

PRONÓSTICO ESPACIOTEMPORAL DE PRECIPITACIONES EN CHILE
UTILIZANDO REDES NEURONALES Y MÉTODOS ESTADÍSTICOS AVANZADOS

POR: MARTÍN EDUARDO MARCHANT FERRADA

SERGIO FELIPE SILVA URIBE

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dr. Mauricio Herrera Marín

Diciembre 2024

SANTIAGO

AGRADECIMIENTO

Martín Marchant

Quiero agradecer a mi amor Melissa por apoyarme y ayudarme durante este proceso y a mis padres que han sido un pilar fundamental en mi educación.

También quiero expresar mi agradecimiento a mi compañero Sergio Silva por toda su dedicación y compromiso no solo en este trabajo, sino también, en todos los proyectos que hemos realizado juntos durante el Magister.

Y por último, agradecer todo el apoyo que me ha brindado mi amigo de toda la vida, Eduardo Ibacache.

Sergio Silva

Agradezco a mis padres y a mi familia por el apoyo y confianza en mi formación académica en general. A Martin por el apoyo durante todo el programa del magíster lo cual ayudó a que sortearemos con éxito muchas dificultades que se nos dieron durante este arduo camino.

TABLA DE CONTENIDO

RESUMEN	1
1. INTRODUCCIÓN	2
2. TRABAJO RELACIONADO	3
3. HIPÓTESIS Y OBJETIVOS	4
4. DATOS Y METODOLOGÍA	5
4.1. DATOS.....	5
4.2. METODOLOGÍA.....	5
5. RESULTADOS	6
6. CONCLUSIÓN Y TRABAJO FUTURO	7
BIBLIOGRAFÍA	8

Resumen

El objetivo de esta investigación es evaluar la efectividad del uso combinado de un Autoencoder (AE) y la Descomposición Modal Dinámica (DMD) en la predicción de patrones espacio-temporales de precipitaciones en Chile. Se propone un enfoque innovador que integra estas metodologías para capturar la complejidad dinámica de los datos climáticos, utilizando una representación compacta en un espacio latente y su posterior descomposición dinámica.

Los resultados obtenidos evidencian una alta precisión en la predicción de precipitaciones, con un destacado desempeño en la identificación de patrones dominantes y recurrentes en las series temporales. Además, se resalta el bajo costo computacional del modelo, ya que todo el proceso se ejecutó exitosamente en un equipo de cómputo estándar, lo que refuerza su aplicabilidad en contextos con recursos limitados. Este enfoque permitió generar predicciones locales, logrando un nivel de detalle y especificidad superior al de los modelos globales tradicionales, crucial para la toma de decisiones en áreas específicas de Chile.

En síntesis, esta investigación demuestra la viabilidad y precisión de esta metodología para predecir precipitaciones en Chile, destacando su potencial para aplicaciones climáticas en entornos específicos, ya que son predicciones locales. Su enfoque basado en la integración de técnicas modernas de machine learning y análisis modal dinámico

resalta como una solución eficiente e innovadora en el campo de la modelización climática.

1. Introducción

En la actualidad, la predicción del clima es una tarea fundamental para enfrentar los desafíos impuestos por el cambio climático y sus efectos en diversas áreas como la agricultura, la gestión hídrica y la planificación urbana. Sin embargo, la predicción precisa de fenómenos climáticos, como las precipitaciones, es particularmente compleja debido a las interacciones espaciales y temporales que caracterizan los datos climáticos.

En este contexto, la modelización y pronóstico de precipitaciones en zonas geográficas específicas, como Chile, requieren de métodos avanzados capaces de capturar las dinámicas subyacentes y las correlaciones espaciales y temporales sin recurrir a recursos computacionales excesivos.

Chile presenta una notable heterogeneidad climática, con diversas zonas geográficas que experimentan patrones de precipitación y variabilidad estacional muy distintos. Esta complejidad hace que los modelos globales de predicción, que funcionan a gran escala, no sean siempre adecuados para generar pronósticos locales de alta precisión. Además, los costos computacionales y monetarios asociados a la implementación de modelos globales en territorios específicos representan un desafío adicional.

