades de adoptar modelos climáticos globales a escalas urbanas sin un entrenamiento exhaustivo. Este resultado resalta la importancia de la heterogeneidad espacial. El análisis por estación muestra que la ventaja de modelos específicos como DCRNN es mayor en zonas de alta variabilidad térmica, como los Chillos y Tumbaco. Por el contrario, en estaciones más estables como Belisario o El Camal, la brecha del error se reduce, lo que indica que la utilización de arquitecturas complejas depende directamente de la variabilidad térmica del punto de monitoreo.

Por otro lado, la robustez de los resultados se sustenta en un análisis y preprocesamiento de datos riguroso. La aplicación de filtros físicos y la imputación bayesiana aseguraron que los modelos se entrenaran con una base de datos limpia y coherente con el comportamiento meteorológico. Este aseguramiento de calidad, alineado con los estándares de la U.S. EPA y la WMO, evitó la influencia de valores atípicos instrumentales durante el entrenamiento de los modelos.

Finalmente, los hallazgos obtenidos permiten concluir que la incorporación de arquitecturas híbridas de aprendizaje profundo constituye una alternativa viable para mejorar la predicción de condiciones climáticas en el Distrito Metropolitano de Quito. La integración de información espacial, dependencias temporales y mecanismos de aprendizaje profundo permitió superar de manera sistemática el desempeño de los modelos estadísticos tradicionales, en concordancia con la hipótesis planteada: la ventaja de la estructura espacial es significativa en el corto plazo y se diluye a medida que el horizonte se extiende, tal como anticipaba la literatura citada.

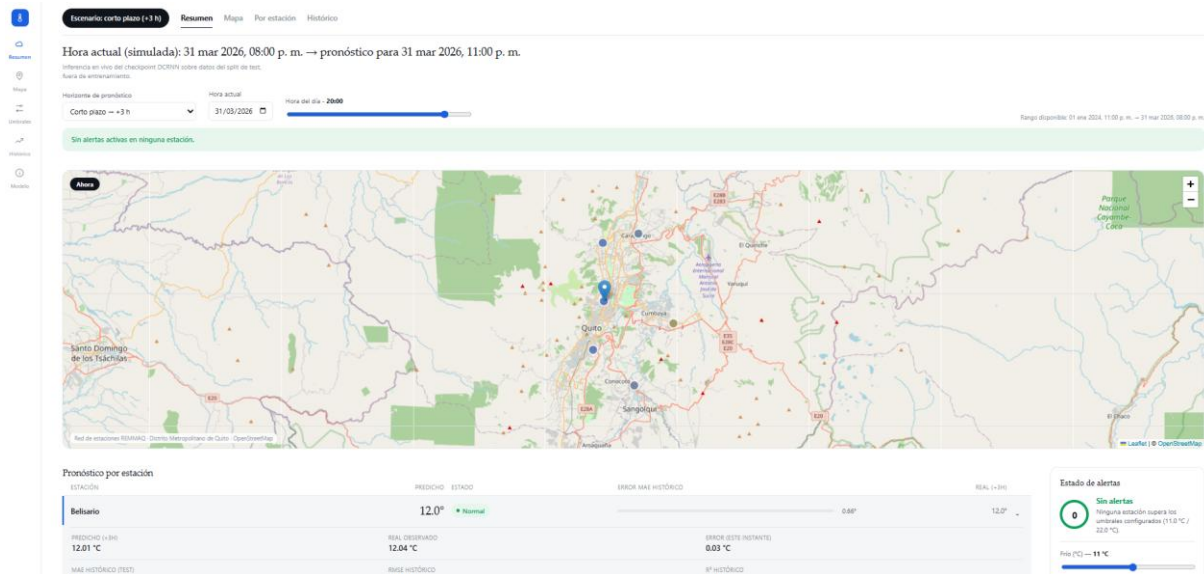
Con base en estos resultados, la siguiente sección presenta la implementación de un sistema de simulación e inferencia, que integra los modelos seleccionados dentro de una aplicación web y demuestra la factibilidad de trasladar los avances obtenidos hacia un entorno operativo orientado al monitoreo climático y la generación de alertas frente a eventos de temperatura extrema.

