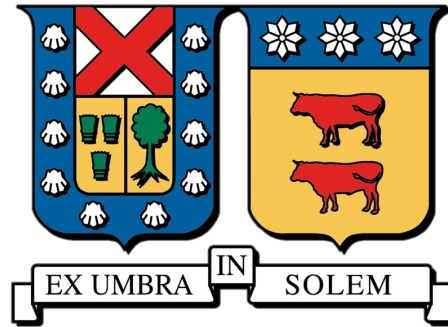




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS

**APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA
DOWNSCALING ESTADÍSTICO Y ADAPTACIÓN LOCAL DE
PRONÓSTICOS DEL MODELO FUNDACIONAL AURORA A
OBSERVACIONES DE UNA ESTACIÓN METEOROLÓGICA EN
VALPARAÍSO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

AUTOR

EDUARDO TAPIA SEPÚLVEDA

PROFESOR GUÍA

Dr. Felipe Escudero

PROFESOR CO-REFERENTE

Dr. Mijail Littin

VALPARAÍSO DE CHILE, 7 DE AGOSTO 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA DOWNSCALING ESTADÍSTICO Y ADAPTACIÓN LOCAL DE PRONÓSTICOS DEL MODELO FUNDACIONAL AURORA A OBSERVACIONES DE UNA ESTACIÓN METEOROLÓGICA EN VALPARAÍSO

Nombre del candidato(a): Eduardo Enrique Tapia Sepúlveda

Carrera / Grado: Ingeniería civil industrial

Campus: Casa Centra **Departamento:** Industrias

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Felipe Andrés Escudero Barros, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 28/08/2026

Firma:

Estudiante o Candidato(a):

Fecha: 28/08/2026

Firma:

Tabla de Contenidos

1. Agradecimientos	4
2. Resumen	5
3. Abstract	6
4. Introducción	7
5. Problema de Investigación	9
6. Objetivos	10
6.1. Objetivo General	10
6.2. Objetivos Específicos	10
7. Alcance	11
8. Marco Teórico	12
8.1. La atmósfera como sistema caótico y sus limitaciones predictivas	12
8.2. Predicción meteorológica y modelación atmosférica	12
8.3. Inteligencia artificial aplicada a meteorología	13
8.4. Modelos fundacionales del sistema terrestre	13
8.5. Modelo Aurora	14
8.6. Resolución espacial y temporal en predicción meteorológica	17
8.7. Downscaling meteorológico	18
8.7.1. Downscaling dinámico	18

8.7.2. Downscaling estadístico	18
8.8. Corrección de sesgo y refinamiento temporal	19
8.9. Estaciones meteorológicas y observaciones locales	19
9. Metodología	21
9.1. Enfoque general	21
9.2. Datos de entrada	22
9.2.1. Datos de reanálisis ERA5	22
9.2.2. Generación de pronósticos con Aurora	23
9.2.3. Observaciones meteorológicas locales	26
9.3. Integración y construcción del conjunto de datos	29
9.4. Modelo de adaptación local	33
9.5. Evaluación y cuantificación de la mejora	36
10. Resultados	38
10.1. Verificación del comportamiento autorregresivo de Aurora	38
10.2. Análisis exploratorio	42
10.2.1. Ciclo diurno y estructura horaria del sesgo	42
10.2.2. Error por horizonte de pronóstico	44
10.2.3. Distribuciones y eventos extremos	45
10.2.4. Pérdida por resolución temporal	46
10.2.5. Persistencia del error y dependencia de las condiciones	48
10.2.6. Velocidad del viento	50
10.2.7. Humedad relativa	52

10.3. Selección del modelo en validación cruzada	53
10.3.1. Desempeño de las líneas base	54
10.3.2. Búsqueda de hiperparámetros	54
10.3.3. Interpretación de los modelos	55
10.3.4. Ensamble RF–Ridge	56
10.4. Evaluación final en el conjunto de prueba	57
10.4.1. Métricas globales	58
10.4.2. Comportamiento horario y episodio extremo	59
10.4.3. Análisis de residuos	61
10.5. Sistema multivariable	62
10.5.1. Desempeño por variable	62
10.5.2. Interpretación por variable	63
10.6. Síntesis	64
 11. Conclusiones	 66
 12. Limitaciones	 69
 13. Anexos	 70

1. Agradecimientos

A mi Sandrita, mi mamá, por criarme desde el amor y la bondad y siempre mantener mi corazón alejado de los rencores y el orgullo.

A mi Tío Juan, por entregarme el cariño, la sabiduría y el abrazo paternal que se heredará a los míos. A mi tía Mónica, mi otra mamá, por su calidez y suavidad, que sin dureza te ayuda a retomar el camino. A mi primo, por el día a día.

Esta es, sin duda, la única hoja en todo el documento que, desde que entré a la universidad, deseaba escribir. Han dejado una huella enorme en mi corazón.

2. Resumen

El presente trabajo desarrolla y evalúa un sistema de adaptación local que corrige, mediante aprendizaje automático, los pronósticos del modelo fundacional Aurora hacia las observaciones de una estación meteorológica de bajo costo instalada en el campus de la Universidad Técnica Federico Santa María. Los modelos de predicción de alcance global entregan sus resultados sobre una grilla extensa y a intervalos de varias horas, por lo que no representan con exactitud las condiciones de un punto específico ni su evolución a resolución horaria. Para abordar esta limitación, el sistema aprende la diferencia entre el pronóstico y la observación, corrigiendo de manera conjunta la falta de representatividad del punto y la limitada resolución temporal. La selección de los modelos se realiza mediante validación cruzada agrupada por días, reservando un conjunto de prueba independiente para la evaluación final.

Sobre dicho conjunto, el sistema reduce el error absoluto medio respecto del pronóstico sin corregir en un 17,2 % en temperatura, 23,2 % en humedad relativa, 63,8 % en velocidad del viento y 83,3 % en presión. Para la temperatura, el coeficiente de determinación se eleva de 0,23 a 0,56. Al descomponer estas cifras frente a correcciones más simples, se observa que una parte importante de las mejoras más altas proviene de procedimientos que no requieren aprendizaje, mientras que el aporte propio del modelo se reparte de forma más homogénea entre las variables, del orden de 12 a 20 % en temperatura, humedad y viento. Los resultados muestran que es posible adaptar los pronósticos de un modelo fundacional a un punto de observación específico empleando recursos computacionales convencionales, con una corrección además interpretable en términos físicos.

3. Abstract

Foundation models for weather forecasting, such as Aurora, provide predictions at grid scale (0.25°), but do not resolve the discrepancy between these predictions and the conditions observed at a specific point. This work develops a statistical downscaling system that uses machine learning to correct Aurora forecasts toward observations from a Davis Vantage Pro 2 weather station located on the UTFSM campus in Valparaíso, addressing both the lack of local representativeness and the 6-hour temporal resolution. The problem is formulated as a residual correction applied to the interpolated forecast, using eleven predictors per hour, with independent models trained for each variable (Ridge, Random Forest, and an ensemble of both), validated on a daily basis and evaluated on a sealed test set. Compared with the uncorrected forecast, the system reduces the error by 17.2 % for temperature, 23.2 % for humidity, 63.8 % for wind, and 83.3 % for pressure, with the contribution of machine learning alone being more homogeneous, ranging from 12 % to 20 %. The results show that a foundation model can be adapted to a local point using conventional computational resources, without retraining the foundation model.

4. Introducción

La predicción meteorológica ha experimentado en los últimos años una transformación impulsada por la aparición de modelos fundacionales de aprendizaje profundo. Entrenados sobre grandes volúmenes de datos atmosféricos históricos, estos modelos han alcanzado una precisión comparable, y en varios casos superior, a la de los sistemas numéricos tradicionales, con un costo de inferencia considerablemente menor. Entre ellos, el modelo Aurora (Bodnar et al., 2024) representa uno de los desarrollos más recientes orientados a la predicción integral del sistema terrestre, capaz de generar pronósticos globales a partir de un estado inicial de la atmósfera.

A pesar de su precisión, estos modelos entregan sus pronósticos sobre una grilla de resolución espacial limitada y a intervalos de varias horas. El valor asociado a cada celda corresponde a un promedio sobre un área extensa, que no coincide necesariamente con las condiciones de un punto específico ubicado dentro de ella, y la separación entre predicciones sucesivas deja sin describir la evolución que ocurre dentro de cada intervalo. Numerosas aplicaciones prácticas requieren, en cambio, información referida a un sitio concreto y a una resolución temporal más fina, lo que restringe el uso directo de estos pronósticos para la toma de decisiones a escala local.

El presente trabajo aborda esa brecha mediante un sistema de adaptación local que ajusta los pronósticos del modelo global a las observaciones de una estación meteorológica específica. El enfoque emplea aprendizaje automático para aprender la diferencia entre el pronóstico y la observación, de modo de corregir de forma simultánea la falta de representatividad del punto y la limitada resolución temporal. Al aprender únicamente la corrección sobre el pronóstico ya existente, el sistema conserva la información de gran escala que aporta el modelo global y concentra el aprendizaje en la discrepancia local, lo que constituye una formulación eficiente y de bajo riesgo.

La combinación de un modelo fundacional reciente con una única estación de costo accesible, ejecutada sobre recursos computacionales convencionales, constituye una configuración escasamente explorada en la literatura. El propósito general de este trabajo es desarrollar y evaluar dicho sistema de corrección y refinamiento, cuantificando la mejora alcanzada respecto del pronóstico sin corregir y de alternativas de corrección más simples, con el fin de distinguir el aporte propio del aprendizaje automático del que se obtendría mediante procedimientos elementales. De este modo, el trabajo busca aportar evidencia sobre la viabilidad de acercar los pronósticos de los modelos de inteligencia artificial al nivel de detalle que demandan las aplicaciones locales.

El documento se organiza de la siguiente manera. El Capítulo de Problema de Investigación delimita la necesidad que motiva el estudio, mientras que los capítulos de Objetivos y Alcance establecen las metas y los límites del trabajo. El Marco Teórico presenta los fundamentos de los modelos fundacionales de predicción y de las técnicas de corrección empleadas. La Metodología describe la construcción del conjunto de datos, la formulación del problema y los modelos evaluados. El capítulo de Resultados expone el desempeño del sistema sobre cada variable y la descomposición de la mejora frente a las líneas base. Finalmente, las Conclusiones recogen los hallazgos principales y el capítulo de Limitaciones acota su alcance y señala las líneas de continuación.

5. Problema de Investigación

Los modelos meteorológicos globales basados en inteligencia artificial han demostrado una alta capacidad para representar la dinámica atmosférica a gran escala. Entre ellos, el modelo Aurora desarrollado por Microsoft Research AI for Science genera predicciones globales con alta precisión y bajo costo computacional (Bodnar et al., 2024). Sin embargo, su resolución espacial aproximada de $0,25^\circ$ y su resolución temporal de 6 horas limitan la representación de las condiciones atmosféricas locales.

Estas limitaciones dificultan capturar fenómenos influenciados por características geográficas, topográficas y microclimáticas específicas, de modo que las predicciones globales pueden no reflejar adecuadamente las condiciones observadas en estaciones meteorológicas locales.

Frente a ello surge la necesidad de metodologías de downscaling meteorológico que adapten la información global de Aurora a condiciones locales específicas. Para ello es posible utilizar registros observacionales de estaciones meteorológicas y entrenar un modelo que aprenda la relación entre las predicciones globales y el comportamiento atmosférico local.

En este contexto, el problema de investigación consiste en determinar de qué manera la integración entre las predicciones de Aurora y los datos de una estación local puede utilizarse para desarrollar un modelo de downscaling estadístico y adaptación local capaz de refinar la resolución temporal de los pronósticos y corregir su representatividad en el punto de observación, mejorando así la representación de las variables atmosféricas a escala local.

6. Objetivos

6.1. Objetivo General

Desarrollar un modelo de downscaling estadístico y adaptación local basado en la integración de predicciones del modelo Aurora y datos observacionales de una estación meteorológica local, con el fin de mejorar la representación de las variables atmosféricas a escala local.

6.2. Objetivos Específicos

- Generar pronósticos globales con el modelo Aurora en un esquema operacional de inicialización diaria sobre la región de estudio, verificando la coherencia física de su despliegue autorregresivo.
- Construir un conjunto de datos supervisado que integre las salidas de Aurora con las observaciones de la estación, homogeneizando unidades, referencias temporales y resolución.
- Caracterizar el error sistemático y condicional de Aurora respecto de las observaciones locales, identificando su estructura horaria, su persistencia y sus variables explicativas.
- Entrenar y seleccionar, mediante validación cruzada, modelos de corrección que aborden de forma conjunta la corrección del sesgo local y el refinamiento temporal de los pronósticos.
- Evaluar el desempeño del modelo sobre un conjunto de prueba independiente y cuantificar la mejora respecto de líneas base de complejidad creciente.

7. Alcance

La presente investigación desarrolla un sistema de adaptación local de pronósticos meteorológicos mediante aprendizaje automático, sobre las predicciones del modelo fundacional Aurora en su versión preentrenada de $0,25^\circ$ de resolución y, como referencia observacional, los registros de una estación Davis Vantage Pro 2 ubicada en el campus de la Universidad Técnica Federico Santa María, en Valparaíso.

El estudio abarca cuatro variables atmosféricas: temperatura, humedad relativa, presión barométrica y velocidad del viento. A partir de los pronósticos de Aurora, de resolución temporal de 6 horas, el sistema realiza dos tareas conjuntas: la corrección del sesgo sistemático del modelo global respecto del punto de observación y el refinamiento temporal de las predicciones hacia una resolución de 1 hora. Ambas se abordan con modelos de aprendizaje automático entrenados sobre la relación entre las salidas de Aurora y las observaciones locales, bajo un esquema operacional de inicialización diaria con horizontes de hasta 24 horas.

El estudio presenta tres limitaciones declaradas. Primero, la adaptación se restringe a la ubicación de la estación: al tratarse de una corrección de grilla a punto, los resultados representan el sitio analizado y no son directamente generalizables a otras zonas. Segundo, el período comprende 40 días de verano (diciembre de 2025 a febrero de 2026), por lo que los modelos caracterizan los regímenes propios de esa estación del año. Tercero, no se contempla el reentrenamiento ni el ajuste fino de Aurora, el desarrollo de modelos globales, la generación de campos de alta resolución espacial ni simulaciones atmosféricas dinámicas: el aporte es la adaptación estadística de predicciones ya generadas, ejecutable con recursos computacionales convencionales.

8. Marco Teórico

8.1. La atmósfera como sistema caótico y sus limitaciones predictivas

La atmósfera terrestre es un sistema dinámico altamente no lineal y caótico, con fuerte sensibilidad a las condiciones iniciales: pequeñas perturbaciones o errores en el estado inicial se amplifican con el tiempo y generan diferencias significativas en los pronósticos futuros. Por esta naturaleza caótica, la predicción meteorológica presenta limitaciones intrínsecas incluso con modelos numéricos avanzados, observaciones de alta resolución y grandes capacidades computacionales (Zhang et al., 2019).

Zhang et al. (2019) demostraron que existe un límite práctico de predictibilidad para los fenómenos de latitudes medias, situado actualmente en torno a los diez días para pronósticos deterministas útiles; los fenómenos de menor escala, como tormentas convectivas o precipitaciones intensas, presentan una predictibilidad aún más reducida por su rápida evolución espacial y temporal. Estas limitaciones han motivado el desarrollo de nuevas metodologías para mejorar la representación de los procesos atmosféricos y las capacidades de pronóstico.

8.2. Predicción meteorológica y modelación atmosférica

La predicción meteorológica es una herramienta fundamental para la planificación y la toma de decisiones en áreas como agricultura, gestión de riesgos, transporte y monitoreo climático. Tradicionalmente se ha desarrollado mediante modelos numéricos de predicción del tiempo (*Numerical Weather Prediction*, NWP), que representan la evolución de la atmósfera mediante ecuaciones diferenciales parciales sobre rejillas discretizadas. Sin embargo, estos métodos presentan importantes limitaciones, por ejemplo, la generación de pronósticos puede requerir horas de

procesamiento en supercomputadores y dependen de parametrizaciones para representar fenómenos no resueltos explícitamente, lo que introduce errores por aproximaciones y simplificaciones físicas.

8.3. Inteligencia artificial aplicada a meteorología

El avance reciente de la inteligencia artificial, en particular del aprendizaje profundo, ha impulsado modelos meteorológicos basados en datos capaces de aprender relaciones espacio-temporales directamente desde grandes volúmenes de datos atmosféricos (Bi et al., 2023). A diferencia de los métodos tradicionales, según Bodnar et al. (2024) estos enfoques reducen considerablemente los tiempos de cálculo mediante redes neuronales profundas y unidades de procesamiento gráfico (GPU). Entre los modelos más relevantes destacan Pangu-Weather y Aurora, con resultados competitivos e incluso superiores a algunos sistemas operacionales. En particular, Pangu-Weather introdujo arquitecturas tridimensionales basadas en transformadores capaces de representar relaciones verticales entre niveles atmosféricos (Bi et al., 2023), mientras que Aurora fue desarrollado como un modelo fundacional orientado a la modelación integral del sistema terrestre (Bodnar et al., 2024).

8.4. Modelos fundacionales del sistema terrestre

Los modelos fundacionales constituyen una nueva aproximación a la modelación del sistema terrestre basada en inteligencia artificial. Se entrenan con volúmenes masivos de datos de reanálisis, observaciones satelitales y simulaciones climáticas, y a diferencia de los modelos especializados en tareas específicas, buscan construir representaciones generales del sistema terrestre

que luego pueden adaptarse a múltiples aplicaciones mediante *fine-tuning* (Bodnar et al., 2024). Dentro de este enfoque, Aurora representa uno de los desarrollos más recientes orientados a la predicción integral del sistema terrestre, incorporando información atmosférica, oceánica y climática en una arquitectura unificada (el detalle de su esquema codificador–decodificador se incluye en la Figura 24 de los Anexos). Esta transferencia de conocimiento entre tareas y dominios mejora la capacidad predictiva y reduce significativamente los costos computacionales respecto de la simulación numérica tradicional.

8.5. Modelo Aurora

Aurora es un modelo fundacional para el sistema terrestre desarrollado por Microsoft Research AI for Science, diseñado para realizar predicciones atmosféricas mediante aprendizaje profundo a gran escala. Con aproximadamente 1,3 mil millones de parámetros, fue preentrenado con más de un millón de horas de datos heterogéneos provenientes de reanálisis meteorológicos, pronósticos operacionales, observaciones satelitales y simulaciones climáticas, buscando representaciones generales transferibles a múltiples tareas (Bodnar et al., 2024).

El funcionamiento general del modelo se ilustra en la Figura 1. Aurora recibe como entrada el estado del sistema terrestre, compuesto por variables atmosféricas definidas en niveles de presión, variables de superficie y variables estáticas, y produce el estado pronosticado seis horas hacia el futuro. Una característica distintiva es su capacidad de ingerir entradas con distintas resoluciones espaciales, números de niveles de presión y conjuntos de variables (Bodnar et al., 2024), flexibilidad que habilita el paradigma de un único modelo base adaptable a diferentes dominios. La aplicación autorregresiva, en la cual cada estado pronosticado se reutiliza como entrada del paso

siguiente, permite extender el horizonte en pasos sucesivos de 6 horas.