Este proyecto tiene como objetivo desarrollar un enfoque innovador para el pronóstico de precipitaciones en Chile mediante el uso combinado de redes neuronales y métodos

estadísticos avanzados, específicamente a través de un modelo híbrido que integre autoencoders (AE) y la descomposición en modos dinámicos (DMD). Este enfoque busca reducir la dimensionalidad del problema de predicción, disminuyendo así la carga computacional sin comprometer la precisión de las predicciones. Al utilizar el autoencoder, se transforma la serie temporal multivariada original en un espacio latente de menor dimensión, donde se pueden aplicar técnicas como la DMD para capturar los patrones dinámicos espacio-temporales de las precipitaciones.

La combinación de estas herramientas permitirá realizar pronósticos de las precipitaciones en puntos específicos de Chile, sin los altos costos asociados a los modelos locales, y generando predicciones que sean no solo computacionalmente eficientes, sino también interpretables. Este trabajo tiene aplicaciones directas en áreas como la gestión hídrica y la adaptación al cambio climático, proporcionando herramientas prácticas y parsimoniosas para abordar la variabilidad climática de un territorio tan diverso como el chileno.

La importancia de este proyecto radica en su capacidad para transformar los datos climáticos multivariados y de alta dimensión en un problema manejable para cualquier computadora personal, ofreciendo una solución accesible para investigadores y profesionales del clima a nivel nacional. A través de este enfoque, se busca no solo mejorar la precisión de las predicciones climáticas, sino también proporcionar una

herramienta interpretativa que facilite la comprensión de los patrones climáticos y su evolución a lo largo del tiempo.

2. Trabajo Relacionado

La predicción del clima es crucial para entender y mitigar los impactos del cambio climático en la producción agrícola. Estudios como el de Lobell et al. (2008) destacan cómo las proyecciones climáticas pueden identificar regiones y cultivos que serán más afectados por el cambio climático. Por ejemplo, se proyecta que Asia del Sur y África Austral sufrirán impactos negativos significativos en varios cultivos importantes si no se implementan medidas de adaptación adecuadas. Estas predicciones permiten priorizar inversiones en adaptaciones agrícolas, como el desarrollo de nuevas variedades de cultivos y la expansión de la irrigación, para asegurar la seguridad alimentaria en el futuro.

DeepAR es una metodología avanzada para producir pronósticos probabilísticos precisos mediante el uso de redes recurrentes autorregresivas. Este enfoque entrena un modelo en una gran cantidad de series temporales relacionadas, mejorando así la precisión de los pronósticos. Utiliza una red neuronal recurrente basada en LSTM para modelar la distribución condicional de las series temporales futuras. El modelo es tanto autorregresivo como recurrente, lo que significa que emplea las observaciones pasadas y las salidas anteriores de la red para predecir los valores futuros.

En contraste, los métodos clásicos de pronóstico, como los modelos ARIMA y el suavizado exponencial, presentan limitaciones significativas cuando se trata de predecir

miles o millones de series temporales relacionadas. DeepAR supera estas limitaciones al aprender de datos históricos de múltiples series temporales.

En el artículo de Salinas, Flunkert y Gasthaus (2019) se observó que DeepAR mejora la precisión de los pronósticos en aproximadamente un 15% en comparación con métodos de última generación como ARIMA, técnicas de suavizado exponencial y modelos de espacio de estados. Esta mejora se demostró de manera consistente en todos los conjuntos de datos evaluados, lo que subraya la robustez y efectividad del modelo.

La predicción de la lluvia a corto plazo es un desafío crucial en la meteorología debido a los peligros que pueden causar las precipitaciones intensas, como inundaciones y deslizamientos de tierra. La capacidad de predecir con precisión la lluvia en un período corto (generalmente de 0 a 6 horas) es esencial para la gestión de recursos hídricos, la planificación urbana y la seguridad pública. En los últimos años, los avances en el aprendizaje automático y las redes neuronales han abierto nuevas posibilidades para la predicción de la lluvia a corto plazo. Modelos como Tiny-RainNet, que combinan redes neuronales convolucionales (CNN) con memoria a corto y largo plazo bidireccional (BiLSTM), han demostrado ser efectivos para predecir la lluvia directamente a partir de mapas secuenciales de ecos de radar (Zhang, Wang, Zeng, Ma, & Guan, 2020).

Las técnicas de inteligencia artificial, como las redes neuronales profundas, son herramientas útiles para el pronóstico de precipitaciones complejas a corto plazo (0-6

horas). Los modelos de aprendizaje profundo, como el LSTM, son capaces de aprender correlaciones espaciales y temporales de manera eficiente, proporcionando pronósticos precisos. Además, el LSTM puede capturar las dinámicas complejas de los sistemas convectivos de lluvia, que resultan difíciles de predecir con otros métodos (Akbari Asanjan et al., 2018).