Implementación del Sistema de Simulación e Inferencia

Se implementó un sistema de simulación e inferencia de los modelos seleccionados para los horizontes de corto, mediano y largo plazo, permitiendo consultar la predicción de la temperatura de manera interactiva mediante una aplicación web. El sistema se diseñó como una plataforma de seguimiento climático, capaz de convertir el conocimiento adquirido durante la fase experimental en un instrumento de apoyo para la supervisión de eventos de temperaturas extremas.

La herramienta desarrollada ofrece una interfaz web en la que el usuario puede seleccionar el horizonte de predicción de interés: corto, mediano o largo plazo, establecer la fecha y la hora de la consulta y elegir cualquiera de las estaciones meteorológicas de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ). A partir de estos parámetros, el sistema ejecuta automáticamente el modelo entrenado correspondiente al horizonte seleccionado y muestra la temperatura estimada junto con el valor observado para el mismo instante, lo que permite comparar de forma inmediata el comportamiento del modelo frente a los datos registrados. Además, la interfaz incorpora información complementaria sobre el rendimiento histórico del modelo, lo que facilita la interpretación de los resultados de la inferencia.

La Figura 31 presenta el diseño de la interfaz principal del sistema e ilustra el entorno de interacción implementado para consultar las predicciones. En él se incluyen los controles necesarios para elegir el escenario de predicción, la fecha, la hora y la estación meteorológica, junto con los componentes empleados para visualizar los resultados generados por los modelos.

Figura 31*Interfaz Principal del Sistema de Simulación e Inferencia*

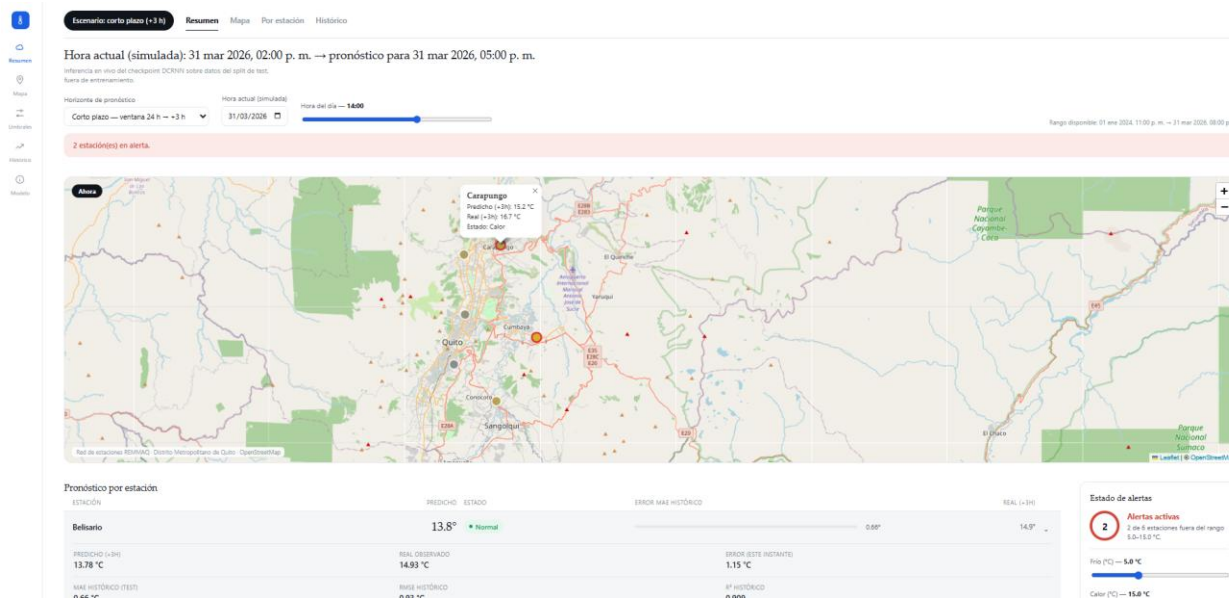
Nota. Elaboración propia a partir del diseño y validación tecnológica del sistema de alerta temprana, considerando pruebas de usabilidad y métricas de evaluación de desempeño.

Para facilitar la interpretación espacial de las predicciones, la aplicación incorpora un mapa interactivo que muestra la ubicación de las estaciones meteorológicas empleadas durante la investigación. Esta funcionalidad permite visualizar de forma simultánea el comportamiento térmico de las distintas áreas supervisadas y seleccionar una estación específica para obtener información detallada sobre la predicción.