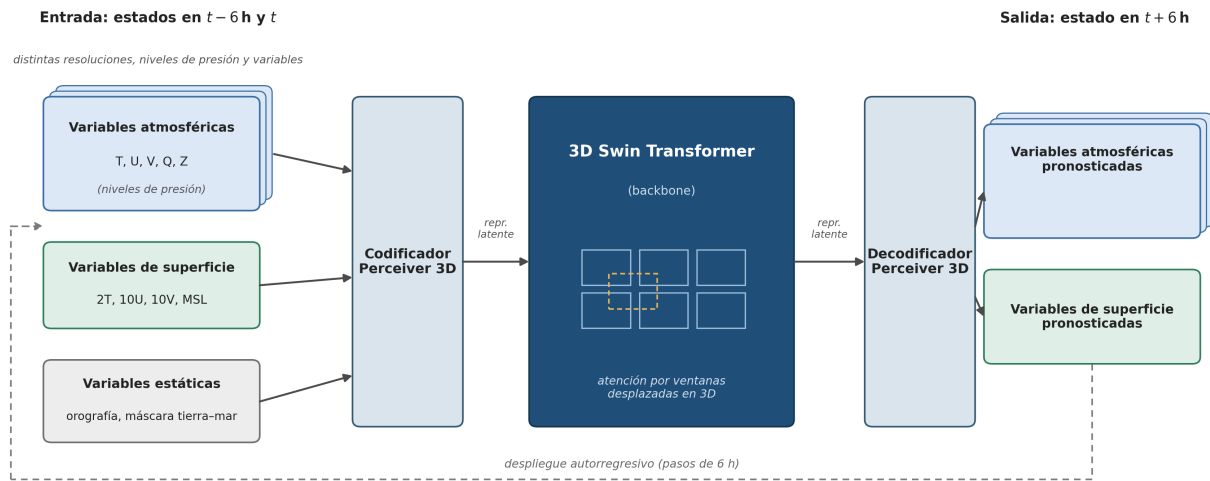


Figura 1: Arquitectura general del modelo Aurora. Las entradas heterogéneas (variables atmosféricas en niveles de presión, variables de superficie y variables estáticas) son transformadas por un codificador Perceiver 3D en una representación latente, procesada por un *backbone* 3D Swin Transformer mediante atención por ventanas desplazadas, y reconstruida por un decodificador Perceiver 3D para producir el estado del sistema en $t + 6$ h. El despliegue autorregresivo permite extender el horizonte de pronóstico en pasos sucesivos de 6 horas. Elaboración propia, basada en Bodnar et al. (2024).

Internamente, Aurora se estructura en tres componentes secuenciales, cuya composición como red neuronal profunda se detalla en la Figura 2. Un codificador de tipo Perceiver 3D divide el estado de entrada en parches tridimensionales, los proyecta a vectores de *embedding* mediante capas lineales específicas por variable y los comprime, mediante atención cruzada, en una representación latente con un número reducido de niveles verticales (Bodnar et al., 2024). Dicha representación es procesada por el *backbone*, un 3D Swin Transformer con estructura U-Net de 48 capas organizadas en tres etapas simétricas de reducción y aumento de resolución con conexiones de salto, que captura la dinámica atmosférica en múltiples escalas; cada capa aplica autoatención local por ventanas tridimensionales, desplazadas de forma alternada para propagar información entre regiones vecinas. Finalmente, un decodificador Perceiver 3D revierte las operaciones del codifica-

dor y reconstruye el estado pronosticado sobre la grilla latitud–longitud y los niveles de presión requeridos.

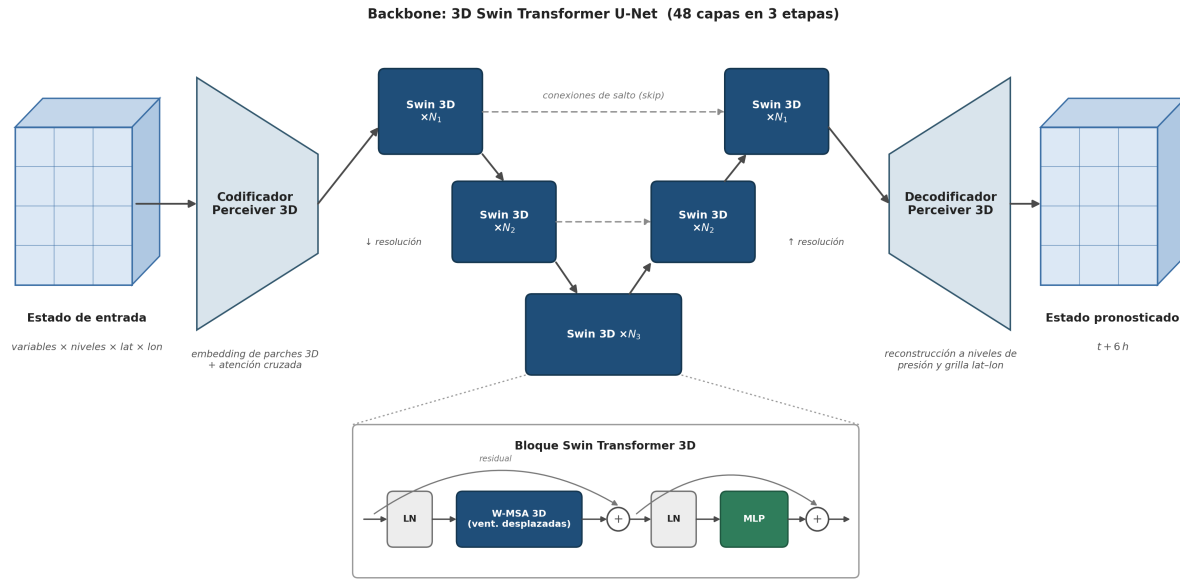


Figura 2: Estructura de red neuronal profunda del modelo Aurora. El estado atmosférico de entrada, representado como un tensor tridimensional (variables $\times$ niveles de presión $\times$ latitud $\times$ longitud), es comprimido por el codificador Perceiver 3D en una representación latente y procesado por un *backbone* de tipo 3D Swin Transformer U-Net, compuesto por 48 capas organizadas en tres etapas simétricas de reducción y aumento de resolución con conexiones de salto, lo que permite el procesamiento multiescala de la dinámica atmosférica. El detalle inferior muestra la composición interna de un bloque Swin Transformer 3D: normalización de capa (LN), autoatención multicabezal por ventanas desplazadas (W-MSA 3D) y perceptrón multicapa (MLP), con conexiones residuales. El decodificador Perceiver 3D reconstruye el estado pronosticado sobre la grilla latitud–longitud y los niveles de presión deseados. Elaboración propia, basada en Bodnar et al. (2024).

El entrenamiento se organiza en dos fases: un preentrenamiento que aprende representaciones generales a partir de diversas fuentes, y un ajuste fino que especializa el modelo en escenarios operacionales a distintas resoluciones. Sus autores reportan configuraciones para química atmosférica y calidad del aire a $0,4^\circ$, oleaje y seguimiento de ciclones tropicales a $0,25^\circ$, y predicción meteorológica de alta resolución a $0,1^\circ$ que, según Bodnar et al. (2024), presentan resultados competitivos e incluso superiores a sistemas operacionales tradicionales y un costo computacional

menor.

En su configuración de propósito general, Aurora opera a una resolución espacial de $0,25^\circ$ y genera pronósticos en pasos de 6 horas. Esta resolución, heredada en buena medida de los datos de reanálisis empleados en su entrenamiento, resulta adecuada para la dinámica atmosférica de gran escala, pero limita la representación de fenómenos locales y de pequeña escala (Park et al., 2026), limitación que motiva precisamente las técnicas de adaptación local abordadas en este trabajo.

No obstante sus ventajas, Aurora presenta limitaciones relevantes. Su carácter de modelo basado en datos dificulta la interpretabilidad de sus predicciones: el conocimiento atmosférico reside en los parámetros aprendidos y no en ecuaciones físicas auditables. Su entrenamiento requiere volúmenes masivos de datos y recursos computacionales al alcance solo de grandes organizaciones, lo que impide a usuarios finales reentrenar o ajustar el modelo (Bodnar et al., 2024). Su despliegue, más económico que el de los sistemas numéricos tradicionales, demanda de todos modos hardware especializado (GPU) para alcanzar tiempos de inferencia prácticos. Estas consideraciones condicionan el diseño de un refinamiento local: la corrección debe operar por fuera del modelo, sobre sus salidas, y con métodos ejecutables en recursos convencionales.

8.6. Resolución espacial y temporal en predicción meteorológica

La resolución espacial y temporal determina el nivel de detalle con que los modelos representan los fenómenos atmosféricos. En modelos globales, la resolución espacial suele expresarse en grados de latitud y longitud (una resolución de $0,25^\circ$ corresponde a celdas de aproximadamente 25 km), y la resolución temporal define el intervalo entre predicciones, típicamente de 1, 3 o 6 horas en sistemas globales.

Los modelos meteorológicos basados en aprendizaje profundo como Aurora, Pangu-Weather y FuXi, obtienen pronósticos globales competitivos con costos computacionales muy inferiores a los de la predicción numérica tradicional (Bodnar et al., 2024; Bi et al., 2023; Park et al., 2026). Sin embargo, permanecen limitados por la resolución de sus datos de entrenamiento, en particular ERA5 (0,25°), lo que dificulta representar fenómenos locales y de pequeña escala como microclimas urbanos, precipitaciones convectivas, circulaciones costeras y efectos topográficos complejos.

8.7. Downscaling meteorológico

El downscaling meteorológico agrupa las técnicas para aumentar la resolución espacial y temporal de pronósticos de baja resolución, permitiendo representar fenómenos locales que los modelos globales no capturan adecuadamente.

8.7.1. Downscaling dinámico

El downscaling dinámico emplea modelos atmosféricos regionales para refinar pronósticos globales mediante simulaciones numéricas de alta resolución sobre regiones específicas; entre los más utilizados destaca WRF (*Weather Research and Forecasting Model*). Aunque representa los procesos físicos y los efectos topográficos con mayor precisión, requiere altos costos computacionales.

8.7.2. Downscaling estadístico

El downscaling estadístico establece relaciones entre variables atmosféricas globales y condiciones locales mediante regresión, corrección de sesgo y aprendizaje automático. Recientemente, modelos de *super-resolution* basados en aprendizaje profundo han permitido transformar pronós-

ticos de $0,25^\circ$ en campos de hasta 1 km de resolución utilizando información satelital de alta resolución (Park et al., 2026).

8.8. Corrección de sesgo y refinamiento temporal

Los modelos globales, tanto físicos como basados en inteligencia artificial, suelen presentar errores sistemáticos asociados a su resolución limitada, a simplificaciones físicas y a la dificultad de representar condiciones locales (Park et al., 2026). En este contexto, las técnicas de corrección de sesgo buscan reducir las diferencias entre pronósticos y observaciones mediante ajustes estadísticos o modelos de aprendizaje automático entrenados con datos históricos.

De manera complementaria, el refinamiento temporal y espacial mejora la representación de las dinámicas locales a partir de pronósticos de baja resolución, mediante interpolación temporal, adaptación regional y modelos capaces de aprender los patrones de una zona geográfica determinada. Estos enfoques han ganado relevancia con el crecimiento de los modelos meteorológicos basados en IA y la disponibilidad de datos satelitales y de estaciones de alta resolución.

8.9. Estaciones meteorológicas y observaciones locales

Las estaciones meteorológicas proporcionan mediciones reales de variables como temperatura, humedad, presión, precipitación y velocidad del viento. Estos datos sirven como referencia (*ground truth*) para validar modelos, evaluar errores de predicción y desarrollar técnicas de corrección de sesgo, y permiten representar fenómenos específicos de una región, especialmente en zonas de geografía compleja o fuerte variabilidad espacial. En los sistemas basados en aprendizaje automático, las observaciones locales se emplean además para entrenar modelos que aprenden

patrones del lugar y adaptan los pronósticos globales a condiciones particulares, constituyendo un componente esencial de la precisión y confiabilidad de la predicción moderna.

9. Metodología

9.1. Enfoque general

Esta investigación propone una metodología basada en aprendizaje automático para adaptar predicciones meteorológicas globales a las condiciones locales de un punto de observación en la ciudad de Valparaíso. A partir de dos fuentes de entrada, los pronósticos globales generados por el modelo fundacional Aurora y las observaciones de una estación meteorológica local, se construye un conjunto de datos con el cual se entrena un modelo de corrección, cuyo producto final es un pronóstico ajustado a la escala local con resolución temporal horaria. El enfoque se inscribe en la tradición de la corrección estadística de pronósticos numéricos (*Model Output Statistics*) (Glahn and Lowry, 1972), actualizada en dos sentidos: el pronóstico base proviene de un modelo fundacional de aprendizaje profundo, y la corrección se aprende mediante técnicas de aprendizaje automático no lineales.

Formalmente, el problema se plantea como el aprendizaje de una función de *corrección residual*. Sea $\tilde{x}_t$ el valor de la variable de interés pronosticado por Aurora e interpolado linealmente al instante horario t . El modelo aprende la función f_θ que estima la discrepancia entre la observación local y dicho pronóstico base:

$$\hat{y}_t = \tilde{x}_t + f_\theta(\mathbf{x}_t^{\text{Aurora}}, \boldsymbol{\tau}_t), \quad (1)$$

donde $\mathbf{x}_t^{\text{Aurora}}$ agrupa las variables pronosticadas por Aurora, interpoladas linealmente al instante t al igual que el pronóstico base, y $\boldsymbol{\tau}_t$ codifica la información temporal (hora del día y posición dentro del intervalo entre pronósticos). Los parámetros θ se ajustan minimizando el error respecto

del residual observado $r_t = y_t - \tilde{x}_t$, donde y_t corresponde a las observaciones de la estación. Esta formulación presenta dos ventajas frente a la predicción directa de y_t : La primera es una división del trabajo ya que Aurora aporta la dinámica atmosférica de gran escala, y el modelo de corrección se especializa únicamente en la discrepancia local, en lugar de aprender desde cero el comportamiento de la variable. La segunda es un piso de desempeño garantizado, en el peor de los casos, una corrección nula devuelve el pronóstico de Aurora interpolado sin alterar, de modo que el sistema nunca puede rendir por debajo de su punto de partida. La inclusión de τ_t es la que permite generar estimaciones horarias a partir de pronósticos disponibles cada 6 horas, abordando simultáneamente la corrección de sesgo y el refinamiento temporal.

La metodología se estructura en cinco etapas: (1) obtención de los datos de reanálisis que alimentan a Aurora, (2) generación de los pronósticos globales, (3) obtención y procesamiento de las observaciones locales, (4) integrar y analizar la construcción del conjunto de datos supervisado, y (5) entrenamiento, evaluación y cuantificación de la mejora del modelo de adaptación local.

9.2. Datos de entrada

9.2.1. Datos de reanálisis ERA5

La ejecución de Aurora requiere como condición inicial el estado de la atmósfera en un instante dado. Para ello se utilizan datos del reanálisis ERA5, obtenidos desde el *Copernicus Climate Data Store* (CDS), plataforma oficial del programa europeo Copernicus.

Se descargan datos entre el 23 de diciembre de 2025 y el 31 de enero de 2026, de modo de disponer de condiciones iniciales para pronosticar los días comprendidos entre el 24 de diciembre de 2025 y el 1 de febrero de 2026 (40 días de pronóstico). Para cada fecha se obtienen un archivo de

variables atmosféricas en 13 niveles de presión (entre 50 y 1000 hPa) y un archivo de variables de superficie; adicionalmente, se descarga un archivo de variables estáticas, que describen características geográficas permanentes utilizadas por el modelo, como la orografía y la máscara tierra-mar (Figura 3).



Figura 3: Tipos de datos de entrada requeridos por el modelo Aurora: variables atmosféricas en niveles de presión, variables de superficie y variables estáticas.

Debido al volumen de información del período, se desarrolla una estructura de almacenamiento y procesamiento por lotes (*batch processing*) que organiza los archivos diarios en una estructura unificada, reduce los tiempos de lectura, facilita la carga secuencial y permite verificar la consistencia entre los pronósticos generados y el reanálisis.

9.2.2. Generación de pronósticos con Aurora

Los pronósticos globales se generan con Aurora en su versión preentrenada a $0,25^\circ$ de resolución espacial (*aurora-0.25-pretrained*), distribuida públicamente por Microsoft a través del repositorio Hugging Face. El modelo se ejecuta en modo de inferencia mediante la librería oficial *aurora* sobre PyTorch, sin adaptadores de ajuste fino (LoRA desactivado, conforme a las recomendaciones de sus autores para el punto de control preentrenado).

El dominio espacial abarca desde los 10°S hasta los 60°S y desde los 100°W hasta los 50°W , región que comprende el territorio chileno y su entorno atmosférico de gran escala en el Pacífico suroriental. A $0,25^{\circ}$ de resolución, el dominio corresponde a una grilla de 201×201 puntos, recortada a 200×200 para satisfacer el requisito de que las dimensiones espaciales sean múltiplos del tamaño de parche del codificador (4 celdas). Las coordenadas se normalizan a la convención de Aurora: latitudes decrecientes y longitudes en $[0^{\circ}, 360^{\circ})$ crecientes.

Cada ejecución se inicializa con dos estados atmosféricos consecutivos, separados por 6 horas, que conforman la historia requerida como entrada. Siguiendo un esquema de tipo operacional, para pronosticar un día D se utilizan como condiciones iniciales los estados de ERA5 de las 12:00 y 18:00 UTC del día $D - 1$, y se ejecuta el modelo de forma autorregresiva durante 4 pasos de 6 horas, obteniendo predicciones para las 00:00, 06:00, 12:00 y 18:00 UTC del día D . Cada valor pronosticado posee así un tiempo de anticipación (*lead time*) de entre 6 y 24 horas, replicando un sistema de pronóstico diario real. El procedimiento se repite para los 40 días del período, registrando para cada predicción el día pronosticado y su tiempo de anticipación.

De cada pronóstico se extraen cinco variables en el punto de grilla más cercano a la estación meteorológica de referencia, situada en la Universidad Técnica Federico Santa María, Valparaíso ($33,03^{\circ}\text{S}$, $71,60^{\circ}\text{W}$): la temperatura a 2 metros (2T), las componentes zonal y meridional del viento a 10 metros (10U, 10V), la presión a nivel del mar (MSL) y la humedad específica en 1000 hPa (Q1000)¹, esta última incorporada por ser el predictor de humedad más cercano a la superficie disponible en las salidas del modelo. Dada la condición costera del punto, se verifica mediante la máscara tierra-mar que el punto de grilla seleccionado corresponda a tierra: la fracción terrestre

¹En el documento se emplea la notación 2T, 10U, 10V, MSL y Q1000, propia de Aurora (Bodnar et al., 2024) y la literatura ECMWF. En la implementación, estos campos corresponden a los identificadores `t2m`, `u10`, `v10`, `msl` y `q` con que ERA5 distribuye las mismas variables, nomenclatura que aparece en el código y sus figuras.

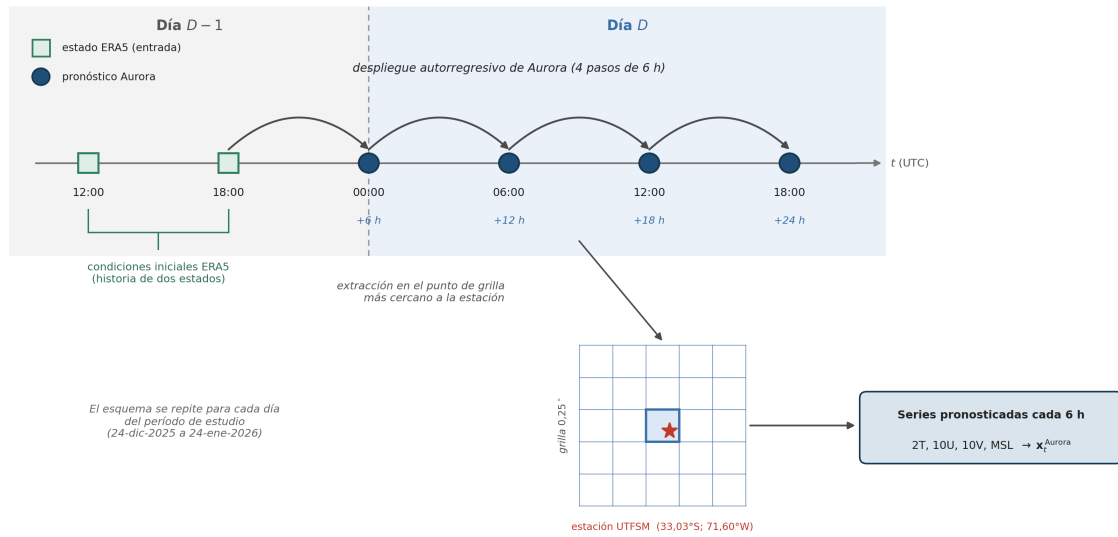


Figura 4: Esquema operacional de pronóstico diario con Aurora. El modelo se inicializa con los estados de ERA5 de las 12:00 y 18:00 UTC del día $D - 1$ y genera, en 4 pasos autorregresivos de 6 horas, las predicciones del día D (*lead times* de 6 a 24 horas), de las cuales se extraen las variables de interés en el punto de grilla más cercano a la estación UTFSM de Valparaíso. Elaboración propia.

resulta de 0,66, o sea, se trata de una celda costera mixta, con cerca de un tercio de superficie marina. Dado que el mar amortigua los cambios de temperatura entre el día y la noche, cabe esperar que la celda oscile menos que el punto de observación, ubicado plenamente en tierra. Las series resultantes constituyen las variables de entrada x_t^{Aurora} de la Ecuación 1.