El estudio "Unleashing the Power of Dynamic Mode Decomposition and Deep Learning for Rainfall Prediction in North-East India" investiga la eficacia de DMD en la predicción de precipitaciones utilizando datos diarios de lluvia recopilados por el Departamento Meteorológico de la India durante un período de 118 años. Los resultados mostraron que DMD puede predecir con precisión las precipitaciones, con valores de error cuadrático medio (RMSE) y error absoluto medio (MAE) que indican un rendimiento razonablemente preciso. Estos hallazgos sugieren que DMD puede mejorar significativamente la precisión de las predicciones de lluvia, ayudando a mitigar el impacto de eventos climáticos extremos y a mejorar la resiliencia de la región de India frente al cambio climático.

El uso de autoencoders en el análisis de datos ha demostrado ser una herramienta poderosa para la predicción meteorológica, especialmente en la modelización de variables múltiples. En el estudio de Man, Zhang, Feng, Li, y Shao (2023), se presenta un modelo de autoencoder enmascarado pre entrenado (W-MAE) que se utiliza para la predicción de variables meteorológicas. El autoencoder enmascarado se entrena para reconstruir las correlaciones espaciales dentro de las variables meteorológicas,

aprendiendo representaciones latentes significativas de los datos. Además, el autoencoder realiza una reducción de ruido en los datos de entrada, lo que mejora la calidad de las representaciones latentes y acelera el proceso de entrenamiento del modelo. Estas representaciones latentes se utilizan luego para predecir los estados futuros de las variables meteorológicas, mejorando la precisión de las predicciones al capturar las dependencias espaciales y temporales complejas presentes en los datos.

Esta tesis se centra en la integración de dos enfoques fundamentales para el análisis meteorológico: el Autoencoder (AE) y la Descomposición en Modos Dinámicos (DMD). En un contexto donde el creciente volumen y complejidad de los datos atmosféricos exige métodos que equilibren precisión y eficiencia computacional, la integración de DMD, AEs y DL no solo mejora la interpretabilidad y robustez de los modelos, sino que también sienta las bases para soluciones más eficientes y escalables en el análisis y pronóstico meteorológico.

3. Hipótesis y Objetivos

Hipótesis:

La utilización en conjunto del AutoEncoder (AE) con Dynamic Mode Decomposition (DMD) ¿Son una metodología precisa para la predicción de precipitaciones con un bajo costo computacional?

Objetivo General:

- Evaluar el rendimiento de la metodología AE + DMD con respecto al desempeño y costo computacional obtenido aplicando métodos de Deep Learning (DL) para las predicciones de precipitaciones en Chile.

Objetivos Específicos:

- Analizar los patrones de precipitaciones en Chile mediante el estudio de datos históricos y la identificación de tendencias y comportamientos climáticos relevantes para el modelado predictivo.
- Preparar y procesar los datos meteorológicos asegurando su calidad y consistencia, mediante técnicas de limpieza, normalización y selección de características relevantes.
- Implementar modelos de Deep Learning como baseline para establecer un punto de comparación en el pronóstico de precipitaciones.

- Desarrollar un modelo integrado basado en Autoencoders (AE) y Descomposición en Modos Dinámicos (DMD), destacando su capacidad para reducir la dimensionalidad y ahorro en costo computacional.
- Evaluar el desempeño de los modelos predictivos mediante métricas de error MAE y RMSE, comparando la eficacia de la metodología propuesta con los modelos baseline y los tiempos computacionales.

4. Datos y Metodología

4.1. Datos

Los datos utilizados en este proyecto provienen del conjunto de datos ERA5, proporcionado por el Servicio de Cambio Climático de Copernicus de la Unión Europea y producido por el Centro Europeo de Pronósticos Meteorológicos a Medio Plazo (ECMWF). ERA5 ofrece datos climáticos re-analizados con alta resolución espacial y temporal, que abarcan desde 1950 hasta el presente. Estos datos incluyen una variedad de variables atmosféricas y meteorológicas, tales como la precipitación, temperatura, y evapotranspiración, entre otras. Para este proyecto, nos centramos principalmente en la variable de precipitaciones, que fue descargada en formato NetCDF (.nc).