Uno de los elementos importantes del sistema es el módulo de generación de alertas, que permite adaptar de forma dinámica los criterios aplicados para identificar posibles eventos de frío o calor. En lugar de recurrir a límites fijos predefinidos, la aplicación permite al usuario establecer los valores mínimos y máximos de temperatura que activarán una notificación. Una

vez realizada la predicción, el sistema compara automáticamente la temperatura proyectada con los umbrales definidos y destaca las estaciones que presentan condiciones fuera del rango establecido mediante indicadores visuales sobre el mapa y mensajes de alerta en la interfaz. Esta característica ofrece un mecanismo flexible para ajustar el comportamiento del sistema a distintos escenarios de monitoreo, y convierte la plataforma en un instrumento de apoyo para la detección temprana de condiciones térmicas potencialmente críticas.

La Figura 32 presenta un ejemplo del funcionamiento del módulo de visualización y generación de alertas. En él se aprecian la ubicación geográfica de las estaciones meteorológicas, las predicciones generadas por los modelos y la identificación automática de las ubicaciones que superan los umbrales de temperatura definidos por el usuario.

Figura 32*Visualización de las Predicciones y Generación de Alertas Sobre el Mapa Interactivo*

Nota. Elaboración propia a partir del diseño, validación tecnológica y pruebas de desempeño del sistema de alerta temprana.

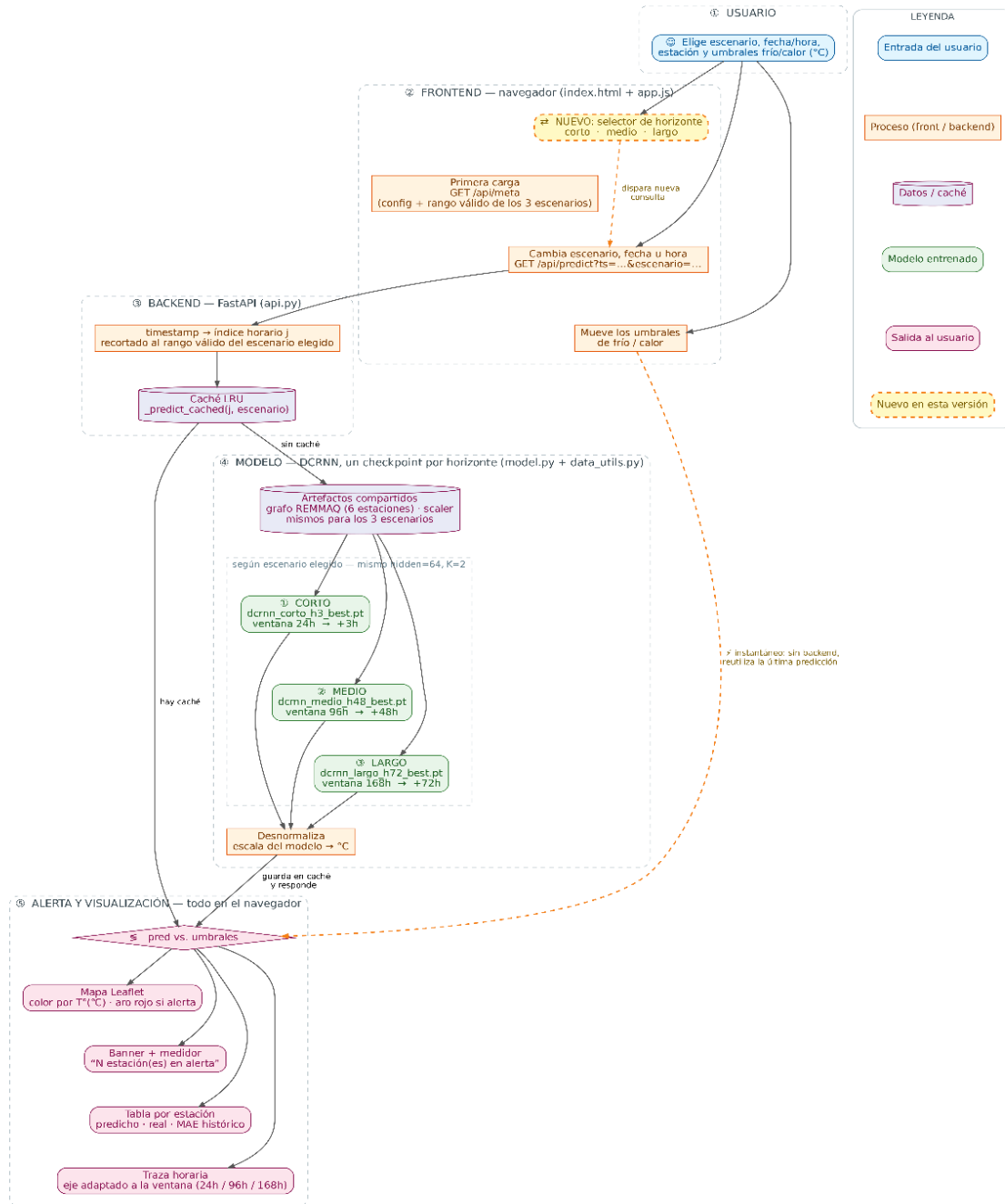
Por último, para facilitar el despliegue y garantizar la reproducibilidad del entorno de ejecución, la aplicación se desplegó mediante Docker, lo que permite encapsular los modelos entrenados, las dependencias del sistema y los componentes necesarios para su ejecución dentro de un entorno homogéneo. Esta estrategia simplifica la implementación del sistema, favorece su portabilidad entre distintos equipos y sienta una base adecuada para futuras implementaciones en contextos productivos.