Verificación del comportamiento autorregresivo. Dado que la aplicación autorregresiva de un modelo de pronóstico acumula errores a medida que se extiende el horizonte (Zhang et al., 2019; Bodnar et al., 2024), previo a la definición del esquema operacional se caracteriza la degradación del pronóstico de Aurora con el tiempo de anticipación: se ejecuta un despliegue autorregresivo extendido de 20 pasos (120 horas) y se compara cada campo pronosticado contra el reanálisis ERA5 del instante correspondiente, sobre la totalidad del dominio. En esta verificación ERA5 actúa como referencia a escala de grilla, rol distinto al de las observaciones de la estación, que

constituyen la referencia local del problema de adaptación.

La degradación se cuantifica mediante el RMSE espacial por paso de pronóstico (Ecuación 5) para las cuatro variables verificadas en grilla (2T, 10U, 10V y MSL), empleando como referencia de destreza la persistencia (el campo del instante inicial mantenido constante). De allí se derivan dos métricas adimensionales: el RMSE normalizado por la desviación estándar espacial del campo de ERA5, y el *skill* respecto de la persistencia, $1 - \text{RMSE}/\text{RMSE}_{\text{pers}}$. Los resultados de esta verificación (Subsección 10.1) respaldan la elección de un horizonte operacional acotado a 24 horas (4 pasos), rango en el cual el error acumulado del despliegue autorregresivo se mantiene en niveles reducidos.

9.2.3. Observaciones meteorológicas locales

Las observaciones locales se obtienen mediante una estación meteorológica Davis Vantage Pro 2², instrumento de uso extendido en monitoreo ambiental. La estación se compone de un conjunto integrado de sensores (ISS, *Integrated Sensor Suite*) que incluye un termohigrómetro alojado en un escudo de radiación pasivo, un anemómetro de copas con veleta, un pluviómetro de balancín y sensores de radiación solar e índice ultravioleta. El ISS transmite las mediciones de forma inalámbrica hacia una consola receptora, que incorpora además un barómetro. La consola se configuró en unidades del Sistema Internacional, velocidad del viento en metros por segundo y presión barométrica en hectopascales.

El registro histórico se realiza mediante un módulo *datalogger* WeatherLink conectado a la consola, configurado con un intervalo de archivo de 5 minutos (con un breve segmento inicial a 1 minuto durante la puesta en marcha, sin efecto sobre los promedios horarios). Los registros

²Davis Instruments Corp., <https://www.davisinstruments.com>

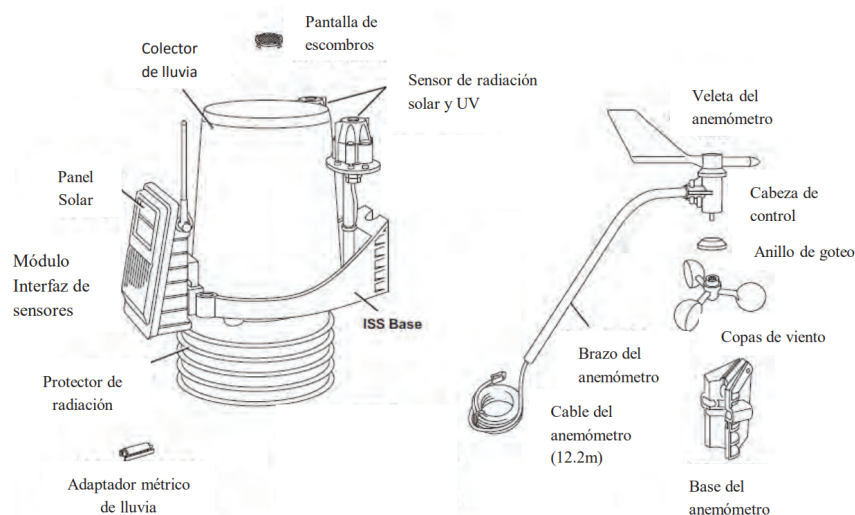


Figura 5: Vista en despiece de los componentes de la estación Davis Vantage Pro 2. A la izquierda, el conjunto integrado de sensores (ISS): colector de lluvia con pantalla de escombros, sensor de radiación solar y UV, panel solar, módulo de interfaz de sensores, protector de radiación y adaptador métrico de lluvia, montados sobre la base del ISS. A la derecha, el anemómetro: veleta, cabeza de control, anillo de goteo, copas de viento, brazo, base de montaje y cable de conexión de 12,2 m. Fuente: Davis Instruments Corp. (2016).

se transfieren periódicamente a un computador mediante el software WeatherLink (Figura 6) y se exportan como archivos de texto delimitados por tabulaciones, con un encabezado de dos filas, cerca de cuarenta variables por registro y la marca --- para mediciones faltantes, convención considerada explícitamente durante la lectura y limpieza.

La estación se instaló en la terraza superior de un edificio de la Universidad Técnica Federico Santa María, en Valparaíso ($33,03^{\circ}\text{S}$, $71,60^{\circ}\text{W}$), un punto elevado y despejado sin muros, estructuras ni vegetación que proyecten sombra sobre los sensores o generen turbulencia y canalizaciones artificiales del viento. El período de medición abarca 41 días, entre el 23 de diciembre de 2025 y el 2 de febrero de 2026, cubriendo íntegramente el período de pronósticos generados con Aurora.

De los campos disponibles se utilizan las cuatro variables coincidentes con las salidas de Aurora: temperatura exterior (Temp_Out), humedad relativa exterior (Out_Hum), velocidad del

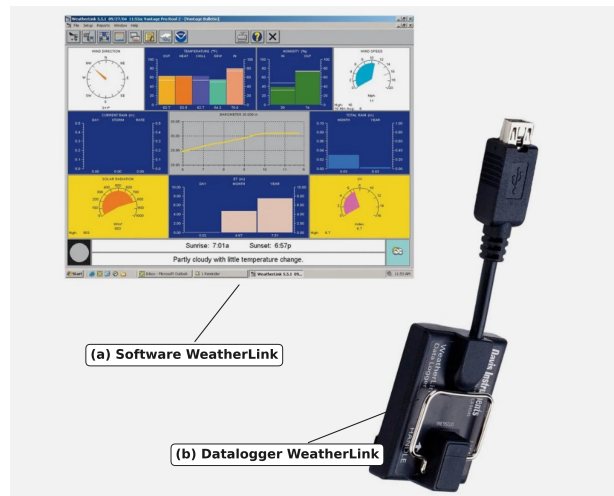


Figura 6: Sistema de registro y descarga de datos de la estación Davis Vantage Pro 2: (a) software WeatherLink, utilizado para la visualización y exportación de los registros históricos; (b) módulo *datalogger* WeatherLink con conexión USB, instalado en la consola receptora. Fuente: Davis Instruments.

viento (*Wind_Speed*) y presión barométrica (*Bar*); la Tabla 1 presenta un extracto de los registros. Estas observaciones constituyen la referencia observacional (*ground truth*) del problema supervisado: los valores objetivo y_t que el modelo aprende a reproducir y contra los cuales se evalúa.

Tabla 1: Extracto de los registros exportados por la estación Davis Vantage Pro 2 (variables seleccionadas).

Fecha	Hora	Temp. ext. (°C)	Hum. (%)	Viento (m/s)	Presión (hPa)
23-12-25	15:51	27,5	40	2,7	1017,7
23-12-25	15:52	27,5	40	4,0	1017,7
23-12-25	15:53	---	40	0,0	1017,7

Sobre los registros crudos se aplica un control de calidad que contempla la identificación de los faltantes, la conversión a formato numérico, la eliminación de registros vacíos y la verificación de consistencia física, incluyendo la detección de valores repetidos consecutivamente (indicio de sensor congelado) y de saltos físicamente improbables. El control arroja un registro de alta calidad: el 99,6 % de las horas cuenta con la totalidad de sus registros de 5 minutos, con solo tres medicio-

nes aisladas de temperatura faltantes y sin marcas temporales duplicadas. Las horas con menos del 50 % de los registros esperados se invalidan por considerarse su promedio poco confiable, criterio que afecta únicamente a la hora final del registro, externa al período de estudio.

9.3. Integración y construcción del conjunto de datos

Los pronósticos globales y las observaciones locales se integran en una única base de datos, diseñada para que el modelo aprenda simultáneamente dos tareas: la *reconstrucción horaria* del comportamiento diario de cada variable a partir de los cuatro estados pronosticados (“rellenar” la evolución entre instantes separados por 6 horas) y la *corrección del sesgo local*, ajustando los pronósticos al comportamiento observado por la estación. Ambas tareas se resuelven de manera conjunta mediante la formulación residual de la Ecuación 1: el residual $r_t = y_t - \tilde{x}_t$ contiene tanto el sesgo de representatividad local como la desviación de la evolución horaria real respecto de la interpolación entre pronósticos.

Homogeneización de unidades y variables derivadas. Las series de Aurora se expresan en las unidades del Sistema Internacional de ERA5, por lo que se aplican conversiones: la temperatura de kelvin a grados Celsius y la presión de pascales a hectopascales. La presión barométrica de la consola Davis ya está reducida al nivel del mar, por lo que es directamente comparable con MSL. En el caso del viento, Aurora entrega las componentes zonal y meridional (10U, 10V) mientras la estación registra la rapidez escalar; se deriva por ello el módulo del viento pronosticado, $|\vec{V}| = \sqrt{u_{10}^2 + v_{10}^2}$, como predictor directo de dicha variable.

La humedad requiere un tratamiento particular, pues las fuentes miden magnitudes distintas: la estación registra humedad relativa (%) y Aurora humedad específica (kg/kg). Para obtener

un predictor comparable, Q1000 se convierte a humedad relativa combinándola con la temperatura y la presión pronosticadas, mediante la presión parcial de vapor $e = q p / (0,622 + 0,378 q)$ y la presión de saturación de la aproximación de Magnus, $e_s = 6,112 \exp[17,67 T / (T + 243,5)]$ con T en grados Celsius, de donde $HR = 100 e / e_s$. La humedad relativa derivada se incorpora como predictor adicional para todas las variables, por tratarse de una combinación no lineal con significado físico (proximidad a la saturación, indicador de nubosidad costera) que los modelos lineales no pueden construir por sí mismos.

Referencia temporal y verificación del alineamiento. Los pronósticos de Aurora están definidos en UTC, mientras que la estación registra en hora local de Chile continental, que durante todo el período (verano austral) corresponde a UTC−3. Antes del emparejamiento, las marcas temporales de los pronósticos se convierten a hora local (el estado de las 18:00 UTC se asocia a las 15:00 locales); omitir esta conversión introduciría un desfase sistemático de 3 horas que distorsionaría el ciclo diurno del error a aprender. El alineamiento se verifica mediante la correlación cruzada entre la temperatura pronosticada y la observada para rezagos de −6 a +6 horas: la correlación máxima se obtiene en rezago cero, confirmando la correcta sincronización de ambas fuentes (Figura 25).

Expansión temporal de los pronósticos. Las series de Aurora poseen cuatro valores por día, mientras que el objetivo es producir estimaciones horarias. El pronóstico base $\tilde{x}_t$ de cada hora se construye mediante interpolación lineal en el tiempo entre los dos estados pronosticados que acotan al instante t , sobre la secuencia continua de pasos de pronósticos consecutivos (Figura 7). La interpolación lineal constituye, además, la primera línea base contra la cual se cuantifica el aporte del aprendizaje automático.

Esta representación se complementa con covariables temporales que indican al modelo la posición de cada hora dentro del día y del intervalo entre pronósticos. La hora local se codifica de forma cíclica mediante el par $(\sin(2\pi h_t/24), \cos(2\pi h_t/24))$, que preserva la vecindad entre el final y el comienzo del día; y el desplazamiento $\delta_t \in \{0, \dots, 5\}$ registra las horas transcurridas desde el último estado pronosticado (las “anclas” de la interpolación). Estas covariables permiten que el modelo aprenda la curvatura sistemática del ciclo diurno que la interpolación lineal no resuelve (por ejemplo, la persistencia de las máximas vespertinas entre los pronósticos de las 15:00 y las 21:00 hora local) en lugar de limitarse a unir los puntos con rectas. El tiempo de anticipación, en cambio, se excluye de los predictores: con una inicialización diaria fija a las 18:00 UTC, cada horizonte cae siempre a la misma hora local, de modo que ambas variables resultan redundantes.

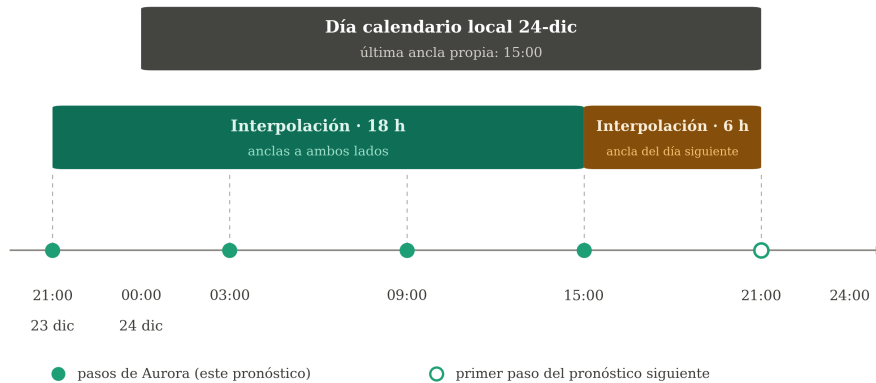


Figura 7: Estructura temporal de la ventana de pronóstico de 24 horas en hora local. Los cuatro pasos generados por Aurora para el día D (21:00 del día $D - 1$ y 03:00, 09:00 y 15:00 del día D , correspondientes a las 00:00, 06:00, 12:00 y 18:00 UTC) actúan como anclas de la expansión horaria. Las primeras 18 horas de la ventana disponen de anclas del propio pronóstico a ambos lados; las 6 horas finales (15:00–21:00) se interpolan utilizando como ancla derecha el primer paso del pronóstico del día siguiente, disponible en un ciclo operacional diario puesto que dicho pronóstico se inicializa con los estados de las 12:00 y 18:00 UTC del día D . Elaboración propia.

Finalmente, se incorpora como predictor el error medio del día de pronóstico anterior, $\bar{e}_{d-1}$, calculado a partir de observaciones ya disponibles al momento de emitir el pronóstico. Su

inclusión responde a la persistencia del error entre días identificada en el análisis exploratorio (autocorrelación diaria de 0,48, Subsubsección 10.2.5): las condiciones atmosféricas que inducen un sesgo determinado tienden a extenderse por varios días.

Agregación de las observaciones. Los registros de la estación, almacenados cada 5 minutos, se agregan a resolución horaria mediante el promedio de cada hora; las discontinuidades breves se completan por interpolación lineal limitada a 3 horas consecutivas, y los huecos mayores se conservan como faltantes, excluyéndose esas horas del conjunto de datos. La variable objetivo y_t representa así la condición media observada durante la hora t , lo que además atenúa el ruido instrumental de alta frecuencia.

Emparejamiento y estructura del conjunto de datos. Ambas fuentes, ya en unidades y referencia temporal comunes, se emparejan mediante una unión interna (*inner join*) sobre el índice horario, conservando las horas en que existen simultáneamente pronóstico y observación. El resultado es un conjunto tabular donde cada fila corresponde a una hora:

$$\mathcal{D} = \{(\mathbf{x}_t, r_t)\}_{t=1}^n, \quad \mathbf{x}_t = \left[\mathbf{x}_t^{\text{Aurora}}, |\vec{V}|_t, \text{HR}_t, \sin \frac{2\pi h_t}{24}, \cos \frac{2\pi h_t}{24}, \delta_t, \bar{e}_{d-1} \right], \quad (2)$$

donde $\mathbf{x}_t^{\text{Aurora}}$ agrupa las variables pronosticadas interpoladas al instante t (2T, 10U, 10V, MSL, Q1000), $|\vec{V}|_t$ el módulo del viento, HR_t la humedad relativa derivada, $(\sin, \cos)$ la codificación cíclica de la hora, δ_t el desplazamiento dentro del intervalo de 6 horas y $\bar{e}_{d-1}$ el error medio del día anterior: 11 predictores por hora. El período comprende 40 días de pronóstico, resultando en 955 ejemplos horarios por variable tras la integración. La Tabla 2 resume la correspondencia entre variables pronosticadas y observadas.

Tabla 2: Correspondencia entre las variables pronosticadas por Aurora y las variables objetivo registradas por la estación.

Variable objetivo (estación)	Unidad	Predictor base Aurora	Transformación aplicada
Temp_Out	°C	2T	$K \rightarrow ^\circ C$
Bar	hPa	MSL	$Pa \rightarrow hPa$
Wind_Speed	m/s	10U, 10V	módulo $ \vec{V} = \sqrt{u_{10}^2 + v_{10}^2}$
Out_Hum	%	Q1000, 2T, MSL	HR derivada (aprox. de Magnus)

Partición del conjunto de datos. El conjunto se divide cronológicamente en un subconjunto de entrenamiento con los primeros 32 días de pronóstico (23 de diciembre de 2025, 21:00 hora local, al 24 de enero de 2026; 768 ejemplos) y un subconjunto de prueba con los 8 días posteriores (25 de enero al 1 de febrero de 2026; 187 ejemplos), proporción aproximada de 80/20. Se evita deliberadamente la partición aleatoria: la autocorrelación de las series meteorológicas implica que una división aleatoria colocaría horas casi idénticas a ambos lados, produciendo fuga de información y una estimación optimista del desempeño. La partición por bloques temporales contiguos obliga, en cambio, a generalizar hacia días no vistos, reproduciendo el uso real del sistema.

Dentro del entrenamiento, toda decisión de modelación (algoritmo, hiperparámetros y predictores) se realiza mediante validación cruzada agrupada por días (*GroupKFold* de 5 pliegues), en la que las 24 horas de cada día de pronóstico se asignan íntegramente al mismo pliegue, extendiendo la lógica anti-fuga al interior del entrenamiento. El subconjunto de prueba se reserva sellado y se utiliza una única vez, al final, para evaluar el sistema ya configurado.

9.4. Modelo de adaptación local

Sobre el conjunto construido se entrena, para cada variable, un modelo de aprendizaje automático supervisado orientado a aprender la corrección residual de la Ecuación 1. Se adopta un modelo independiente por variable, descartándose la formulación multi-salida conjunta por la

heterogeneidad de las variables objetivo, que demandan estructuras de corrección distintas, y por la pérdida de interpretabilidad de compartir una única estructura de decisión entre fenómenos físicos disímiles. Todos los modelos se implementan con la librería `scikit-learn` (Pedregosa et al., 2011), con semilla aleatoria fija para garantizar la reproducibilidad.

Líneas base. Previo al entrenamiento se definen tres líneas base de complejidad creciente, que delimitan el aporte real del aprendizaje automático:

- **B1:** el pronóstico de Aurora interpolado linealmente a resolución horaria, sin corrección alguna (corrección residual nula).
- **B2:** B1 más una corrección constante, igual al sesgo medio del conjunto de entrenamiento.
- **B3:** B1 más una corrección dependiente de la hora del día, calculada como el sesgo medio de cada hora local en el conjunto de entrenamiento.