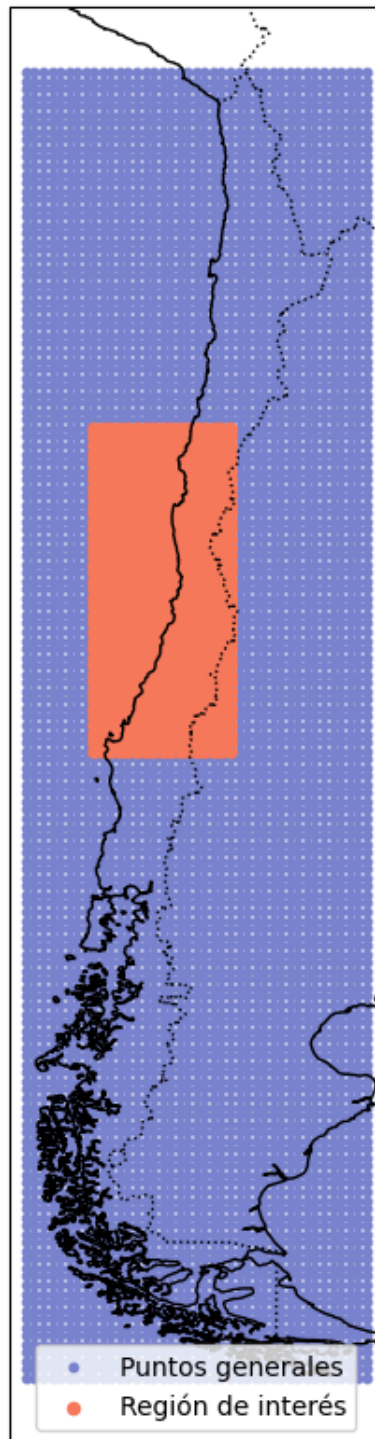
Los datos se estructuran en una malla de puntos de coordenadas geográficas, con longitudes entre -76.0° y -66.0° y latitudes entre -17.5° y -56.0° , con una resolución espacial de $0.25^{\circ} \times 0.25^{\circ}$. Esto da lugar a una malla de 6355 puntos espaciales, donde cada punto contiene series temporales diarias de datos climáticos entre 1980 y 2022. Sin embargo, para facilitar el procesamiento y abordar la complejidad computacional asociada con el trabajo con grandes volúmenes de datos, se decidió realizar un preprocesamiento para reducir el tamaño y la dimensionalidad de los datos, ajustándose al foco de este estudio: El pronóstico de precipitaciones en Chile.

Conjunto de datos para el Baseline

Para establecer una base de comparación con otros modelos, se seleccionaron los últimos cinco años de datos (de 2018 a 2022) del archivo NetCDF original, y se filtraron las estaciones dentro de las coordenadas geográficas correspondientes a las latitudes entre $-32^{\circ} 00' S$ y $-33^{\circ} 50' S$, y longitudes entre $-70^{\circ} W$ y $-72^{\circ} W$. Este recorte resultó en un conjunto de datos con 1825 filas (una por cada día en el periodo seleccionado) y 392 estaciones, con cada columna representando una combinación de latitud y longitud (es decir, las estaciones) que forman una malla espacial. Este conjunto de datos fue utilizado para entrenar los modelos de referencia (baseline) y evaluar su desempeño en la predicción de precipitaciones.

Conjunto de datos para la estrategia AE + DMD

Para la estrategia de Autoencoder (AE) y Descomposición en Modos Dinámicos (DMD), se utilizó un conjunto de datos más extenso, que abarca los últimos 40 años de datos (desde 1982 hasta 2022). Además, para capturar mejor las dinámicas espaciales y temporales de las precipitaciones en la zona central de Chile, se filtraron las estaciones correspondientes a latitudes entre $-28^{\circ} 00' S$ y $-33^{\circ} 50' S$, y longitudes entre $-70^{\circ} W$ y $-74^{\circ} W$. El conjunto final resultó en 14610 filas (una por cada día entre 1982 y 2022) y 663 estaciones, cada una representando una combinación de latitud y longitud. Este conjunto de datos fue estructurado como una malla espacial, siendo adecuado para la aplicación de técnicas avanzadas de reducción de dimensionalidad y pronóstico, como el Autoencoder y la Descomposición en Modos Dinámicos.



Mapa con los puntos generales y puntos de interés - Elaboración propia.