El funcionamiento integral del sistema se resume en la Figura 33, que ilustra el recorrido de cada solicitud, desde la interacción del usuario con la interfaz hasta la generación de la predicción y la emisión de la alerta correspondiente. El esquema muestra la interacción entre la

interfaz web, el servicio de inferencia encargado de gestionar los modelos entrenados y el módulo responsable de procesar los resultados antes de presentarlos al usuario, lo que evidencia la consolidación de todos los componentes desarrollados durante este estudio en una única plataforma dedicada al seguimiento climático.

Figura 33

Diagrama de Funcionamiento de la Aplicación de Simulación e Inferencia



Nota. Elaboración propia a partir del diseño arquitectónico, pruebas de integración y validación tecnológica del sistema de alerta temprana.

CAPÍTULO V

Conclusiones

El presente estudio demostró que la incorporación de arquitecturas híbridas de aprendizaje profundo constituye una alternativa sólida para el pronóstico de la temperatura en entornos urbanos complejos como el Distrito Metropolitano de Quito. El análisis comparativo entre modelos estadísticos clásicos, redes neuronales de grafos espaciotemporales y modelos fundacionales adaptados a series climáticas evidenció que cada paradigma ofrece ventajas específicas según el horizonte de predicción.

En el corto plazo, las redes neuronales de grafos, DCRNN, mostraron un rendimiento superior para capturar la variabilidad térmica y reducir el error absoluto medio, validando la hipótesis de que la integración explícita de la topología espacial mejora la capacidad predictiva en pronósticos inmediatos. Sin embargo, en horizontes de largo plazo (72h), el modelo SARIMA alcanzó un desempeño estadísticamente equivalente con un menor costo computacional, lo que plantea un balance entre complejidad técnica y eficiencia operativa. Este resultado respalda que la influencia espacial disminuye conforme se amplía el horizonte de predicción, dando mayor protagonismo al componente temporal.

Asimismo, se resaltó la importancia de realizar un análisis y preprocesamiento robusto de datos, mediante filtros físicos y técnicas de imputación bayesiana, así como la construcción de un grafo espacial que permitió representar las interdependencias topográficas y meteorológicas de la región. Estas etapas del desarrollo fueron determinantes para asegurar la validez de los resultados.

Finalmente, el modelo ganador fue integrado en un Sistema de Simulación e Inferencia provisto en un contenedor de Docker, proporcionando una herramienta tecnológica funcional para la toma de decisiones operativas y la gestión del riesgo climático.

De manera complementaria, este estudio establece un marco metodológico reproducible para el desarrollo de modelos de predicción meteorológica basados en inteligencia artificial, ya que la combinación de un preprocesamiento orientado a series temporales, el ajuste fino de modelos fundacionales y la incorporación de un mecanismo de recuperación vectorial constituye una estrategia que puede extenderse a otros problemas de pronóstico ambiental donde exista disponibilidad de información histórica y dependencia espacio-temporal. En conjunto, estos aportes contribuyen tanto al avance científico en predicción climática como al fortalecimiento de la gestión pública frente al riesgo ambiental.