B3 representa el límite de lo alcanzable sin aprendizaje automático: captura toda la estructura sistemática del sesgo que depende únicamente de la hora, pero no puede adaptar la corrección a las condiciones de cada día. La distancia entre el modelo propuesto y B3 cuantifica, por tanto, el valor que aporta el aprendizaje automático.

Algoritmos candidatos. Se evalúan dos familias situadas en posiciones opuestas del compromiso sesgo–varianza. La primera es la regresión lineal con regularización de cresta (Ridge) (Hoerl and Kennard, 1970), que modela la corrección como combinación lineal de los predictores estandarizados y penaliza la magnitud de los coeficientes mediante el hiperparámetro α , explorado en la grilla $\{0,01; 0,1; 1; 10; 100\}$; su carácter lineal le otorga capacidad de extrapolación e interpretabilidad, a costa de no representar interacciones. La segunda es el *Random Forest* (Breiman, 2001),

ensamble de 300 árboles de regresión entrenados sobre muestras *bootstrap* con selección aleatoria de predictores por división, cuyo promedio reduce la varianza de los árboles individuales; sus hiperparámetros de regularización (profundidad máxima y tamaño mínimo de hoja) se exploran en las grillas $\{4, 6, 8\}$ y $\{5, 10, 20\}$, complementadas con una extensión de control hacia configuraciones de mayor capacidad (profundidades 10, 12 y sin límite; hojas mínimas de 2) para verificar que el óptimo no quede en el borde, seleccionando entre configuraciones estadísticamente equivalentes la más restrictiva (criterio de parsimonia).

Ensamble final. Adicionalmente se evalúa la combinación convexa de las predicciones de ambos algoritmos,

$$\hat{r}_t = w \hat{r}_t^{\text{RF}} + (1 - w) \hat{r}_t^{\text{Ridge}}, \quad (3)$$

fundamentada en la complementariedad estructural de sus errores: al provenir de familias distintas, sus desviaciones están parcialmente descorrelacionadas y tienden a cancelarse en el promedio. Como punto de comparación se evalúa además un *Gradient Boosting*, de la misma familia de árboles que el *Random Forest*, lo que permite contrastar si la ganancia del ensamble proviene de combinar familias de modelos distintas o del simple promedio de estimadores similares. El peso w se trata como un hiperparámetro adicional, seleccionado por validación cruzada sobre los valores $w \in \{0,4; 0,5; 0,6; 0,7; 0,8\}$, complementados con $w = 1,0$ (*Random Forest* puro) como referencia, con desempate por desempeño en el decil de correcciones de mayor magnitud.

Restricciones físicas y tratamiento del sesgo instrumental. Las predicciones reconstruidas se restringen a sus rangos físicamente admisibles (humedad relativa en $[0, 100]$ % y viento no negativo). En la presión barométrica, el análisis exploratorio identifica un desplazamiento aproximada-

mente constante entre el sensor y la referencia de Aurora, atribuible a la calibración del ajuste a nivel del mar de la consola. Dicho desplazamiento no se remueve de los datos: al ser una componente constante del residual, los modelos de corrección lo absorben trivialmente junto con el resto del sesgo. Su origen instrumental se considera en la interpretación de los resultados.

9.5. Evaluación y cuantificación de la mejora

El desempeño se evalúa comparando las predicciones corregidas $\hat{y}_i$ con las observaciones y_i del conjunto de prueba, mediante métricas estándar en verificación de pronósticos meteorológicos. Sea n el número de observaciones y $\bar{y}$ el promedio observado:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$\text{ME} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

El error medio (ME, Ecuación 6) se incluye por ser la métrica que cuantifica el sesgo sistemático, cuya reducción es uno de los objetivos centrales del trabajo. El coeficiente R^2 admite una interpretación de referencia útil: un valor nulo indica desempeño equivalente a predecir el promedio observado, y un valor negativo, desempeño inferior a esa referencia trivial.

La selección entre algoritmos, hiperparámetros y el ensamble se realiza comparando el MAE de validación cruzada agrupada dentro del entrenamiento, siempre en contraste con las líneas base B1–B3 evaluadas bajo el mismo esquema (con sus parámetros calculados exclusivamente en los pliegues de entrenamiento de cada iteración, para asegurar una comparación equitativa). El sistema seleccionado se reentrena con la totalidad del entrenamiento y se evalúa una única vez sobre el conjunto de prueba.

Finalmente, el beneficio de la adaptación local se cuantifica comparando el modelo contra las líneas base. Para cada métrica de error E se calcula la mejora porcentual:

$$\Delta E (\%) = \frac{E_{\text{base}} - E_{\text{modelo}}}{E_{\text{base}}} \times 100, \quad (8)$$

reportándose respecto de B1 (mejora total sobre el pronóstico global sin corregir) y respecto de B3 (aporte específico del aprendizaje automático por sobre la corrección climatológica), de modo que un valor positivo indica que el modelo reduce el error respecto de la referencia correspondiente.

10. Resultados

10.1. Verificación del comportamiento autorregresivo de Aurora

Previo a la adopción del esquema operacional de 4 pasos (24 horas), esta sección cumple con el propósito de verificar de extremo a extremo que la implementación realizada de Aurora (dominio, coordenadas e inicialización) produce pronósticos físicamente coherentes sobre la región de estudio, y respaldar empíricamente la elección del horizonte operacional, ubicándolo dentro de la curva de degradación del despliegue autorregresivo. El experimento consiste en un despliegue extendido (*rollout*) partiendo desde una única inicialización (estados de las 00:00 y 06:00 UTC del primer día) proyectada 20 pasos hacia el futuro (120 horas), sin información nueva de la atmósfera real. Cada paso se verifica contra el ERA5 de su propio instante de validez. A diferencia del esquema operacional, que se reinicializa a diario y no supera el cuarto paso, el despliegue extendido muestra qué ocurriría sin esa reinicialización. Como referencia de destreza se emplea la persistencia — mantener congelado el campo inicial como pronóstico de todos los horizontes — cuyo error mide cuánto cambió la atmósfera en cada horizonte.

La Figura 8 muestra dos comportamientos. Viento y presión tienen un crecimiento sostenido los primeros 2–3 días y estabilización posterior, con Aurora siempre bajo la persistencia. La temperatura presenta curvas en serrucho: la persistencia alcanza máximos en +12, +36 y +60 horas (el campo congelado de madrugada se compara contra una tarde, el momento opuesto del ciclo diurno) y mínimos en +24, +48 y +72 (misma hora de un día posterior: el ciclo diurno se cancela y solo queda el cambio de las condiciones de gran escala entre un día y el siguiente), mientras la curva de Aurora oscila porque su error espacial está modulado por la hora del instante verificado. En ambos casos el ciclo diurno domina sobre la acumulación autorregresiva.

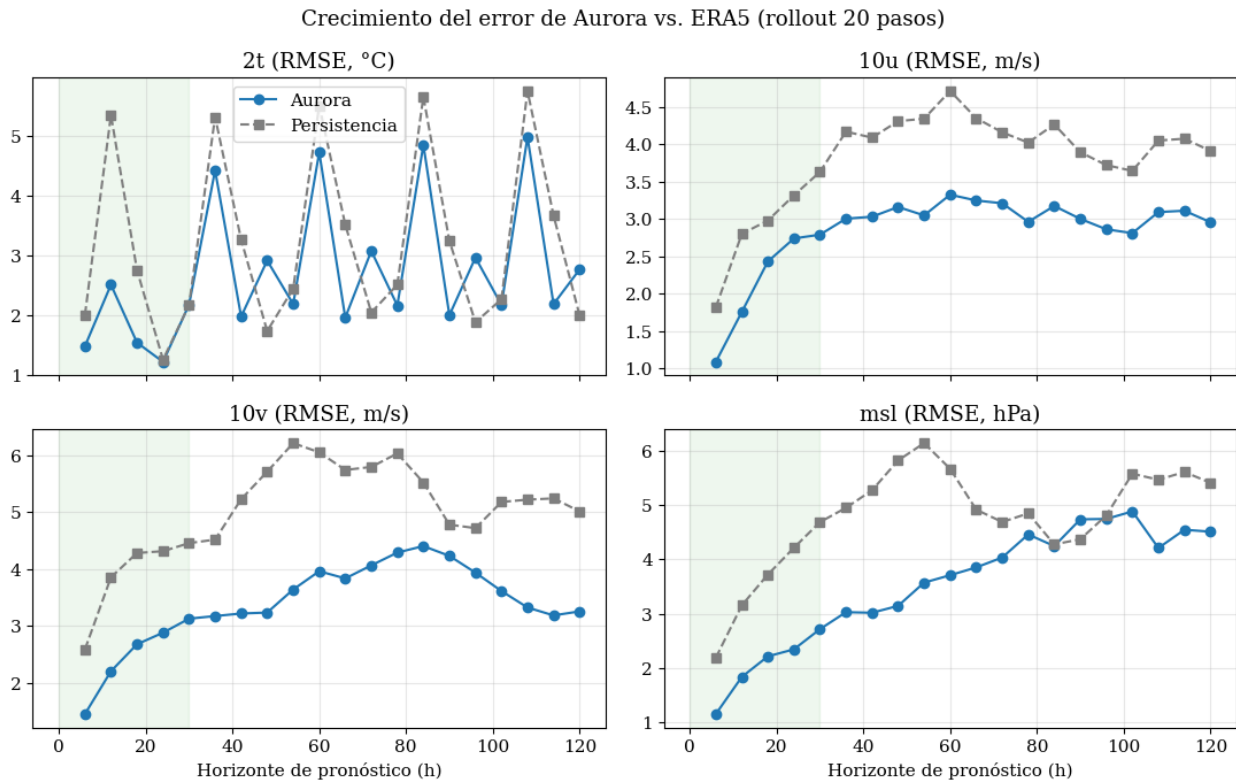


Figura 8: RMSE espacial del pronóstico de Aurora y de la persistencia contra ERA5, en función del horizonte, para las cuatro variables verificadas (2T, 10U, 10V y MSL) (rollout de 20 pasos). La franja sombreada indica la ventana operacional de este trabajo. Elaboración propia.

Los mapas de la Figura 9 agregan la dimensión espacial: diferencias pequeñas y desorganizadas en los primeros horizontes, anomalías regionales hacia +48 h, y en +84 h la pérdida de un episodio cálido de gran escala completo sobre el centro-sur del continente (banda azul del panel de diferencias), desarrollado por la atmósfera real tras cinco días sin información nueva para el pronóstico. A +120 h el patrón general sigue plausible pero magnitudes y localizaciones divergen: el comportamiento típico de los pronósticos generados paso a paso, que pierden primero la amplitud de los eventos y después su estructura.

La Tabla 3 y la Figura 10 cuantifican la ventana operacional con dos métricas adimensionales: el RMSE normalizado por la desviación estándar espacial del campo de ERA5 (1 indicaría un error tan grande como la variabilidad del propio campo) y el *skill* respecto de la persistencia,

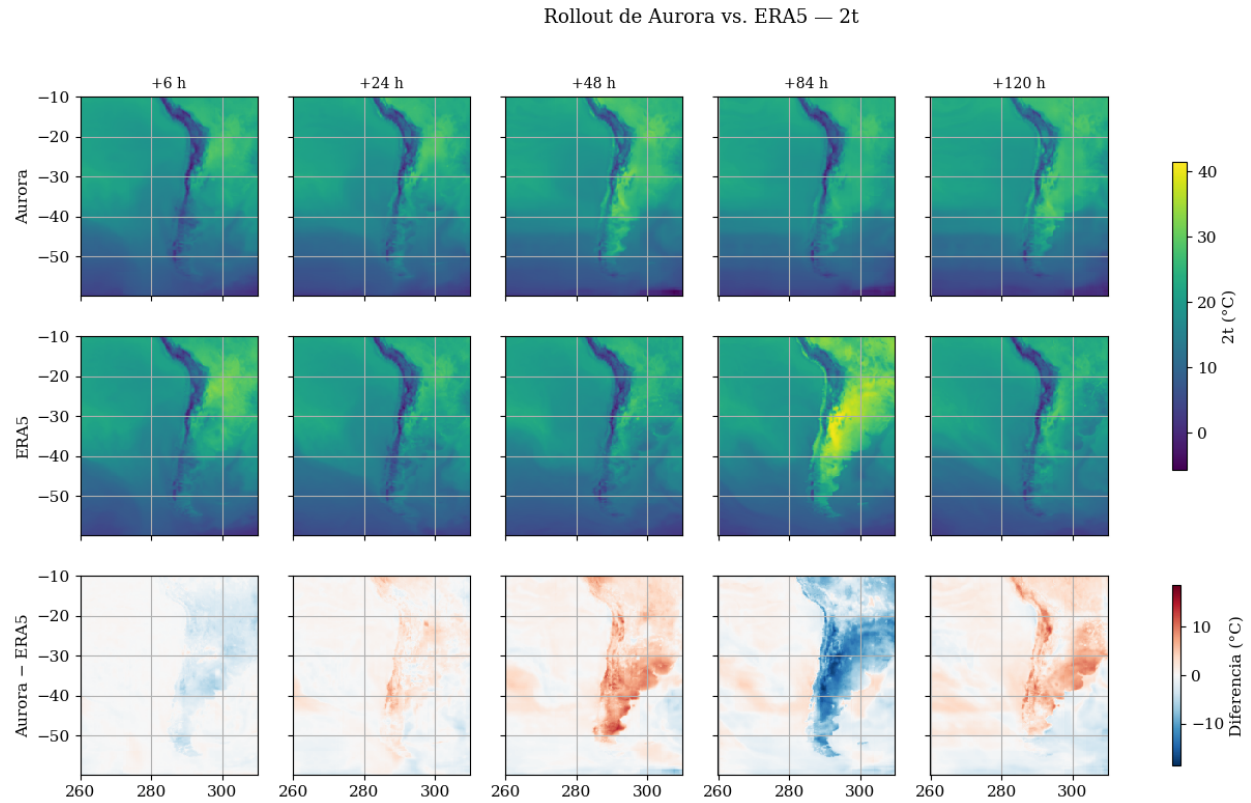


Figura 9: Campos de temperatura a 2 metros pronosticados por Aurora (arriba), reanálisis ERA5 (centro) y su diferencia (abajo) para horizontes seleccionados del rollout de 20 pasos. Elaboración propia.

1 — $RMSE/RMSE_{pers}$, positivo cuando Aurora supera a la línea base.

Tabla 3: Error normalizado y *skill* respecto de la persistencia del pronóstico de Aurora contra ERA5, para los cuatro pasos de la ventana operacional.

Horizonte	RMSE normalizado				Skill vs. persistencia			
	2T	10U	10V	MSL	2T	10U	10V	MSL
6 h	0,225	0,255	0,313	0,165	0,261	0,402	0,438	0,472
12 h	0,315	0,391	0,483	0,253	0,530	0,373	0,428	0,417
18 h	0,239	0,551	0,634	0,327	0,440	0,184	0,374	0,404
24 h	0,209	0,602	0,691	0,342	0,029	0,174	0,332	0,443

Dos hechos sustentan el horizonte elegido. Primero, el error normalizado queda muy por debajo de la unidad en los 16 casos (máximo de 0,69 en 10V a las 24 horas; 0,21 y 0,34 para 2T y MSL): el error de Aurora es una fracción menor de la variabilidad espacial de los campos.

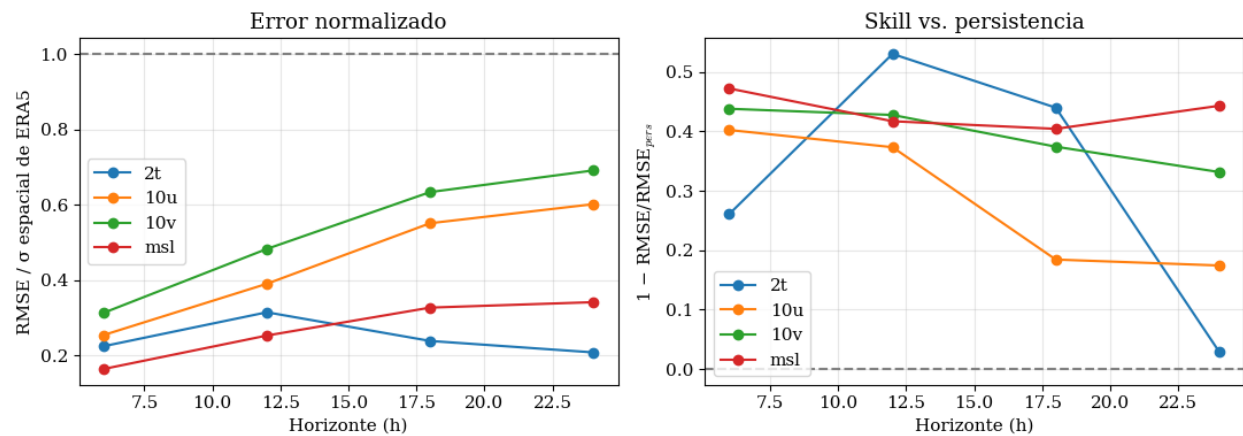


Figura 10: Error normalizado (izquierda) y *skill* respecto de la persistencia (derecha) para los cuatro pasos de la ventana operacional. La línea segmentada marca, respectivamente, el error igual a la variabilidad del campo y el empate con la persistencia. Elaboración propia.

Segundo, el *skill* es positivo en los 16 casos: Aurora supera a la persistencia en todas las variables y pasos de la ventana.

El *skill* de 2T a las 24 horas (0,029, prácticamente neutro) se explica por el serrucho ya descrito: en ese horizonte la persistencia compara la misma hora del día y hereda gratis el ciclo diurno, quedando como una línea base excepcionalmente exigente. El pronóstico no se degradó (el error normalizado de 2T es ahí el más bajo de la ventana, 0,209); que Aurora iguale el truco de copiar el día anterior por la vía de simular la atmósfera confirma la consistencia del despliegue.

En conjunto, a horizontes de hasta 24 horas el despliegue mantiene errores reducidos, campos coherentes y ventaja sistemática sobre la persistencia; la degradación sustancial (crecimiento del error de viento, pérdida de episodios de gran escala) aparece recién desde el segundo y tercer día, fuera de la ventana utilizada.

10.2. Análisis exploratorio

Previo al entrenamiento de los modelos de corrección, se caracteriza el error de Aurora respecto de las observaciones locales, con doble propósito: cuantificar la magnitud y estructura de la discrepancia a corregir, y fundamentar empíricamente las decisiones de diseño declaradas en la metodología. La Figura 11 presenta las series horarias completas de temperatura y anticipa los dos rasgos centrales del análisis: Aurora reproduce adecuadamente la variabilidad de escala de días (los períodos cálidos y fríos evolucionan sincronizados en ambas series), pero atenúa sistemáticamente el ciclo diurno, sin alcanzar los máximos vespertinos observados.