Para ambas estrategias, la elección de las coordenadas geográficas se debe a que el área es de particular interés debido a su importancia climática, económica y social. La zona central concentra la mayor parte de la población del país y es clave para la producción agrícola y la gestión de recursos hídricos, siendo responsable de una proporción significativa de los cultivos irrigados y los reservorios de agua potable. Además, el clima de esta región está fuertemente influenciado por fenómenos atmosféricos como el Anticiclón del Pacífico Sur y eventos como El Niño y La Niña, que afectan directamente las precipitaciones y, por ende, la disponibilidad de agua. Dado su papel estratégico en la economía y el bienestar de Chile, es crucial realizar predicciones de precipitaciones para esta zona, particularmente en un contexto de cambio climático que ha intensificado fenómenos como la mega sequía experimentada en las últimas décadas. La selección de esta región no solo busca capturar patrones espaciales y temporales relevantes, sino también aportar información valiosa para la planificación hídrica y la toma de decisiones en sectores clave.

AutoEncoder

Los autoencoders son redes neuronales diseñadas para aprender representaciones eficientes de los datos de entrada, comprimiéndolos en una representación de menor dimensión (codificador) y luego reconstruyéndolos (decodificador), manteniendo la mayor similitud posible con los datos originales. Durante el entrenamiento, el autoencoder minimiza la diferencia entre los datos de entrada y los datos reconstruidos,

utilizando una función de pérdida que mide esta diferencia (Bank, Koenigstein, & Giryas, 2021).

En el contexto de la predicción meteorológica, los autoencoders pueden capturar las correlaciones espaciales y temporales presentes en los datos meteorológicos, lo que es crucial para mejorar la precisión de las predicciones. Además, los autoencoders son especialmente útiles para la eliminación de ruido en los datos. Al entrenar el modelo con datos ruidosos y su correspondiente versión limpia, el autoencoder aprende a filtrar el ruido y a reconstruir una versión más precisa de los datos originales. Este proceso mejora la calidad de las representaciones latentes y, por ende, la precisión de las predicciones meteorológicas (Bank et al., 2021).

Otra aplicación importante de los autoencoders es la detección de anomalías. Al ser entrenados con datos normales, los autoencoders pueden identificar datos que no se ajustan a los patrones aprendidos, es decir, anomalías. Cuando se presentan datos anómalos, la reconstrucción del autoencoder es deficiente, lo que permite detectar estas irregularidades. Esta capacidad es particularmente valiosa en la predicción meteorológica, donde la identificación temprana de eventos anómalos puede ser crucial para la toma de decisiones (Bank et al., 2021).

Descomposición en Modos Dinámicos

La Descomposición en Modos Dinámicos (DMD) es una técnica de análisis de datos que se utiliza para descomponer sistemas dinámicos complejos en modos dinámicos simples. Esta técnica es particularmente útil para el análisis de datos temporales y espaciales, permitiendo identificar patrones y estructuras subyacentes en los datos. El DMD se basa en la descomposición de matrices y utiliza la Descomposición en Valores Singulares (SVD) para identificar modos dinámicos en los datos. Este método sigue los siguientes pasos:

- **Recolección de Datos:** Se recopilan datos de un sistema dinámico en diferentes instantes de tiempo. Estos datos se organizan en una matriz, donde cada columna representa el estado del sistema en un instante de tiempo específico.
- **Descomposición en Valores Singulares (SVD):** Se aplica la SVD a la matriz de datos para descomponerla en tres matrices: una matriz de modos espaciales U , una matriz diagonal de valores singulares Σ y una matriz de modos temporales V^* . Matemáticamente, esto se expresa como:

$$X = U\Sigma V^*$$

- **Construcción de la Matriz de DMD:** Se utiliza la SVD para construir una matriz de DMD que captura la dinámica del sistema. Esta matriz se descompone

en sus valores propios y vectores propios, que representan los modos dinámicos y sus correspondientes tasas de crecimiento o decaimiento.

- **Reconstrucción del Sistema:** Los modos dinámicos y sus coeficientes se utilizan para reconstruir el comportamiento del sistema en el tiempo, permitiendo predecir su evolución futura.

La capacidad de DMD para descomponer datos en modos dinámicos simples permite identificar patrones subyacentes y predecir la evolución futura del sistema. Además, al combinar la DMD con técnicas de aprendizaje profundo, como los autoencoders, se pueden obtener modelos aún más precisos y eficientes, lo que abre nuevas oportunidades para aplicaciones en diversas áreas, desde la predicción meteorológica hasta la biología y la ingeniería (Kutz et al., 2016).

4.2. Metodología

Construcción del Baseline

Para la construcción del baseline, se utilizaron los datos previamente descritos, estructurados como una malla espacial de 1825 filas (días) y 392 columnas (estaciones), correspondientes al período de 2018 a 2022 dentro de la zona geográfica de estudio.