Recomendaciones

En primer lugar, se recomienda profundizar en la optimización de los modelos fundacionales especializados para el dominio meteorológico y de series temporales. Dado que la ejecución actual de ClimaX para el horizonte de largo plazo mostró una convergencia temprana que derivó en predicciones planas, se sugiere realizar un reentrenamiento con una mayor tolerancia en el mecanismo de *early stopping* y una exploración más exhaustiva de hiperparámetros. Asimismo, el marco comparativo debería ampliarse incorporando arquitecturas de vanguardia como Chronos, Lag-Llama o TimeGPT, lo que permitiría evaluar si modelos

preentrenados con diversas estrategias de tokenización se adaptan de mejor manera a la compleja orografía y variabilidad térmica del Distrito Metropolitano de Quito.

En segundo lugar, se sugiere aprovechar la capacidad de las arquitecturas STGNN para manejar la incertidumbre y los vacíos de información sin depender únicamente de una reconstrucción sintética previa. En lugar de aplicar imputación bayesiana para cubrir los huecos de información generados por fallos en las estaciones de la red REMMAQ, futuras implementaciones podrían utilizar máscaras de datos y funciones de pérdida ponderadas (como *MaskedMSELoss*), de modo que el modelo aprenda directamente de la intermitencia instrumental. Este enfoque permitiría reflejar el comportamiento real de las fallas en los sensores y establecer un horizonte mínimo de completitud, evitando el sesgo introducido por valores imputados en estaciones con alta variabilidad.

En tercer lugar, se recomienda ampliar la aplicación de este marco metodológico a otras ciudades de la región interandina con características topográficas y altitud similares a las de Quito, como Bogotá o La Paz. Esta expansión permitiría verificar si la pérdida de la señal espacial observado a partir de las 48 horas corresponde a una propiedad atmosférica general de las urbes andinas.

Finalmente, para la transición del prototipo hacia un entorno operativo robusto, se recomienda implementar estrategias de aprendizaje incremental y actualización periódica de los modelos. Esto permitiría al sistema de simulación e inferencia adaptarse a cambios en los patrones climáticos de largo plazo y reducir la degradación del desempeño.

BIBLIOGRAFÍA

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M.** (2019). *Optuna: A next-generation hyperparameter optimization framework*. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2623–2631). Association for Computing Machinery. <https://doi.org/10.1145/3292500.3330701>
- Ali, A., Naeem, H. M. Y., Ali, R., Asad, M., Heyat, M. B. B., & Alsarhan, T.** (2025). *An efficient algorithm for weather forecasting using causal graph neural network*. ICCCK Transactions on Advanced Computing and Systems, 1(4), 258–274. <https://doi.org/10.62762/TACS.2025.619794>
- Alvarez, C. I., Ulloa Vaca, C. A., & Echeverria Llumipanta, N. A.** (2025). *Machine learning for urban air quality prediction using Google AlphaEarth Foundations satellite embeddings: A case study of Quito, Ecuador*. Remote Sensing, 17(20), Artículo 3472. <https://doi.org/10.3390/rs17203472>
- Bennett, D. A.** (2001). *How can I deal with missing data in my study?* Australian and New Zealand Journal of Public Health, 25(5), 464–469. <https://doi.org/10.1111/j.1467-842X.2001.tb00294.x>
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M.** (2015). *Time series analysis: Forecasting and control (5.ª ed.)*. John Wiley & Sons. <https://doi.org/10.1111/jtsa.12194>
- Campbell, S. D., & Diebold, F. X.** (2005). *Weather forecasting for weather derivatives*. Journal of the American Statistical Association, 100(469), 6–16. <https://doi.org/10.1198/016214504000001057>
- Cerqueira, V., Torgo, L., & Mozetič, I.** (2021). *Model Selection for Time Series Forecasting: Empirical Analysis of Different Estimators* [Preprint]. arXiv preprint arXiv:2104.00584. <https://arxiv.org/abs/2104.00584>
- Chai, T., & Draxler, R. R.** (2014). *Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature*. Geoscientific Model Development, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Dettmers, T., Lewis, M., Shleifer, S., & Zettlemoyer, L.** (2022). *8-bit optimizers via block-wise quantization*. En International Conference on Learning Representations (ICLR). <https://openreview.net/forum?id=N8KaSl00w->

Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2024). *QLoRA: Efficient finetuning of quantized LLMs*. Advances in Neural Information Processing Systems, 36, 10088–10115. <https://arxiv.org/abs/2305.14314>

Diebold, F. X. (2015). *Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests*. Journal of Business & Economic Statistics, 33(1), 1–1. <https://doi.org/10.1080/07350015.2014.983236>

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Housby, N. (2021). *An image is worth 16x16 words: Transformers for image recognition at scale*. En International Conference on Learning Representations (ICLR). <https://openreview.net/forum?id=YicbFdNTtl>

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). *Deep learning for time series classification: A review*. Data Mining and Knowledge Discovery, 33(4), 917–963. <https://doi.org/10.1007/s10618-019-00619-1>

González González, G. L., Rodríguez Martínez, A., Cuevas Figueroa, G., & Lezama López, U. (2018). *Pronóstico de generación eléctrica de corto plazo de un parque eólico, utilizando técnicas numéricas de predicción del clima*. Pistas Educativas, 40(130), 142–151. Tecnológico Nacional de México en Celaya. <https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/download/1681/1338>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <https://www.deeplearningbook.org/>

Grange, S. K. (2014). *Technical note: Averaging wind speeds and directions*. University of Auckland. https://www.researchgate.net/profile/Stuart-Grange/publication/262766424_Technical_note_Averaging_wind_speeds_and_directions/links/54f6184f0cf27d8ed71d5bd4/Technical-note-Averaging-wind-speeds-and-directions.pdf

Gruver, N., Finzi, M., Qiu, S., & Wilson, A. G. (2024). *Large language models are zero-shot time series forecasters*. Advances in Neural Information Processing Systems, 36, 19623–19645. <https://doi.org/10.48550/arXiv.2310.07820>

Guerrero Muñoz, G. J., & Armijos Toro, L. M. (2025). *Pronóstico de la contaminación por material particulado PM10 con series de tiempo*. Revista Ciencia y Tecnología, 18(2), 86–96. <https://doi.org/10.18779/cyt.v18i2.944>

Hendrycks, D., & Gimpel, K. (2016). *Gaussian error linear units (GELUs)* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.1606.08415>

- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). *LoRA: Low-rank adaptation of large language models*. En International Conference on Learning Representations (ICLR). <https://openreview.net/forum?id=nZeVKeeFYf9>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>
- Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Renard Lavaud, L., Lachaux, M.-A., Stock, P., Le Scao, T., Lavril, T., Wang, T., Lacroix, T., & El Sayed, W. E. (2023). *Mistral 7B* [Preprint]. arXiv. <https://arxiv.org/abs/2310.06825>
- Jiang, J., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. de las, & Sayed, W. E. (2026). *A parallel attention-based framework for multi-step multivariate runoff forecasting in mountainous watersheds*. Journal of Hydrology: Regional Studies, 63, 103045. <https://doi.org/10.1016/j.ejrh.2025.103045>
- Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., Chen, J., & Wen, Q. (2024). *Time-LLM: Time series forecasting by reprogramming large language models*. En International Conference on Learning Representations (ICLR). <https://openreview.net/forum?id=Bl8z7ja18d>
- Johnson, J., Douze, M., & Jégou, H. (2019). *Billion-scale similarity search with GPUs*. IEEE Transactions on Big Data, 7(3), 535–547. <https://doi.org/10.1109/TBDATA.2019.2921572>
- Kalamkar, D., Mudigere, D., Mellempudi, N., Das, D., Banerjee, A., Avancha, S., Vuduc, R., Dubey, P., & Smelyanskiy, M. (2019). *A study of BFLOAT16 for deep learning training* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.1905.12322>
- Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., & Hoyer, S. (2024). *Neural general circulation models for weather and climate*. Nature, 632(8024), 1060–1066. <https://doi.org/10.1038/s41586-024-07744-y>
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirsberger, P., Fortunato, M., Pritzel, A., Ravuri, S., Ewalds, T., Alet, F., Eaton-Rosen, Z., Hu, W., Merzbacher, A., Hoyer, S., Battaglia, P., Wu, E., & Kohli, P. (2023). *Learning skillful medium-range global weather forecasting with GraphCast*. Science, 382(6677), 1416–1421. <https://doi.org/10.1126/science.adi2336>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-augmented generation for*