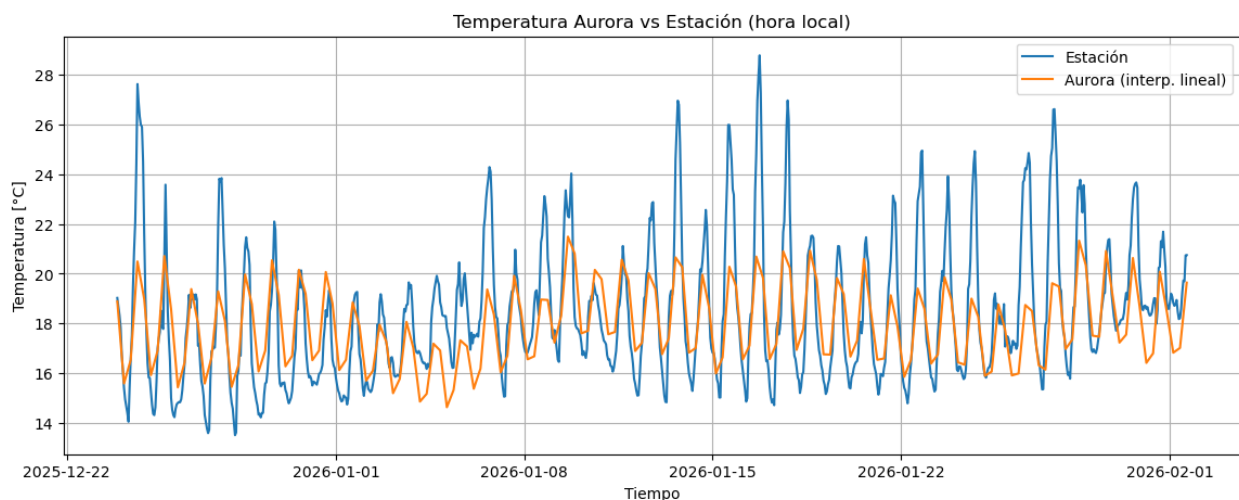


Figura 11: Series horarias de temperatura de la estación y de Aurora interpolada linealmente (hora local), para el período completo de estudio. Elaboración propia.

10.2.1. Ciclo diurno y estructura horaria del sesgo

La Figura 12 descompone la discrepancia en su estructura horaria. El panel (a) compara los ciclos diurnos promedio: la estación registra una amplitud de 6,36 °C (máximo de 22,1 °C a las 18:00 y mínimo de 15,8 °C a las 06:00), mientras el ciclo de Aurora alcanza solo 3,49 °C, el modelo global captura el 55 % de la amplitud diurna local. La atenuación es consistente con el

carácter mixto de la celda (fracción terrestre de 0,66), se espera que el mar tarda mucho más que la tierra en calentarse y enfriarse, y el tercio marino de la celda modera el ciclo respecto del punto de observación. El desfase aparente entre máximos (15:00 en Aurora, 18:00 en la estación) combina el suavizamiento real con un efecto de muestreo, pues el máximo de la serie interpolada solo puede ocurrir sobre una hora-ancla.

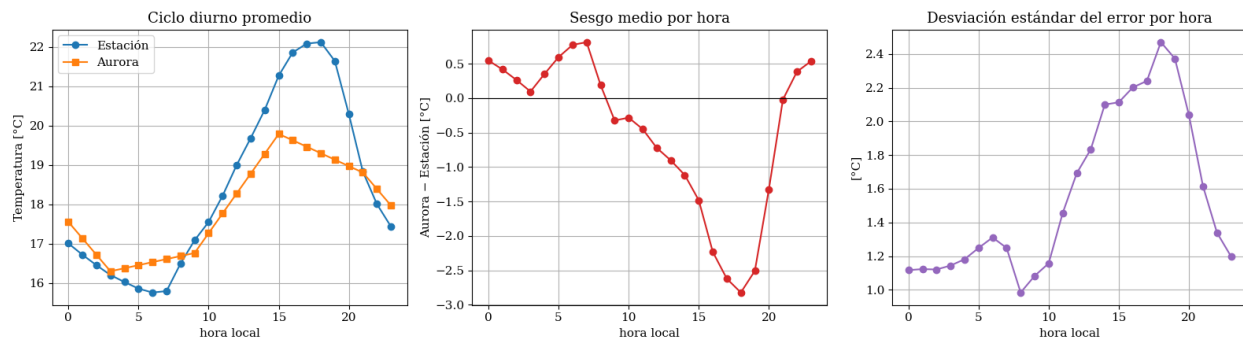


Figura 12: Estructura horaria del error de temperatura de Aurora respecto de la estación. (a) Ciclo diurno promedio de ambas fuentes. (b) Sesgo medio por hora del día (Aurora – estación). (c) Desviación estándar del error por hora. Elaboración propia.

El panel (b) revela la propiedad más relevante del sesgo, este cambia de signo con la hora. Aurora resulta cálida durante la noche y madrugada (hasta $+0,8\text{ }^{\circ}\text{C}$) y marcadamente fría durante la tarde (mínimo de $-2,82\text{ }^{\circ}\text{C}$ a las 18:00), por la posición de la celda, se espera que el océano impida que la celda se enfríe tanto como el punto continental de noche, y que se caliente tanto de tarde. El sesgo global, de apenas $-0,48\text{ }^{\circ}\text{C}$, enmascara por compensación esta estructura, lo que anticipa el fracaso de una corrección constante ya que puede mejorar las tardes a costa de empeorar las noches, que son más horas.

El panel (c) muestra que la dispersión del error también depende de la hora, duplicándose entre la madrugada ($\sim 1,1\text{ }^{\circ}\text{C}$) y la tarde ($\sim 2,5\text{ }^{\circ}\text{C}$). Esta dependencia delimita el aporte de cada método: el sesgo medio por hora es eliminable con una corrección climatológica (B3), pero la dispersión alrededor de él, que varía día a día según las condiciones meteorológicas, solo es accesible

para un modelo capaz de leerlas. El MAE del pronóstico interpolado sin corrección (B1) asciende a 1,47 °C, referencia inicial de todo el desarrollo posterior.

10.2.2. Error por horizonte de pronóstico

El esquema operacional asigna a cada hora un tiempo de anticipación de entre 6 y 24 horas, lo que plantea si el error crece con el horizonte. La Tabla 4 y la Figura 13 presentan el error evaluado en los instantes exactos de predicción, agrupado por *lead time*.

Tabla 4: Error de la temperatura de Aurora por horizonte de pronóstico, evaluado en los instantes exactos de predicción ($n = 40$ por horizonte). Cada horizonte corresponde siempre a la misma hora local debido a la inicialización diaria fija.

Lead [h]	Hora local	Sesgo [°C]	MAE [°C]	Desv. est. [°C]
6	21:00	-0,02	1,33	1,62
12	03:00	+0,10	0,92	1,14
18	09:00	-0,32	0,89	1,08
24	15:00	-1,48	1,84	2,11

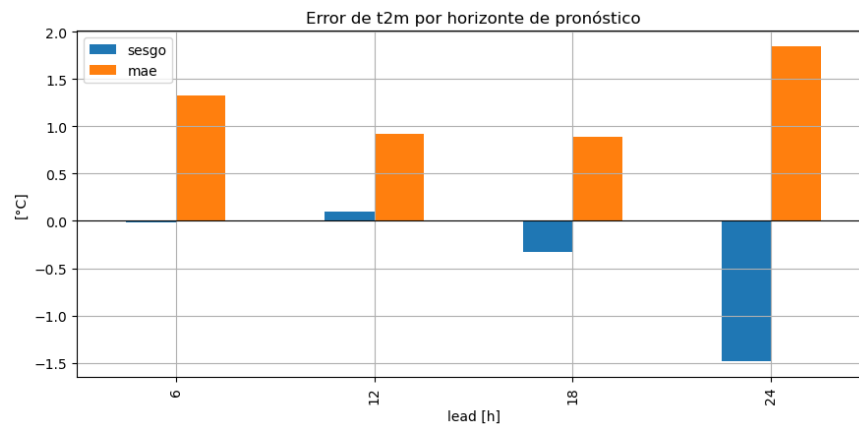


Figura 13: Sesgo y MAE de la temperatura de Aurora según horizonte de pronóstico. El patrón no monótono indica que el error está dominado por la hora local asociada a cada horizonte y no por la degradación autorregresiva. Elaboración propia.

Si la degradación autorregresiva dominara, el error crecería de forma monótona con el horizonte. El patrón observado no lo es (MAE de 1,33, 0,92, 0,89 y 1,84 °C para 6, 12, 18 y 24

horas), y la explicación reside en la tercera columna de la tabla: con inicialización diaria fija, cada horizonte cae siempre a la misma hora local, y el error reproduce el mapa de dificultad horaria de la Figura 12: los horizontes de 12 y 18 horas caen en madrugada y mañana, horas de sesgo reducido, mientras el de 24 cae a las 15:00, en la antesala del máximo sesgo vespertino. A horizontes de hasta 24 horas, el error está dominado por la representatividad local y no por el tiempo de anticipación, en concordancia con la baja pérdida de precisión de los modelos globales modernos en horizontes cortos. De aquí se sigue la decisión de diseño ya incorporada en la metodología: horizonte y hora local son información redundante, y el primero se excluye de los predictores. El horizonte de 6 horas ilustra además la matiz de sesgo casi nulo ($-0,02\text{ }^{\circ}\text{C}$) con desviación elevada ($1,62\text{ }^{\circ}\text{C}$), es decir, al atardecer Aurora no se equivoca de manera sistemática sino distinta cada día.

10.2.3. Distribuciones y eventos extremos

Se examina cómo se distribuyen los valores de ambas fuentes y qué ocurre en los extremos. La Figura 14 presenta la comparación de cuantiles (Q-Q), la dispersión hora a hora y los histogramas; la Tabla 5 cuantifica la comparación por percentiles.

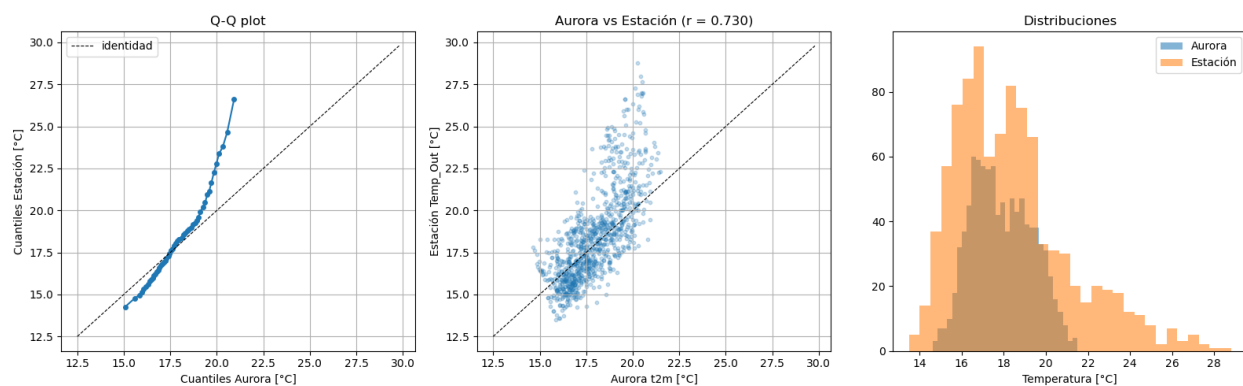


Figura 14: Comparación de la distribución entre la temperatura de Aurora y la de la estación. (a) Diagrama cuantil-cuantil; la línea segmentada indica la identidad. (b) Dispersión hora a hora ($r = 0,730$). (c) Histogramas de ambas series. Elaboración propia.

Tabla 5: Percentiles de temperatura de ambas fuentes durante el período de estudio.

Percentil	Aurora [°C]	Estación [°C]	Diferencia [°C]
p1	15,1	14,3	+0,8
p5	15,8	15,0	+0,9
p25	16,8	16,3	+0,4
p50	17,8	18,0	-0,2
p75	19,0	19,6	-0,6
p95	20,3	23,8	-3,5
p99	20,9	26,6	-5,7

La estructura del cuadro es elocuente: en el centro de la distribución (p25–p75) las diferencias no superan $\pm 0,6$ °C (Aurora está bien calibrada para condiciones típicas), pero ambas colas aparecen comprimidas. En la fría, Aurora queda +0,8 a +0,9 °C más cálida; en la caliente, hasta -5,7 °C demasiado fría. El despegue de la curva Q-Q en el extremo derecho del panel (a) indica que la corrección requerida no es un desplazamiento constante, sino que crece con la temperatura y alcanza su máximo en los valores más altos.

El resultado más relevante es el conteo de eventos cálidos: la estación registró **43 horas con temperaturas superiores a 24 °C; Aurora no predijo ninguna**. Los eventos de mayor relevancia práctica resultan invisibles a la escala de la celda. El panel (b) revela la causa: la temperatura de Aurora se satura en torno a los 21 °C, emitiendo valores casi idénticos en un día caluroso y en uno extremo. La temperatura pronosticada, por sí sola, no contiene información para reconstruir los extremos locales; la viabilidad de rescatarlos depende de que otras variables de Aurora los discriminen, hipótesis examinada en la Subsubsección 10.2.5.

10.2.4. Pérdida por resolución temporal

Este análisis aísla la segunda fuente de error del sistema base, la resolución temporal de 6 horas. El experimento elimina a Aurora de la ecuación considerando solo las observaciones de la

estación en las horas-ancla (21:00, 03:00, 09:00 y 15:00) y se interpolan linealmente a resolución horaria, se mide la desviación de la observación real respecto de esa reconstrucción. El residual cuantifica lo que se pierde *exclusivamente* por unir con rectas estados separados por 6 horas.

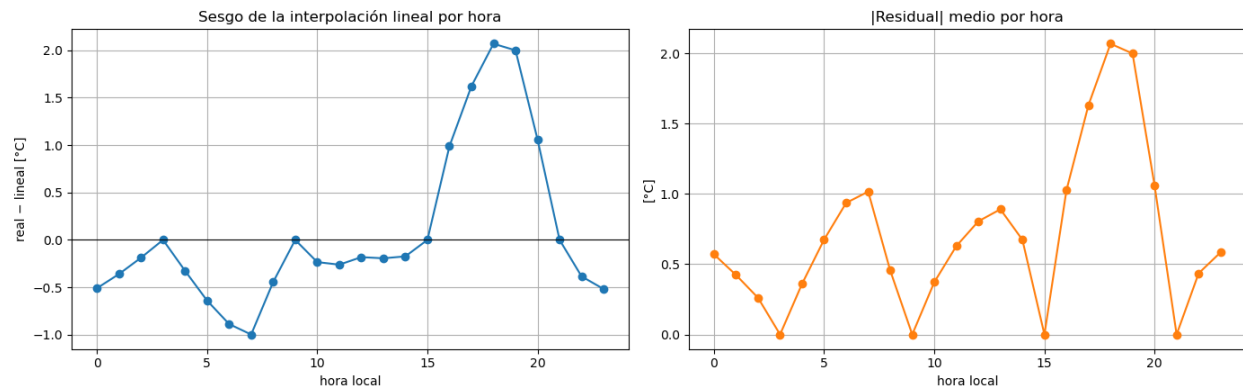


Figura 15: Residual de la interpolación lineal entre observaciones reales muestreadas en las horas-ancla. (a) Sesgo del residual por hora del día (observación – interpolación). (b) Magnitud media del residual por hora. Los valores nulos en las horas-ancla verifican la construcción del experimento. Elaboración propia.

El análisis muestra que la interpolación lineal “perfecta” incurre en un MAE de **0,70 °C**. Aun si Aurora predijera sin error en sus cuatro pasos, el sistema se equivocaría en promedio 0,70 °C por hora solo por la resolución temporal. Entonces, el refinamiento temporal no es un complemento accesorio de la corrección de sesgo ya que por sí solo explica una porción importante del error total.

La Figura 15(a) localiza dónde falla la recta, con dos esquinas sistemáticas. La mayor es entre las anclas de las 15:00 y 21:00, la temperatura real se mantiene elevada hasta las 18:00–19:00 mientras la recta descende uniforme, con residual máximo de +2,07 °C a las 18:00. La segunda es el mínimo del amanecer: entre las anclas de las 03:00 y 09:00, la recta asciende directamente mientras la temperatura real sigue bajando hasta las 06:00–07:00 (residual de –1,0 °C). Ambas son funciones sistemáticas de la hora y, por tanto, aprendibles mediante las covariables temporales. El mediodía muestra el caso contrario: residual medio casi nulo (–0,2 °C) con desviación de 1,0–

1,2 °C, es decir, la recta falla allí de forma distinta cada día, componente accesible solo para un modelo que considere las condiciones meteorológicas.

Las dos fuentes de error identificadas se superponen, además, en la misma franja: a las 18:00 confluyen el máximo sesgo frío de Aurora ($-2,82$ °C) y el máximo residual de interpolación ($+2,07$ °C). La tarde concentra los dos déficits que el modelo debe resolver simultáneamente, lo que explica que sea la franja de mayor error del sistema base y, como se verá, la de mayor mejora del sistema propuesto.

10.2.5. Persistencia del error y dependencia de las condiciones

La corrección climatológica por hora puede eliminar el sesgo sistemático del ciclo diurno, pero no la variabilidad condicional alrededor de él. Esta subsección examina esa estructura condicional: si el error persiste entre días, qué variables de Aurora se relacionan con él, y si los extremos de la Tabla 5 poseen una firma reconocible en el espacio de predictores. La Figura 16 reúne los seis paneles.

Persistencia. La autocorrelación horaria (panel a) exhibe máximos periódicos en los rezagos de 24, 48 y 72 horas, el error “respira” con el ciclo diurno y persiste entre días. El panel (b) la cuantifica a escala diaria mostrando que la autocorrelación de primer orden del error medio diario alcanza $r = 0,48$, reflejo de que las condiciones atmosféricas que inducen un sesgo determinado suelen extenderse por varios días. Este hallazgo fundamenta la inclusión del error medio del día anterior ($\bar{e}_{d-1}$) como predictor, decisión declarada en la metodología y operacionalmente legítima: al emitir el pronóstico, las observaciones del día previo ya están disponibles.

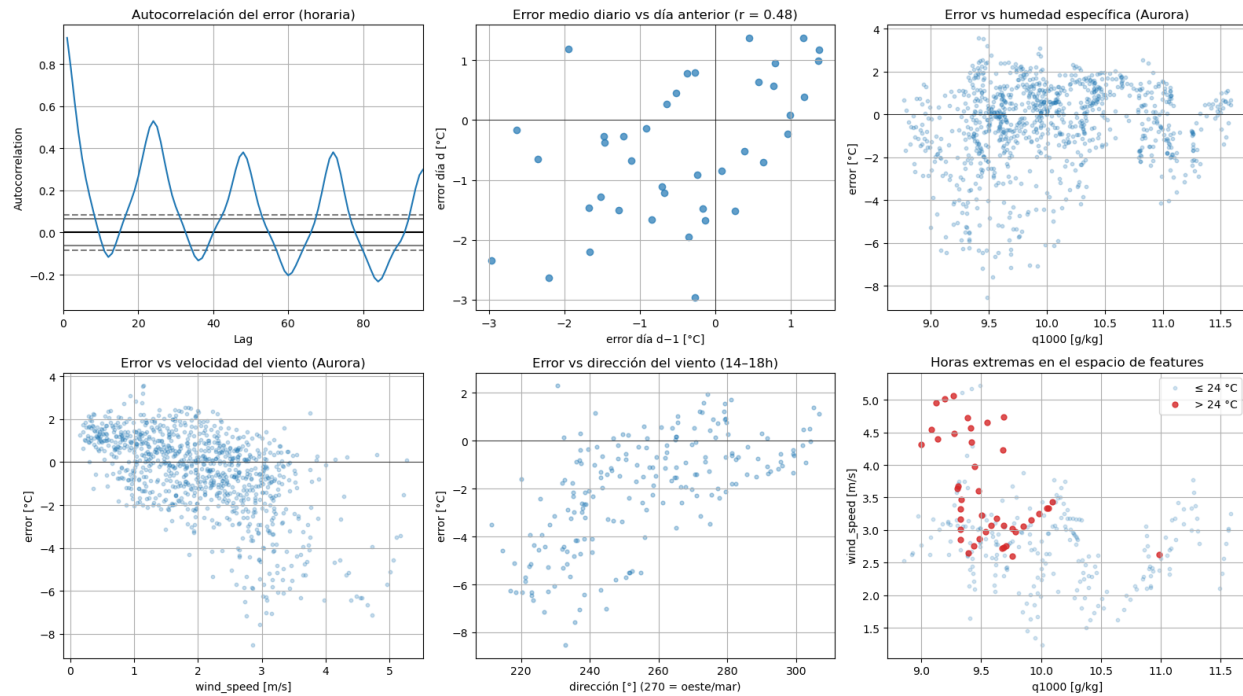


Figura 16: Estructura condicional del error de temperatura. (a) Autocorrelación horaria del error. (b) Error medio diario contra el del día anterior ($r = 0,48$). (c) Error horario contra la humedad específica pronosticada. (d) Error horario contra la velocidad del viento pronosticada ($r = -0,57$). (e) Error vespertino (14:00–18:00) contra la dirección del viento pronosticada. (f) Horas extremas ($> 24^\circ\text{C}$, en rojo) y horas normales de las mismas franjas horarias en el plano humedad–viento de Aurora. Elaboración propia.