	Time	(-32.00, -74.00)	(-32.00, -73.75)	(-32.00, -73.50)	(-32.00, -73.25)	(-32.00, -73.00)	(-32.00, -72.75)	(-32.00, -72.50)	(-32.00, -72.25)	(-32.00, -72.00)
Date										
2022-12-26	26658	0.004314	0.009168	0.007910	0.001618	0.000000	0.000000	0.000000	0.000000	0.000000
2022-12-27	26659	0.001078	0.000000	0.000000	0.000000	0.000539	0.000899	0.000899	0.000000	0.000000
2022-12-28	26660	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2022-12-29	26661	0.015100	0.015279	0.025168	0.034516	0.046740	0.064177	0.066515	0.048717	0.032899
2022-12-30	26662	0.253114	0.153702	0.116490	0.130691	0.050874	0.030560	0.024089	0.018695	0.019236

Estructura de datos - Elaboración propia.

A partir de este conjunto, se realizó una división temporal para entrenar y evaluar el modelo:

- **Train:** Se incluyó los datos de los primeros cuatro años del período (2018–2021).
- **Test:** Corresponde al año 2022, que fue completamente reservado para evaluar la calidad de las predicciones.

El modelo seleccionado fue DeepAR, una herramienta basada en aprendizaje profundo de la biblioteca GluonTS, diseñada para modelar series temporales multivariadas.

DeepAR utiliza redes neuronales recurrentes (RNNs) y distribuciones probabilísticas, lo que lo hace especialmente adecuado para manejar datos de conteo, como precipitaciones. Además, su capacidad para capturar patrones temporales en múltiples puntos simultáneamente lo posiciona como una línea base robusta para este estudio.

Esta elección se fundamenta en sus capacidades diferenciadoras:

- DeepAR puede manejar múltiples series temporales simultáneamente, capturando relaciones temporales y espaciales entre estaciones de manera eficiente, algo que modelos tradicionales como ARIMA no logran.
- Su capacidad para generar intervalos de confianza mediante distribuciones probabilísticas es crucial para fenómenos climáticos con alta variabilidad, como las precipitaciones.
- DeepAR está diseñado para trabajar con datos de alta dimensionalidad, como las precipitaciones en múltiples estaciones, superando las limitaciones de enfoques como Random Forest o ARIMA en este tipo de problemas.

Por estas razones, DeepAR es el modelo más indicado para establecer un baseline sólido y competitivo, garantizando una comparación adecuada con el modelo AE+DMD.

El modelo fue configurado para realizar predicciones diarias de precipitación durante todo el año 2022, reconstruyendo las series temporales para los 392 puntos de la malla. Este proceso generó predicciones detalladas, permitiendo evaluar el desempeño tanto a nivel global como en puntos específicos.

Selección de dos puntos representantes

Además de analizar las métricas en toda la región que comprende la malla, se seleccionaron dos puntos geográficos representativos: (-35° 75' S, -71°50' W) y (-33° S, -70°50' W). Estos puntos fueron identificados como el representante de clusters

obtenidos a partir de un análisis de agrupamiento basado en series temporales de precipitación, utilizando la métrica Dynamic Time Warping (DTW). Este método, es descrito en un estudio interno universitario, segmentó las series de precipitación de múltiples ubicaciones en clusters de comportamiento similar. Los puntos seleccionados son denominados representantes porque resumen el comportamiento promedio del cluster al que pertenecen, lo que significa que las series temporales de precipitación de los puntos cercanos a ellos muestran dinámicas muy similares. Esta representatividad es clave para simplificar el análisis y reducir la dimensionalidad del problema, sin perder de vista los patrones climáticos relevantes dentro de la región de estudio. La selección de los puntos en cuestión permitió realizar un análisis más detallado de las predicciones, evaluando cómo el modelo captura las dinámicas temporales específicas en un contexto controlado, además de servir como una referencia para validar los resultados obtenidos en toda la región.