- knowledge-intensive NLP tasks*. Advances in Neural Information Processing Systems, 33, 9459–9474. <https://dl.acm.org/doi/abs/10.5555/3495724.3496517>
- Li, Y., Yu, R., Shahabi, C., & Liu, Y.** (2018). *Diffusion convolutional recurrent neural network: Data-driven traffic forecasting*. En International Conference on Learning Representations (ICLR). <https://openreview.net/forum?id=SJiHXGWAZ>
- Loshchilov, I., & Hutter, F.** (2017). *SGDR: Stochastic gradient descent with warm restarts*. En International Conference on Learning Representations (ICLR). <https://openreview.net/forum?id=Skq89Scxx>
- Ma, M., Xie, P., Teng, F., Wang, B., Ji, S., Zhang, J., & Li, T.** (2023). *HiSTGNN: Hierarchical spatio-temporal graph neural network for weather forecasting*. Information Sciences, 648, Artículo 119580. <https://doi.org/10.1016/j.ins.2023.119580>
- Mancheno, G., & Jorquera, H.** (2025). *High spatial resolution WRF-Chem modeling in Quito, Ecuador*. Environmental Science: Advances, 4, 1310–1332. <https://doi.org/10.1039/D5VA00059A>
- Meta.** (2024). *The Llama 3 herd of models* [Preprint]. arXiv. <https://arxiv.org/abs/2407.21783>
- Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., & Wu, H.** (2018). *Mixed precision training* [Preprint]. arXiv. <https://openreview.net/forum?id=r1gs9JgRZ>
- Morera Campos, J. R.** (2025). *Predicción de series temporales meteorológicas mediante técnicas de aprendizaje profundo* [Trabajo de fin de grado, Universidad de La Laguna]. <http://riull.ull.es/xmlui/handle/915/43055>
- Muñoz Mandujano, M.** (2024). *Algoritmo para pronóstico de precipitaciones en tiempo real basado en el modelo predictivo CRHUDA* [Tesis de doctorado, Universidad Autónoma de Querétaro]. <https://ri-ng.uaq.mx/handle/123456789/10423>
- Naquintero Garavito, N. A.** (2025). *Elaboración de un modelo predictivo de temperatura en Barrancabermeja mediante machine learning* [Trabajo de grado, Universidad Nacional Abierta y a Distancia]. <https://repository.unad.edu.co/handle/10596/74388>
- Nguyen, Q. V., Delgado Fernandez, J., & Potenciano Menci, S.** (2025). *Spatiotemporal Graph Neural Networks in short-term load forecasting: Does adding Graph Structure in Consumption Data Improve Predictions?* arXiv. <https://doi.org/10.48550/arXiv.2502.12175>

Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., & Grover, A. (2023). *ClimaX: A foundation model for weather and climate*. En Proceedings of the 40th International Conference on Machine Learning (ICML) (pp. 25904–25931). PMLR.

<https://proceedings.mlr.press/v202/nguyen23a.html>

Nurcahyo, A. W., & Putranto, B. P. D. (2026). *Spatio-temporal graph neural network for solar irradiance prediction: A case study in Nganjuk, Indonesia*. Journal of Information Systems and Informatics, 8(1), 617–635. <https://doi.org/10.63158/journalisi.v8i1.1424>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.

<https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

Quispe Galindo, L. (2024). *Modelo de predicción climatológica con inteligencia artificial en la región Ica* [Tesis de grado, Universidad Nacional San Luis Gonzaga].

<https://hdl.handle.net/20.500.13028/5629>

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). *Green AI*. Communications of the ACM, 63(12), 54–63. <https://doi.org/10.1145/3381831>

Secretaría de Ambiente del Distrito Metropolitano de Quito. (2024). *Reporte anual de calidad del aire en el DMQ-2024 (Informe Técnico F-PG-09-07, Ed. 03)*. Municipio de Quito. <https://ambiente.quito.gob.ec/wp-content/uploads/2025/12/INFORME-DE-LA-CALIDAD-DEL-AIRE-EN-EL-DMQ-2024.pdf>