Dependencia de las condiciones pronosticadas. La velocidad del viento de Aurora resulta el predictor individual más fuerte del error (panel d, $r = -0,57$), cuando Aurora pronostica viento intenso, su temperatura tiende a quedar muy por debajo de la observada. La humedad específica muestra una relación lineal débil (panel c, $r = 0,17$), con indicios de estructura no lineal en la cola de errores grandes. El panel (e) agrega la dimensión direccional para las horas vespertinas: los errores fríos de mayor magnitud se concentran en direcciones del suroeste ($220\text{--}250^\circ$), de modo que las componentes del viento (10U, 10V), y no solo su magnitud, portan información sobre el régimen del error, lo que respalda su inclusión como predictores separados.

El perfil de los eventos extremos. El panel (f) responde la pregunta abierta en la subsección de distribuciones ¿contienen las variables de Aurora información para reconocer los eventos cálidos que su temperatura saturada no distingue? La respuesta es afirmativa, las 43 horas sobre 24 °C se separan con claridad en el plano humedad–viento de Aurora, caracterizándose por aire más seco (9,6 frente a 10,1 g/kg) y viento más intenso (3,6 frente a 2,8 m/s), producto del aire continental seco y cálido que llega desde el interior durante los episodios de calor de la costa central. Aunque la temperatura pronosticada no discrimine los extremos, el conjunto multivariable de predictores sí contiene la información para anticiparlos. Esto justifica la implementación del modelo de corrección multivariable.

En síntesis, el análisis identifica tres estructuras condicionales explotables por el aprendizaje automático (persistencia diaria, dependencia del régimen de viento y la firma seco-ventosa de los extremos), ninguna accesible para las líneas base B2 y B3.

10.2.6. Velocidad del viento

Las dos subsecciones siguientes extienden la caracterización a las restantes variables objetivo, comenzando por el viento, a priori la de mayor dificultad por su dependencia de las condiciones del punto de medición. La Figura 17 compara el módulo pronosticado por Aurora con la rapidez registrada por la estación.

El resultado más favorable es la concordancia de fase del panel (a): ambos ciclos comparten la forma, aunque en escalas distintas. Aurora representa correctamente el fenómeno de la brisa; lo que no representa es la atenuación local del punto. La correlación horaria de $r = 0,611$ indica una relación aprendible pese a la discrepancia de magnitudes.

Dicha discrepancia es severa y estructurada. El sesgo global asciende a +1,11 m/s sobre

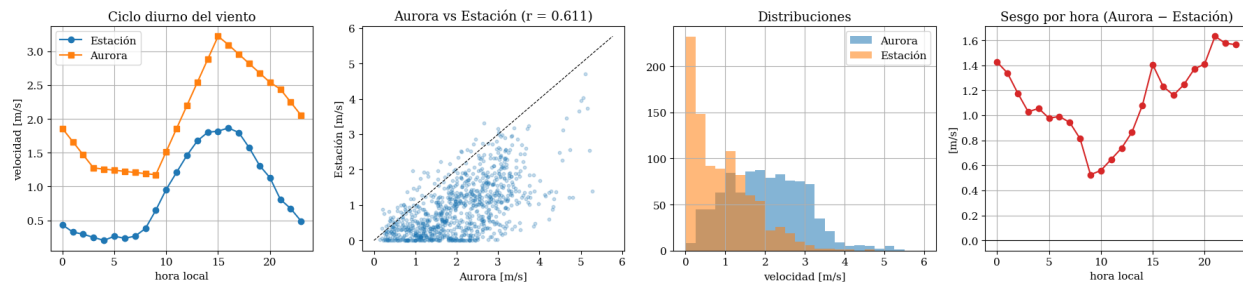


Figura 17: Comparación de la velocidad del viento entre Aurora y la estación. (a) Ciclo diurno promedio de ambas fuentes. (b) Dispersión hora a hora ($r = 0,611$). (c) Histogramas de ambas series. (d) Sesgo medio por hora del día (Aurora – estación). Elaboración propia.

una media observada de apenas 0,91 m/s: el viento de la celda, representativo de un área de 25 km con un tercio de mar y sin obstáculos, duplica con creces al del anemómetro, frenado por los edificios y la rugosidad del entorno urbano. El panel (d) muestra que el sesgo depende de la hora (mínimo de $\sim 0,55$ m/s a media mañana, máximo de $\sim 1,6$ m/s en la noche), y la lectura relativa es aún más reveladora: de noche la razón Aurora/estación se acerca a 6 (1,5 contra 0,25 m/s) y en la tarde baja a 1,7. La corrección requerida es más multiplicativa que aditiva y dependiente de la hora, lo que motiva evaluar también la predicción directa de la magnitud, comparación presentada en el sistema multivariable.

El panel (c) expone el límite estructural, el 40 % de las horas observadas corresponde a calmas ($< 0,5$ m/s), frente al 6 % en Aurora. Esta concentración en el cero, atribuible al abrigo del punto y a la velocidad mínima que el anemómetro alcanza a registrar, define un piso de predictibilidad: ningún modelo de regresión continua reproduce exactamente el valor nulo, y la restricción de no negatividad declarada en la metodología opera sobre este régimen.

Que Aurora sobreestime el viento local no contradice que su viento sea el mejor predictor del error de temperatura ($r = -0,57$, Subsubsección 10.2.5). Como predictor, una variable no requiere ser exacta sino informativa del régimen atmosférico: un mismo campo puede ser, a la vez,

un pronóstico deficiente de su propia variable y un excelente predictor del error de otra.

10.2.7. Humedad relativa

La caracterización emplea como referencia la humedad relativa derivada de las propias variables de Aurora (Q1000, 2T y MSL) mediante la aproximación de Magnus, según la metodología.

La Figura 18 la compara con los registros del sensor de humedad de la estación.

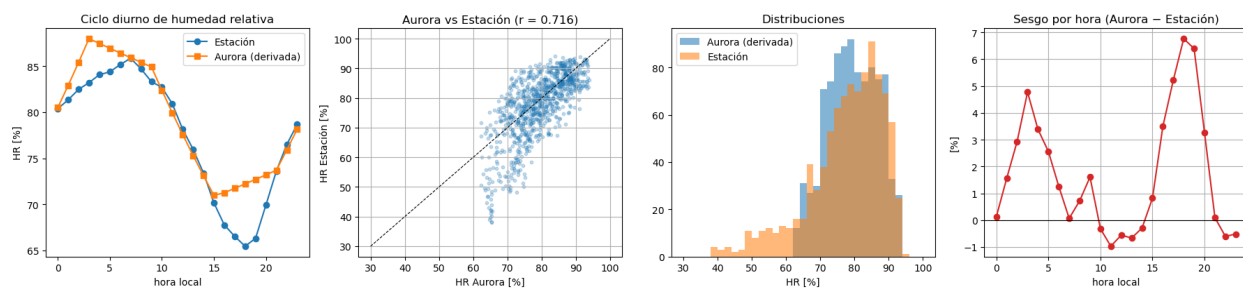


Figura 18: Comparación de la humedad relativa entre la estación y la derivada de las variables de Aurora. (a) Ciclo diario promedio. (b) Dispersión hora a hora ($r = 0,716$). (c) Histogramas de ambas series. (d) Sesgo medio por hora del día (Aurora – estación). Elaboración propia.

Pese a la doble aproximación de la derivación (niveles cercanos pero no idénticos y fórmula empírica de saturación), la humedad derivada alcanza $r = 0,716$ con la observada, prácticamente equivalente a la de la temperatura ($r = 0,730$), y un sesgo global de apenas $+1,7\%$. La variable resulta tan tratable como la temperatura, validando su uso como predictor derivado.

La estructura horaria del sesgo (panel d) confirma que el máximo sesgo húmedo ($+6,8\%$) ocurre a las 18:00, exactamente la hora del máximo sesgo frío de la temperatura. La coincidencia no es causal ya que la humedad relativa mide cuán cerca está el aire de saturarse, y esa capacidad depende de la temperatura. Al subestimar la temperatura vespertina, Aurora describe un aire que admite menos vapor del que el aire real admite, y sobreestima en consecuencia la humedad relativa. Una parte sustancial del sesgo húmedo es, así, la sombra del sesgo térmico, por lo que su corrección

se beneficia de los mismos predictores que la de temperatura.

El panel (c) revela el fenómeno espejo en el extremo caliente: la *cola seca* (humedades bajas) está comprimida. La estación registra hasta 38 % de humedad, mientras la serie derivada nunca baja de 62 %. Son físicamente los mismos eventos: las tardes de aire continental seco que producen las 43 horas sobre 24 °C llevan la humedad bajo el rango alcanzable por la celda mixta. Un único fenómeno, la incapacidad de la celda costera de reproducir condiciones continentales extremas, explica las colas comprimidas de dos variables. En el extremo opuesto, Aurora detecta 59 de las 78 horas con humedad sobre el 90 % (indicador de niebla y nubosidad costera baja), una tasa del 76 % que confirma que ese régimen está razonablemente representado.

Síntesis del análisis exploratorio. Aurora reproduce la variabilidad de escala de días y la fase de los ciclos diurnos de todas las variables, pero atenúa sus amplitudes y comprime sus extremos, con un sesgo que cambia de signo según la hora (lo que invalida las correcciones constantes) y una dispersión condicional inaccesible para la corrección meteorológica (B3). La resolución temporal de 6 horas aporta por sí sola una fracción comparable del error total, con esquinas sistemáticas aprendibles. El error exhibe, además, tres estructuras condicionales explotables (persistencia diaria, régimen de viento y condiciones secas y ventosas de los extremos) que justifican el conjunto de predictores adoptado. Las secciones siguientes cuantifican ese espacio.

10.3. Selección del modelo en validación cruzada

Esta sección presenta la selección del modelo de corrección para la temperatura, la variable de desarrollo del sistema. Todas las comparaciones emplean el MAE de validación cruzada agrupada por días (5 pliegues) dentro del entrenamiento; el conjunto de prueba permanece sellado

durante todo el proceso.

10.3.1. Desempeño de las líneas base

La validación cruzada arroja MAE de 1,465 °C para B1, 1,505 °C para B2 y 1,218 °C para B3. Dos lecturas se desprenden. Primero, **B2 no solo no mejora a B1: lo empeora**, confirmando la predicción del análisis exploratorio, con un sesgo que cambia de signo según la hora, una corrección constante de $-0,48$ °C mejora las tardes al costo de empeorar las noches, que son más horas. Segundo, B3 establece la vara real del problema: el sesgo climatológico horario ya reduce el error un 17 %, y todo aporte del aprendizaje automático debe medirse desde 1,218 °C.

10.3.2. Búsqueda de hiperparámetros

La Tabla 6 resume la búsqueda en grilla. Ridge es insensible a la regularización en el rango bajo ($\alpha \leq 1$, MAE entre 0,994 y 1,001 °C) y se deteriora para valores altos; se selecciona $\alpha = 0,1$ (0,994 °C). En el bosque, el MAE mejora con la profundidad hasta estabilizarse en la configuración de profundidad 8 y hoja mínima 5 (0,957 °C). Como control de borde se evalúa una extensión hacia mayor capacidad (profundidades 10, 12 y sin límite, hojas de 2) donde el mejor valor alcanzable es 0,950 °C, una ganancia de solo 0,007 °C ($< 1\%$). Conforme al criterio de parsimonia declarado, se conserva la configuración más restrictiva, estadísticamente equivalente y de menor riesgo de sobreajuste.

El ordenamiento global contiene un resultado estructural: **todas** las configuraciones de ambos algoritmos, incluida la peor ($\alpha = 100$, 1,081 °C), superan a la mejor línea base sin aprendizaje (B3, 1,218 °C). La distancia entre el bosque seleccionado y B3 (0,957 contra 1,218 °C, una reducción adicional del 21 %) cuantifica el valor de las relaciones condicionales identificadas en el

Tabla 6: MAE de validación cruzada agrupada [°C] para las grillas de hiperparámetros de Ridge y Random Forest (variable: temperatura).

Ridge (regularización α)	
$\alpha = 0,01$	0,999
$\alpha = 0,1$	0,994
$\alpha = 1$	1,001
$\alpha = 10$	1,032
$\alpha = 100$	1,081

Random Forest (MAE según profundidad y hoja mínima)			
	hoja ≥ 5	hoja ≥ 10	hoja ≥ 20
prof. 4	1,015	1,015	1,017
prof. 6	0,970	0,979	0,995
prof. 8	0,957	0,970	0,991

análisis exploratorio. El bosque supera además a Ridge (0,957 contra 0,994), diferencia atribuible a su capacidad de representar interacciones y respuestas no lineales.

10.3.3. Interpretación de los modelos

La Figura 19 presenta los coeficientes estandarizados del Ridge y las importancias del bosque.

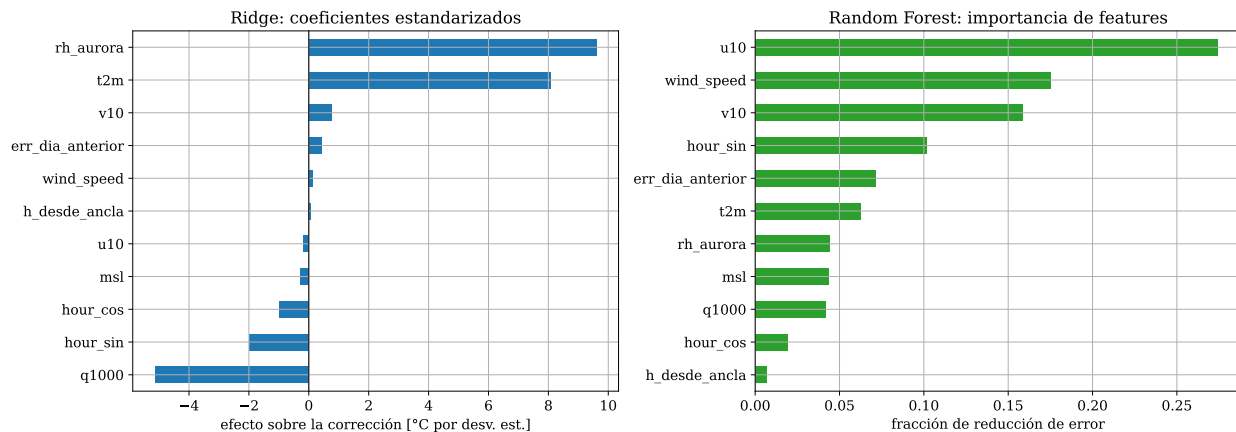


Figura 19: Interpretación de los modelos de corrección de temperatura. Izquierda: coeficientes estandarizados de Ridge ($\alpha = 0,1$), en °C de corrección por desviación estándar del predictor. Derecha: importancia de predictores del Random Forest. Elaboración propia.

En el bosque, los tres primeros lugares corresponden a variables de viento: la componente zonal 10U (0,274), el módulo (0,175) y la meridional 10V (0,158), que en conjunto concentran cerca del 60 % de la importancia. El modelo dedica la mayor parte de su capacidad a leer el régi-

men de viento, en consistencia con el análisis exploratorio, la velocidad del viento era el predictor más correlacionado con el error ($r = -0,57$) y la componente zonal distingue si el aire llega desde el mar o desde tierra, la alternancia que gobierna el sesgo costero. Siguen la codificación horaria (hour_sin , 0,102), el refinamiento temporal, y el error del día anterior (0,072), la persistencia del error entre días.

Los coeficientes de Ridge requieren una advertencia: los tres predictores termodinámicos (HR derivada +9,6, 2T +8,1 y Q1000 -5,1) están fuertemente correlacionados entre sí, por lo que sus coeficientes no admiten interpretación aislada, de modo que el modelo distribuye entre ellos un efecto conjunto. Sí es interpretable el par horario: hour_sin (-1,98) y hour_cos (-0,97) definen una senoide de período diario con amplitud $\sqrt{1,98^2 + 0,97^2} \approx 2,2$ °C: **el modelo lineal aprendió, en dos coeficientes, la onda del sesgo diurno que el análisis exploratorio midió empíricamente** (Figura 12b). El desplazamiento δ_t (coeficiente 0,07) resulta marginal una vez codificada la hora, según lo anticipado en la metodología.

10.3.4. Ensamble RF-Ridge

La etapa final evalúa si combinar modelos mejora al mejor individual. La Tabla 7 compara el MAE global de validación cruzada y el restringido al decil de correcciones verdaderas de mayor magnitud, criterio de desempate declarado en la metodología.

Los resultados ilustran en qué condiciones un ensamble aporta. El *Gradient Boosting*, que pertenece a la misma familia de árboles que el bosque, no logra mejorarlo (0,965 contra 0,957 °C), pues dos modelos de la misma familia comparten sesgos estructurales y carecen de errores descorrelacionados que puedan cancelarse entre sí. La mezcla convexa RF-Ridge, en cambio, reduce el MAE a 0,905 °C, un 5,4 % por debajo del mejor modelo individual, ya que las desviaciones de

Tabla 7: Validación cruzada del ensamble (temperatura): MAE global y MAE en el decil de correcciones de mayor magnitud [°C].

Modelo	MAE global	MAE extremos
Ridge ($\alpha = 0,1$)	0,994	2,171
Random Forest	0,957	1,886
Gradient Boosting	0,965	1,864
Mezcla RF–Ridge ($w = 0,5$)	0,905	1,985
<i>Barrido del peso w de la mezcla</i>		
$w = 0,4$	0,911	2,016
$w = 0,5$	0,905	1,985
$w = 0,6$	0,905	1,953
$w = 0,7$	0,910	1,922
$w = 0,8$	0,919	1,903
$w = 1,0$	0,957	1,886

un modelo de árboles y uno lineal provienen de mecanismos distintos y se cancelan parcialmente al promediarse. El barrido del peso muestra un valle plano entre $w = 0,5$ y $w = 0,6$ (ambos 0,905 °C), y se selecciona $w = 0,6$ por su mejor desempeño en el decil extremo (1,953 frente a 1,985 °C). La tabla también revela el costo de la combinación, ya que en los extremos el bosque puro supera a la mezcla (1,886 frente a 1,953 °C) porque la componente lineal modera las correcciones grandes; la mezcla intercambia así una pérdida acotada en el decil extremo por una ganancia global mayor.

Queda así configurado el sistema para la temperatura: **mezcla 60/40 de Random Forest y Ridge sobre el residual**, con MAE de validación cruzada de 0,905 °C, una reducción del 38 % respecto de B1 y del 26 % respecto de B3.

10.4. Evaluación final en el conjunto de prueba

Configurado el sistema, se procede a la evaluación única sobre el conjunto de prueba: los 8 días finales (25 de enero al 1 de febrero, 187 horas), nunca utilizados en ninguna decisión.

10.4.1. Métricas globales

La Tabla 8 presenta los resultados de todos los modelos y líneas base sobre las mismas 187 horas.

Tabla 8: Evaluación final de la temperatura en el conjunto de prueba (8 días, 187 horas).

Modelo	MAE [°C]	RMSE [°C]	Sesgo [°C]	R^2
B1 Aurora interpolada	1,496	2,164	−1,255	0,227
B2 sesgo constante	1,379	2,006	−0,958	0,335
B3 sesgo por hora	1,415	1,825	−1,009	0,450
Ridge ($\alpha = 0,1$)	1,406	1,815	−0,955	0,456
Random Forest	1,228	1,626	−0,786	0,563
Mezcla 60/40 (sistema final)	1,239	1,639	−0,854	0,557

El sistema final alcanza un MAE de 1,239 °C: mejora del 17,2 % sobre el pronóstico sin corregir y del 12,4 % sobre la corrección climatológica, con el sesgo reducido de −1,26 a −0,85 °C y el R^2 elevado de 0,23 a 0,56. Tres observaciones contextualizan estos valores.