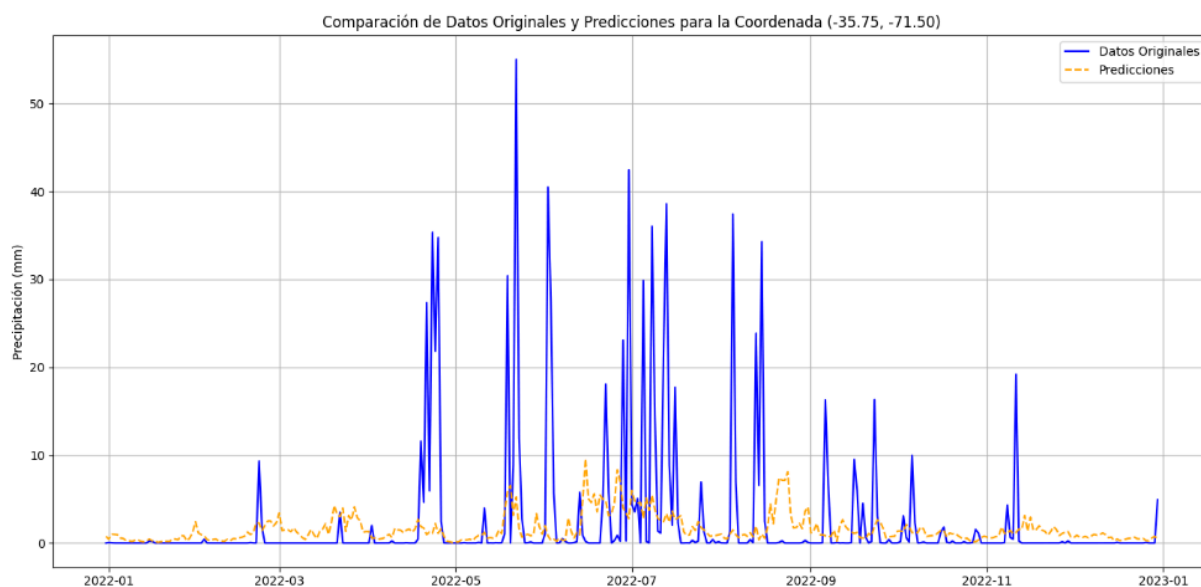
Para este punto representativo, además de las predicciones generales, se calcularon métricas específicas, incluyendo:

- Error absoluto promedio (MAE).
- Raíz del error cuadrático (RMSE).

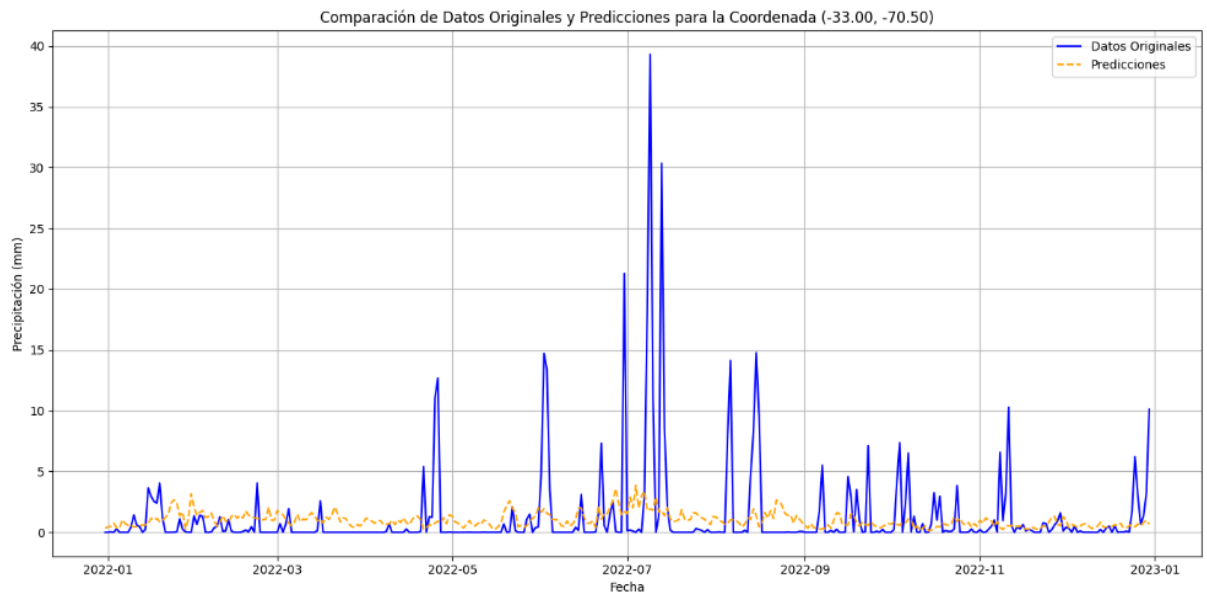
Estas métricas fueron analizadas en conjunto con los valores obtenidos para toda la malla, proporcionando una visión detallada del desempeño del baseline.