Secretaría de Ambiente del Distrito Metropolitano de Quito. (2025). *Marco QA/QC para el procesamiento y validación de datos de la REMMAQ*. Municipio de Quito.

https://datosambiente.quito.gob.ec/datos/Marco_QAQC_REMMAQ.pdf

Secretaría de Ambiente del Distrito Metropolitano de Quito. (2026). *Red de Monitoreo de Calidad del Aire de Quito (REMQA): Histórico de datos meteorológicos y de calidad del aire* [Dataset]. Municipio de Quito. <https://datosambiente.quito.gob.ec/datos/>

Smith, S. L., Kindermans, P. J., Ying, C., & Le, Q. V. (2018). *Don't decay the learning rate, increase the batch size* [Preprint]. arXiv. <https://openreview.net/forum?id=B1Yy1BxCZ>

Tang, H., Yang, H., & Zhang, W. (2025). *DAHG: A dynamic augmented heterogeneous graph framework for precipitation forecasting with incomplete data*. Information, 16(11), Artículo 946. <https://doi.org/10.3390/info16110946>

U.S. Environmental Protection Agency. (2017). *Quality assurance handbook for air pollution measurement systems, Volume II: Ambient air quality monitoring program* (EPA-454/B-17-001). Office of Air Quality Planning and Standards.
<https://www3.epa.gov/ttnamti1/files/ambient/pm25/qa/Final%20Handbook%20Document%2017.pdf>

Van Buuren, S. (2018) *Flexible Imputation of Missing Data* (2nd ed.). CRC Press, Boca Raton.
<https://doi.org/10.1201/9780429492259>

World Meteorological Organization. (2011). *Guide to meteorological instruments and methods of observation* (WMO-No. 8, 7.^a ed.). Ginebra, Suiza.

World Meteorological Organization. (2014). *Guide to meteorological instruments and methods of observation* (WMO-No. 8). Ginebra, Suiza.

Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020). *Connecting the dots: Multivariate time series forecasting with graph neural networks. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 753–763). Association for Computing Machinery. <https://doi.org/10.1145/3394486.3403118>

Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2022). *Are transformers effective for time series forecasting?* *Advances in Neural Information Processing Systems*, 36, 11121–11138.
<https://doi.org/10.48550/arXiv.2205.13504>

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2019). *T-GCN: A temporal graph convolutional network for traffic prediction. IEEE Transactions on Intelligent Transportation Systems*, 21(9), 3848–3858. <https://doi.org/10.1109/TITS.2019.2935152>

Zhu, J., Song, Y., Zhao, L., & Li, H. (2021). *A3T-GCN: Attention temporal graph convolutional network for traffic forecasting. ISPRS International Journal of Geo-Information*, 10(7), Artículo 485. <https://doi.org/10.3390/ijgi10070485>

APÉNDICE A

Repositorio de Código Fuente

El código fuente desarrollado para la implementación de las arquitecturas híbridas de aprendizaje profundo utilizadas en la presente investigación se encuentra disponible en un repositorio público de GitHub. El repositorio contiene los scripts, notebooks y demás archivos necesarios para la preparación de los datos, construcción, entrenamiento, validación y evaluación de los modelos desarrollados, así como los procedimientos empleados para la generación de los resultados experimentales.

Referencia del recurso:

Jácome Guanoluisa, R. X., Molineros Negrete, P. A., & Márquez Chávez, C. A. (2026).

Hybrid forecasting architectures [Source code]. GitHub.

<https://github.com/pablolinerosn/hybrid-forecasting-architectures.git>

APÉNDICE B

Repositorio de la Base de Datos y Artefactos Generados

La base de datos procesada y los artefactos generados durante las diferentes etapas de la investigación se encuentran disponibles en un repositorio de Google Drive. Este repositorio contiene los conjuntos de datos utilizados para el desarrollo de los modelos, archivos derivados

del proceso de preparación y transformación de los datos, modelos entrenados, resultados experimentales y demás materiales complementarios empleados durante la ejecución de la metodología propuesta.

Referencia del recurso:

Jácome Guanoluisa, R. X., Molineros Negrete, P. A., & Márquez Chávez, C. A. (2026).

Repositorio de la base de datos y artefactos generados [Data set and supplementary materials].

Google Drive.

<https://drive.google.com/drive/folders/1UQry9P621pf8Y2eRoEzirUJY8c8ttNTl?usp=sharing>