Primero, todos los modelos rinden peor en test que en validación cruzada (el sistema pasa de 0,905 a 1,239 °C). La degradación es generalizada, afecta también a las líneas base, y se explica por el régimen del período de prueba, más cálido y con más extremos que el promedio del entrenamiento; al ser común a todos los métodos, las comparaciones relativas se mantienen, pero las magnitudes absolutas son condicionales al régimen evaluado.

Segundo, el mismo corrimiento produce una inversión aparente: B2 supera a B3 en MAE (1,379 contra 1,415 °C), al contrario que en validación cruzada. En un período dominado por el sesgo frío vespertino, la corrección constante resulta accidentalmente favorecida. Las métricas cuadráticas conservan el orden esperado (RMSE de 2,006 contra 1,825; R^2 de 0,335 contra 0,450).

Tercero, el bosque puro obtiene en test un MAE marginalmente inferior al de la mezcla (1,228 contra 1,239 °C). El protocolo exige reportar como sistema final el configurado por valida-

ción cruzada, la mezcla, pues elegir retrospectivamente al ganador del test constituiría precisamente la fuga de información que el diseño evita; la diferencia, inferior al 1 %, es consistente con la variabilidad muestral de 8 días.

10.4.2. Comportamiento horario y episodio extremo

La Figura 20 muestra las series del período de prueba y la Figura 21 el MAE por hora del día. La mejora se concentra exactamente donde el análisis exploratorio localizó el problema: en la franja 16:00–19:00 el sistema reduce el error de B1 de 2,9–3,2 °C a 1,7–2,4 °C, mientras que en las horas fáciles de la madrugada paga un sobreprecio pequeño respecto del pronóstico crudo, allí casi insesgado.

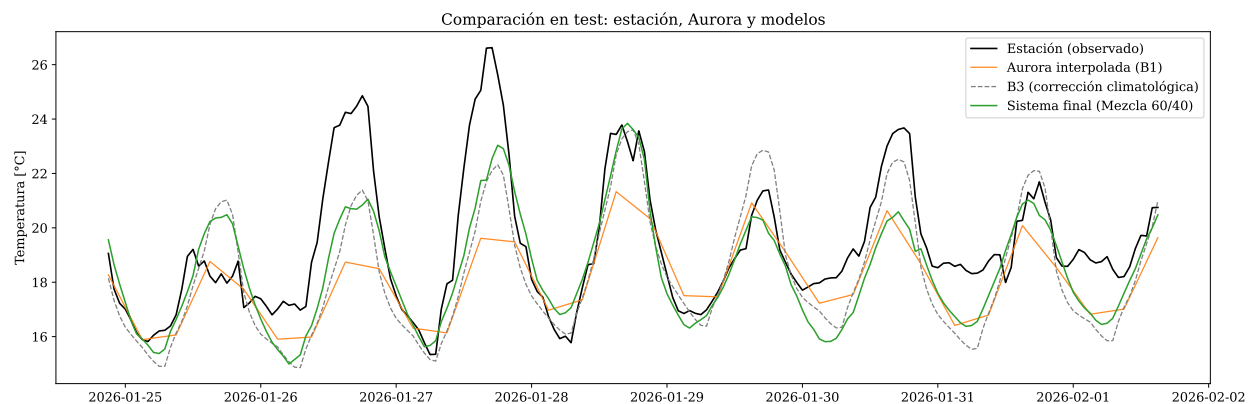


Figura 20: Series horarias del período de prueba: observación, Aurora interpolada (B1), corrección climatológica (B3) y sistema final. Elaboración propia.

El período de prueba incluye el episodio más exigente del registro: la máxima absoluta de 26,6 °C del 27 de enero (Figura 22). Aurora, saturada, pronosticó 19,6 °C para esa hora; el sistema corrigió hasta 22,5 °C, recuperando 2,9 de los 7,0 °C de déficit. En el conjunto de horas extremas del test (> 24 °C), el MAE se reduce de 5,93 °C (Aurora) a 3,51 °C (sistema): una mejora sustancial que no alcanza la magnitud plena de los eventos. La limitación es estructural y

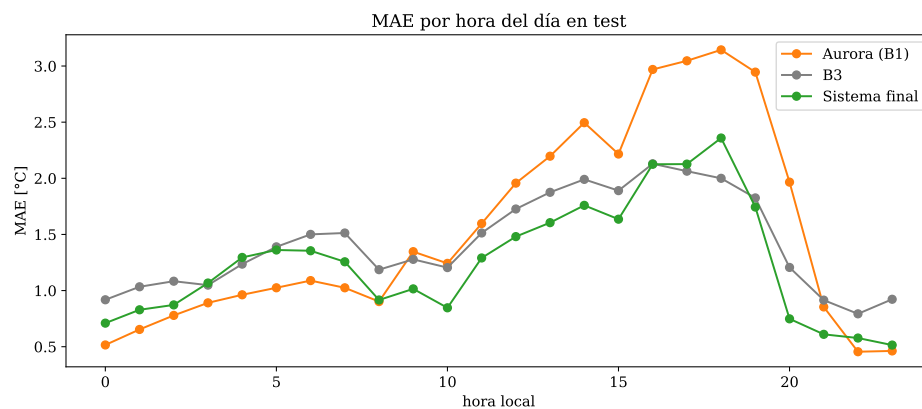


Figura 21: MAE por hora del día en el conjunto de prueba para B1, B3 y el sistema final. Elaboración propia.

doble: el bosque no extrapola correcciones más allá del rango de entrenamiento, y la firma seccionada de los predictores discrimina la *ocurrencia* del evento mejor que su *magnitud* exacta. El día siguiente, de máxima más moderada (23,7 °C), fue capturado casi exactamente (23,8 °C): la pérdida se concentra solo en los valores más allá del rango de entrenamiento.

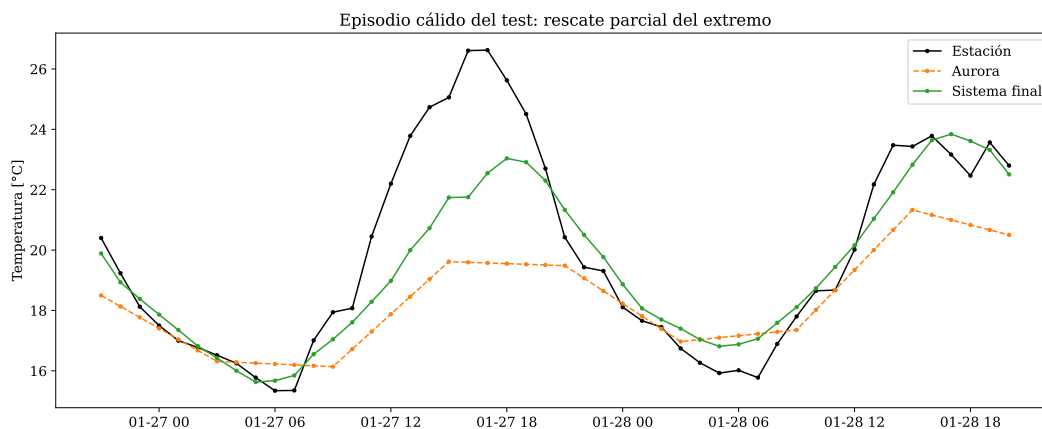


Figura 22: Episodio cálido del período de prueba (27–28 de enero): observación, Aurora interpolada y sistema final. El sistema recupera parcialmente el déficit en la máxima absoluta (26,6 °C) y captura casi exactamente la máxima moderada del día siguiente (23,7 °C). Elaboración propia.

10.4.3. Análisis de residuos

La Figura 23 examina los residuos del sistema final. La correlación entre lo predicho y lo observado alcanza $r = 0,823$ (panel a), frente al 0,730 del pronóstico crudo. El residuo medio de $-0,85$ °C (panel c) confirma que el sesgo remanente es frío y, según el panel (b), creciente con la magnitud observada: la subcorrección se concentra en las horas más cálidas, en coherencia con el episodio extremo.

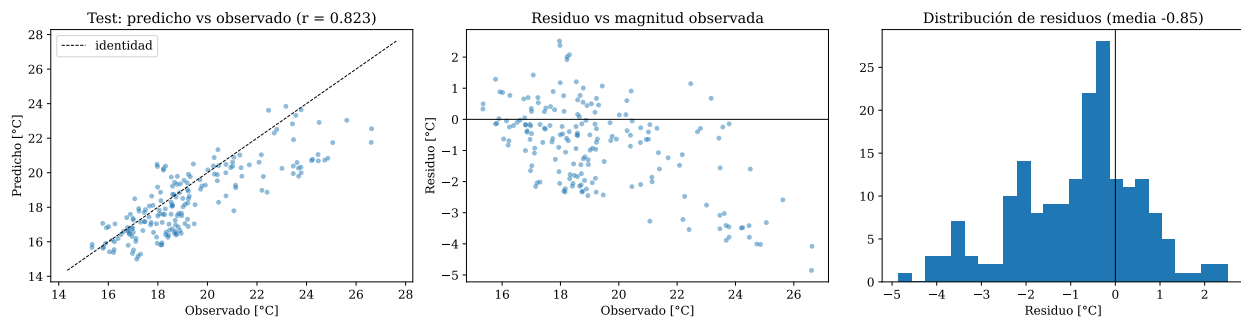


Figura 23: Análisis de residuos del sistema final en el conjunto de prueba. (a) Predicho contra observado ($r = 0,823$). (b) Residuo contra magnitud observada. (c) Distribución de los residuos (media $-0,85$ °C). Elaboración propia.

Para evaluar si quedó señal sin explotar, se calcula la correlación del residuo con cada predictor. La mayor corresponde al error del día anterior ($\bar{e}_{d-1}$, $+0,31$), correlación modesta que sugiere que la persistencia del error entre días no queda explotada por completo; las restantes no superan $|r| = 0,21$ (la presión MSL, $-0,21$). La única estructura apreciable es contra la propia magnitud observada (panel b), variable que el modelo no conoce por construcción: el sistema alcanzó el techo de la información de sus predictores, y el error restante requiere información que las salidas de Aurora no portan a esta escala, no un modelo más capaz sobre las mismas entradas.

10.5. Sistema multivariable

La receta completa (formulación residual, 11 predictores, validación cruzada agrupada, líneas base y evaluación única en test) se aplica de manera idéntica a las cuatro variables. En validación cruzada, el bosque resulta seleccionado para humedad, viento y presión; en humedad la diferencia con Ridge es un empate técnico (4,878 contra 4,845 %, un 0,7 %) resuelto por uniformidad del sistema, y en temperatura se emplea la mezcla RF–Ridge según lo establecido. La Tabla 9 presenta la evaluación final completa.

10.5.1. Desempeño por variable

Tabla 9: Evaluación final del sistema multivariable en el conjunto de prueba (187 horas). En negrita, el modelo final adoptado para cada variable: el seleccionado por validación cruzada, salvo en humedad, donde Ridge y RF empatan en validación cruzada y se adopta RF por uniformidad del sistema.

Variable	Modelo	MAE	RMSE	Sesgo	R^2
Temp. [°C]	B1 Aurora	1,496	2,164	−1,255	0,227
	B3 sesgo/hora	1,415	1,825	−1,009	0,450
	Ridge	1,406	1,815	−0,955	0,456
	RF (residual)	1,228	1,626	−0,786	0,563
	Mezcla 60/40	1,239	1,639	−0,854	0,557
Hum. rel. [%]	B1 Aurora	6,133	7,825	+4,120	0,475
	B3 sesgo/hora	5,897	7,445	+3,115	0,524
	Ridge	4,893	6,664	+2,469	0,619
	RF (residual)	4,709	6,203	+2,268	0,670
	RF (directo)	4,403	6,111	+2,005	0,641
Viento [m/s]	B1 Aurora	1,091	1,255	+1,069	−1,166
	B3 sesgo/hora	0,453	0,582	−0,033	0,535
	Ridge	0,426	0,540	−0,024	0,599
	RF (residual)	0,395	0,496	+0,044	0,662
	RF (directo)	0,403	0,511	+0,005	0,641
Presión [hPa]	B1 Aurora	6,414	6,551	−6,414	−10,654
	B3 sesgo/hora	1,090	1,336	−0,468	0,515
	Ridge	0,917	1,106	−0,244	0,668
	RF (residual)	1,068	1,243	−0,335	0,580
	RF (directo)	1,068	1,243	−0,335	0,580

El sistema final reduce el MAE respecto del pronóstico sin corregir en un 17,2 % para la

temperatura, 23,2 % para la humedad, 63,8 % para el viento y 83,3 % para la presión. La heterogeneidad de estas cifras no es ruido: refleja la naturaleza distinta del error de cada variable.

10.5.2. Interpretación por variable

Humedad relativa. La ganancia (de 6,13 a 4,71 %, con el R^2 de 0,48 a 0,67) sigue el patrón de la temperatura, en coherencia con el origen compartido de ambos errores donde una parte sustancial del sesgo húmedo es la proyección del sesgo térmico (Subsubsección 10.2.7), y el modelo la corrige con los mismos predictores. El sesgo se reduce de +4,1 a +2,3 %.

Viento. Es la variable de mayor mejora relativa (63,8 %) y la única de error genuinamente atmosférico donde el pronóstico crudo resulta peor que predecir el promedio observado ($R^2 = -1,17$): su sesgo de +1,07 m/s supera a la media observada del período (0,91 m/s). La presión exhibe un R^2 aún más negativo ($-10,65$), pero por una causa distinta —el desplazamiento instrumental constante de la consola, analizado más adelante— y no por una deficiencia del pronóstico atmosférico. La descomposición de la ganancia es ilustrativa: la mayor parte es capturada por la corrección climatológica (B3 reduce el MAE a 0,453 m/s, eliminando el sesgo horario de la Subsubsección 10.2.6), y el aprendizaje automático aporta un 12,8 % adicional (0,395 m/s). La formulación directa de la magnitud, evaluada por el carácter multiplicativo de la corrección, resulta equivalente a la residual (0,403 contra 0,395 m/s); se conserva la residual por uniformidad y por su degradación controlada.

Presión barométrica. La mejora del 83,3 % requiere la lectura anatómica comprometida en la metodología. El error de B1 es esencialmente puro sesgo: su MAE (6,414 hPa) coincide con el valor absoluto de su sesgo, característica de un desplazamiento constante atribuible a la calibración del ajuste a nivel del mar de la consola (equivalente a unos 50 m de diferencia de altitud).

Removerlo no requiere aprendizaje: la corrección implícita en B3 reduce el MAE a 1,090 hPa, y el aporte del modelo sobre esa referencia es marginal (RF: 1,068 hPa, un 2 % bajo B3). Por transparencia se consigna que en el test Ridge obtuvo un MAE inferior al del bosque seleccionado (0,917 contra 1,068 hPa), invirtiendo el orden de la validación cruzada (0,748 contra 0,718 hPa); conforme al protocolo, se reporta el modelo elegido por validación cruzada, y la inversión se atribuye a variabilidad esperable en una variable de error remanente pequeño. Para la presión, la adaptación local consiste esencialmente en una recalibración instrumental, valiosa para el uso operacional de la estación, más un refinamiento menor del aprendizaje automático.

10.6. Síntesis

La Tabla 10 resume el sistema final (ilustración del sistema de temperatura disponible en anexo Figura 27) y su desempeño, distinguiendo la mejora total sobre el pronóstico sin corregir (vs. B1) de la atribuible específicamente al aprendizaje automático por sobre la corrección climatológica (vs. B3).

Tabla 10: Resumen del sistema final de adaptación local. El modelo se selecciona por validación cruzada (en humedad, donde Ridge y RF empatan en CV, se adopta RF por uniformidad). La mejora se reporta respecto del pronóstico sin corregir (B1) y respecto de la corrección climatológica horaria (B3), que aísla el aporte del aprendizaje automático.

Variable	Modelo final adoptado	MAE test	Mejora vs. B1	Mejora vs. B3
Temperatura [°C]	Mezcla 60/40 RF–Ridge	1,239	17,2 %	12,4 %
Humedad rel. [%]	RF residual	4,709	23,2 %	20,1 %
Viento [m/s]	RF residual	0,395	63,8 %	12,8 %
Presión [hPa]	RF residual	1,068	83,3 %	2,0 %

La lectura transversal admite una formulación compacta: **el valor de la adaptación local, medido como mejora total sobre el pronóstico crudo (vs. B1), es proporcional a la magnitud de la discrepancia entre la escala de la celda y la escala del punto.** Es máximo en el viento, don-

de la representatividad local domina por completo al pronóstico crudo; sustancial en la temperatura y la humedad, donde la celda costera atenúa los ciclos y extremos continentales; y, descontado el componente instrumental, marginal en la presión, variable de gran escala que la celda ya representa bien.

Al aislar el aporte propio del aprendizaje automático (vs. B3), el orden se reordena de forma reveladora: la contribución del modelo por sobre la corrección climatológica es del orden de 12–20 % en temperatura, humedad y viento, y de apenas 2 % en la presión, donde la mejora es esencialmente recalibración instrumental. Es decir, gran parte de la mejora total del viento y casi toda la de la presión provienen de correcciones que no requieren aprendizaje (interpolación y climatología horaria), mientras que el valor diferencial del modelo se concentra de manera más pareja en las cuatro variables. Esta descomposición, que la arquitectura de líneas base hace posible, es la distinción sin la cual las cifras agregadas resultarían engañosas.

11. Conclusiones

El presente trabajo desarrolló y evaluó un sistema de downscaling estadístico y adaptación local que, mediante aprendizaje automático, corrige los pronósticos del modelo fundacional Aurora hacia las condiciones observadas en una estación Davis Vantage Pro 2 instalada en el campus de la Universidad Técnica Federico Santa María. El sistema aborda de manera conjunta las dos limitaciones que separan un pronóstico global de una observación puntual, la falta de representatividad local de la celda de $0,25^\circ$ y su resolución temporal de 6 horas. Sobre el conjunto de prueba independiente, reduce el error absoluto medio respecto del pronóstico sin corregir en un 17,2 % en temperatura, 23,2 % en humedad relativa, 63,8 % en velocidad del viento y 83,3 % en presión barométrica, cumpliendo así el objetivo general planteado.

Los resultados sobre la temperatura, variable de desarrollo del sistema, sintetizan el desempeño alcanzado. El error absoluto medio se redujo de $1,496^\circ\text{C}$ en el pronóstico sin corregir a $1,239^\circ\text{C}$, el coeficiente de determinación se elevó de 0,23 a 0,56 y el sesgo se atenuó de $-1,26$ a $-0,85^\circ\text{C}$. La mejora se concentró en la franja vespertina, donde se acumulaban las dos fuentes de error identificadas en el diagnóstico, mientras que en las horas de madrugada, en que el pronóstico ya resultaba prácticamente insesgado, el sistema introdujo un costo menor. Este comportamiento confirma que la corrección actúa donde existe estructura aprovechable y no de manera indiscriminada.

La descomposición de la mejora frente a las líneas base constituye el aporte metodológico más relevante. Una parte sustancial de las reducciones más llamativas no proviene del aprendizaje automático, sino de operaciones que no lo requieren, como la remoción de un desplazamiento instrumental constante en la presión y la corrección climatológica horaria en el viento. Al aislar la

contribución propia del modelo respecto de esa referencia, su valor se vuelve notablemente parejo, del orden de 12 a 20 % en temperatura, humedad y viento, y de apenas 2 % en la presión. Es decir, gran parte de la mejora total del viento y casi toda la de la presión provienen de correcciones sin aprendizaje, mientras que el valor diferencial del modelo se reparte de manera más homogénea entre las cuatro variables. Esta distinción, posible gracias a la arquitectura de líneas base, es la que impide que las cifras agregadas resulten engañosas.

De la lectura conjunta de las cuatro variables se desprende una conclusión transversal. El valor de la adaptación local, medido como mejora total sobre el pronóstico crudo, es proporcional a la magnitud de la discrepancia entre la escala de la celda y la escala del punto. Resulta máximo en el viento, donde la representatividad local domina por completo al pronóstico sin corregir; sustancial en la temperatura y la humedad, donde la celda atenúa los ciclos y los extremos del punto de observación; y marginal en la presión, una variable de gran escala que la celda ya representa adecuadamente. La selección de los modelos se realizó en todo momento mediante validación cruzada agrupada por días, con el conjunto de prueba sellado hasta la evaluación final, lo que respalda que las mejoras reportadas reflejan capacidad de generalización y no ajuste a los datos de evaluación.

En conjunto, el trabajo demuestra que la información de un modelo fundacional global puede adaptarse a un punto de observación específico con métodos ejecutables en recursos computacionales convencionales, sin reentrenar ni ajustar el modelo base. La corrección obtenida es además interpretable, ya que cada ganancia admite un mecanismo físico identificable, lo que la distingue de una mejora puramente estadística y refuerza la confianza en su comportamiento. La combinación de un modelo fundacional reciente con una única estación de costo accesible constituye, hasta donde alcanza la revisión realizada, una configuración no abordada previamente en la

literatura, y sugiere una vía económica para acercar los pronósticos de los modelos de inteligencia artificial al nivel de detalle que requieren las aplicaciones locales.

Los alcances del estudio acotan estas conclusiones a su contexto, una corrección de grilla a punto sobre un único emplazamiento, un período acotado a la estación de verano y un límite estructural en la reconstrucción de los eventos extremos, aspectos que se desarrollan en el Capítulo 10.6. Lejos de invalidar los resultados, estas fronteras delimitan el dominio de validez del sistema y trazan las direcciones naturales para su extensión, entre ellas la incorporación de múltiples estaciones, la cobertura de las restantes estaciones del año y la exploración de formulaciones capaces de extrapolar más allá del rango de valores observado durante el entrenamiento.

12. Limitaciones

Generalización espacial. El sistema es una corrección de grilla a punto entrenada sobre una única estación, por lo que aprende la relación entre la celda y ese emplazamiento específico, no transferible a otra ubicación sin reentrenamiento. No constituye, por tanto, un método de downscaling espacial general, sino una adaptación local sitio-específica. La incorporación de múltiples estaciones permitiría estudiar la transferibilidad de la corrección.

Cobertura temporal. El período abarca 40 días de verano, de modo que los modelos caracterizan los regímenes propios de esa estación y no los de las restantes. Las relaciones condicionales aprendidas podrían no sostenerse fuera del verano. La propia evaluación lo sugiere: dentro del mismo verano, el corrimiento de régimen del conjunto de prueba ya degradó el desempeño de 0,905 a 1,239 °C en temperatura, por lo que un cambio estacional completo tendría un efecto mayor.

Reconstrucción de los eventos extremos. La limitación más relevante afecta a los eventos de mayor interés práctico, el sistema recuperó solo 2,9 de los 7,0 °C de déficit en la máxima absoluta del período. Esto producto que el Random Forest no extrapola más allá del rango de entrenamiento. Sumado a lo anterior, la cantidad de datos no era suficiente para que el modelo aprendiese a ajustarse a tales eventos.

Robustez de la evaluación y supuestos operacionales. El conjunto de prueba es adecuado para una estimación pero reducido para sustentar diferencias pequeñas entre modelos. Por último, la inicialización emplea el reanálisis ERA5, no disponible con la latencia de un sistema operacional real (existe un desfase de 5 días), y el predictor de error del día anterior supone observaciones previas siempre disponibles y depuradas.

13. Anexos

Arquitectura Aurora

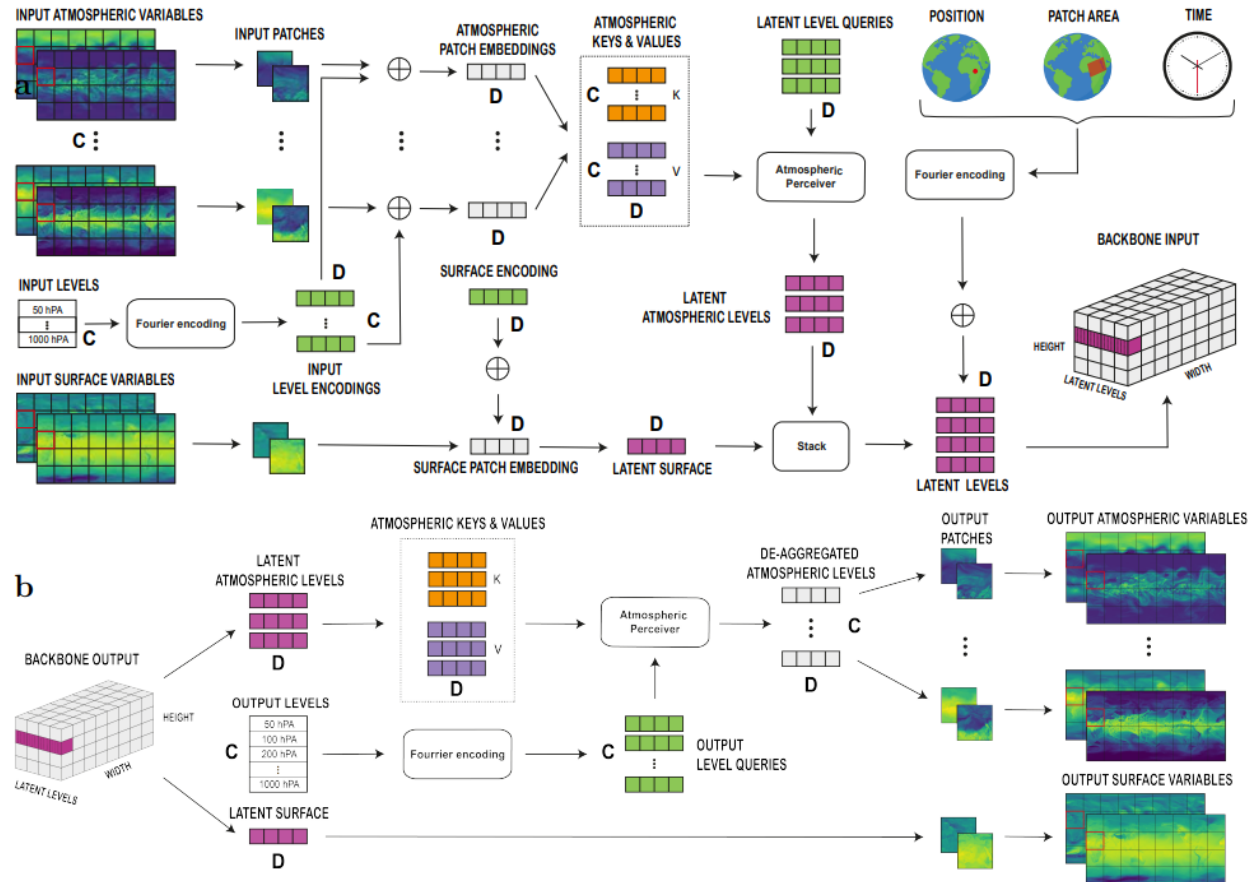


Figura 24: Arquitectura codificador–decodificador de Aurora con representación latente tridimensional. (a) Codificador: los estados meteorológicos de entrada (variables atmosféricas en niveles de presión y variables de superficie) se dividen en parches y se proyectan a *embeddings*, que bloques de atención cruzada de tipo Perceiver comprimen en una representación latente 3D con un número reducido de niveles verticales. Los tokens latentes se aumentan con codificaciones de posición, área de parche y tiempo mediante codificación de Fourier, que aportan la información espacial, temporal y de escala. (b) Decodificador: revierte las operaciones del codificador, deagregando los niveles latentes al conjunto de niveles de presión deseado mediante una capa Perceiver y decodificando dinámicamente a parches con capas lineales específicas por variable, para reconstruir las variables de salida sobre la grilla latitud–longitud. En el esquema, C denota el número de variables o niveles y D la dimensión del *embedding*. Fuente: Bodnar et al. (2024).

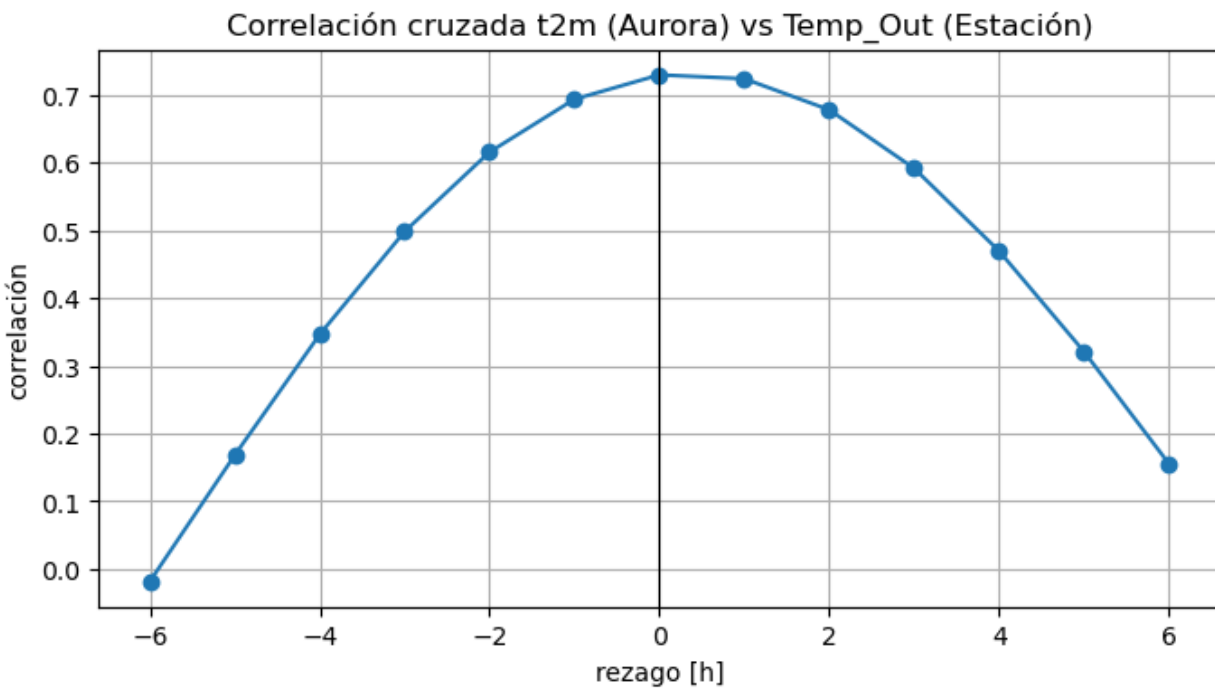
Correlación cruzada: temperatura Aurora – temperatura estación

Figura 25: Verificación del alineamiento temporal entre fuentes mediante correlación cruzada. Cada punto corresponde a la correlación de Pearson entre la temperatura pronosticada por Aurora y la observada por la estación, con la serie observada desplazada en el rezago indicado. El máximo en rezago cero ($r = 0,730$) confirma la correcta conversión de la referencia temporal de UTC a hora local (UTC−3). Elaboración propia.

Campos de temperatura: cuatro pasos del rollout

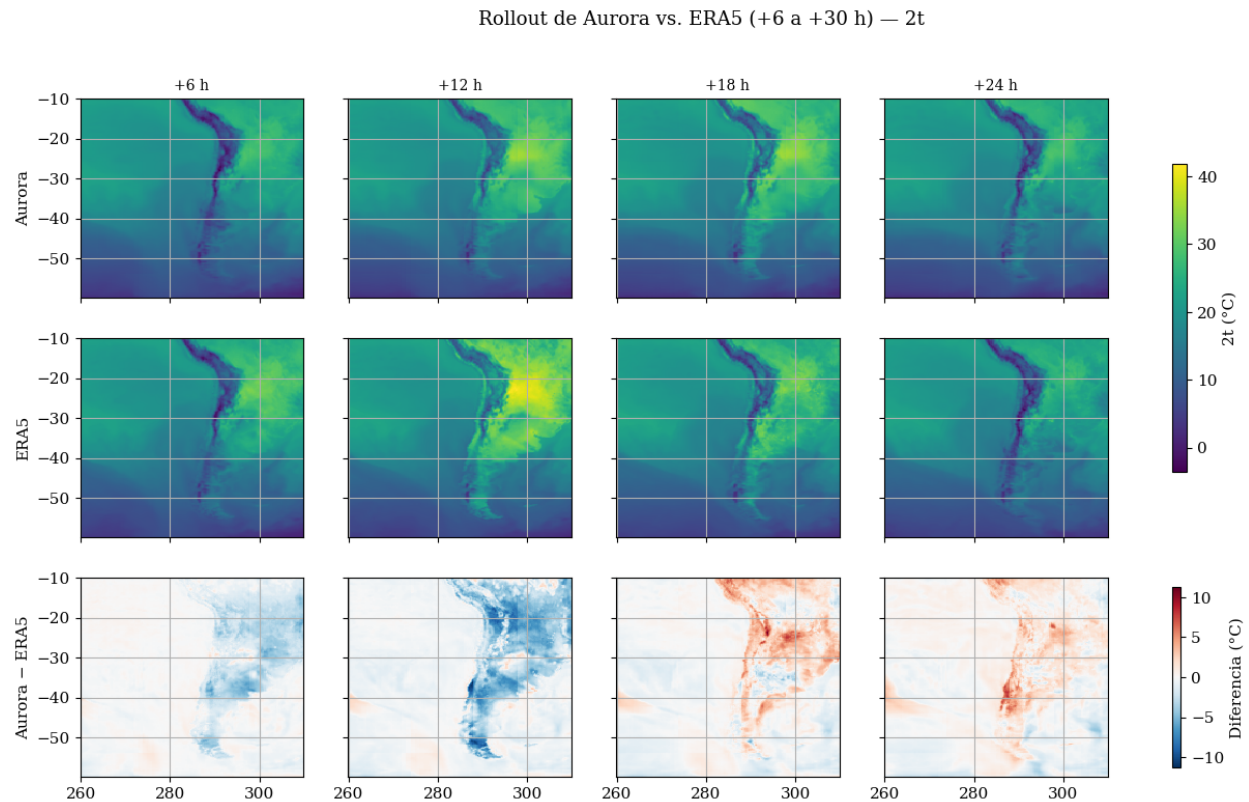


Figura 26: Campos de temperatura a 2 metros sobre el dominio de estudio durante los cuatro pasos de la ventana operacional (+6 a +24 h) del despliegue autorregresivo de 20 pasos. Fila superior: pronóstico de Aurora; fila central: reanálisis ERA5 de referencia; fila inferior: diferencia (Aurora – ERA5). En los cuatro horizontes, Aurora reproduce la estructura de gran escala de la temperatura (el gradiente cordillerano, el contraste continente–océano y el núcleo cálido continental), con diferencias acotadas y espacialmente desorganizadas: el RMSE normalizado de 2T se mantiene entre 0,21 y 0,32 en toda la ventana, muy por debajo de la variabilidad espacial del campo. La evolución del signo de la diferencia no responde a una degradación autorregresiva, sino al ciclo diurno del instante verificado: tonos fríos (azules) hacia +12 h, cuando Aurora subestima levemente el núcleo cálido continental, y tonos cálidos (rojos) hacia +18 y +24 h sobre el centro del continente. Elaboración propia.

Representación del sistema de corrección de temperatura

Modelo de corrección de temperatura: pesos aprendidos

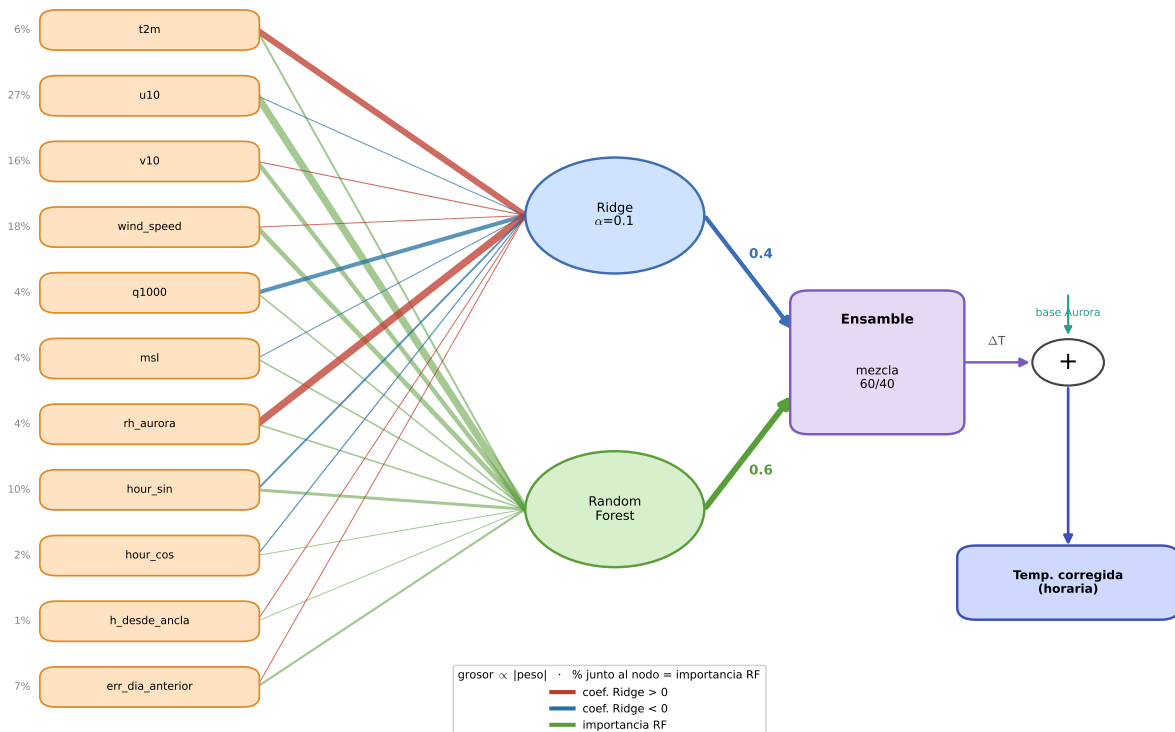


Figura 27: Representación integrada del sistema de corrección de temperatura y de los pesos aprendidos por sus dos componentes. Los once predictores alimentan de forma simultánea un modelo lineal (Ridge, $\alpha = 0,1$) y un Random Forest, cuyas correcciones sobre el residuo (ΔT respecto del pronóstico base de Aurora) se combinan en la mezcla convexa 60/40 que constituye el sistema final, entregando la temperatura corregida a resolución horaria. El grosor de cada conexión es proporcional a la magnitud del peso del predictor; el porcentaje junto a cada nodo indica su importancia en el Random Forest, y el color distingue el signo del coeficiente de Ridge. El esquema sintetiza en una sola vista el predominio de las variables de viento (la componente zonal 10U con un 27 %, el módulo con un 18 % y la componente meridional 10V con un 16 %, que concentran cerca del 60 % de la importancia del bosque), seguidas de la codificación horaria (10 %) y del error del día anterior (7 %). Elaboración propia.

Referencias

- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q. (2023). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619:533–538.
- Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Vaughan, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J. A., Dong, H., Gupta, J. K., Thambiratnam, K., Archibald, A. T., Wu, C.-C., Heider, E., Welling, M., Turner, R. E., and Perdikaris, P. (2024). A foundation model for the earth system. *arXiv preprint arXiv:2405.13063*.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Davis Instruments Corp. (2016). *Vantage Pro2 Console Manual / Manual de la Estación Vantage Pro2*. Davis Instruments Corp., Hayward, CA, EE.UU.
- Glahn, H. R. and Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, 11(8):1203–1211.
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.
- Park, H., Park, S., Kang, D., and Kim, J.-H. (2026). A super-resolution framework for downscaling machine learning weather prediction toward 1-km air temperature. *npj Climate and Atmospheric Science*, 9:56.
- 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., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Zhang, F., Sun, Y. Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H., and Emanuel, K. (2019). What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, 76(4):1077–1091.