El diseño del baseline tiene como objetivo proporcionar un marco de referencia sólido para comparar el desempeño del modelo AE + DMD en etapas posteriores. La reconstrucción completa del año 2022 permitió evaluar las predicciones del modelo en toda la malla, destacando su desempeño tanto a nivel regional como en puntos clave.

La selección de los puntos representantes, extraídos de esta reconstrucción, refuerza el análisis al ofrecer un enfoque detallado sobre puntos estratégicos previamente identificados como representativos del comportamiento climático. Este enfoque no sólo valida la calidad de las predicciones generales, sino que también establece una base sólida para evaluar las ventajas y limitaciones de las metodologías avanzadas introducidas en este trabajo.



Predicción baseline punto representante (-35° 75' S, -71°50' W) - Elaboración propia.



Predicción baseline punto representante (-33° S, $-70^{\circ}50'$ W) - Elaboración propia.

Construcción del AE + DMD

Para la construcción combinada de estos métodos que permiten capturar la complejidad temporal y espacial de los datos de precipitaciones se siguieron una secuencias de pasos, los cuales involucran una mayor cantidad de años de datos (10, 20, 30 y 40 años) y de estaciones de medición (663 estaciones) que el baseline y continuas iteraciones para determinar empíricamente las mejores combinaciones posibles de los resultados.

Preprocesamiento AutoEncoder

Inicialmente se extrajeron los datos en csv que contienen datos desde 1982 hasta 2022, datos que se utilizaron como train y test, dejando el último año (2022) como test y los demás para train.

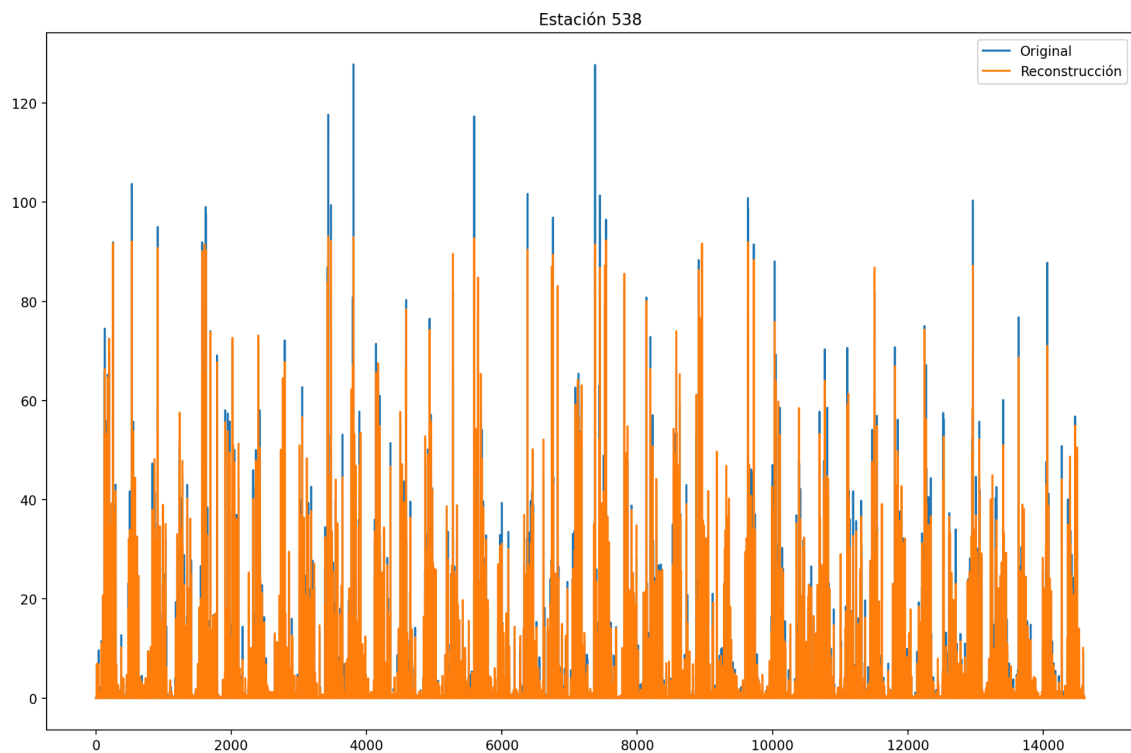
Se usó la normalización de los datos usando la librería de Scikit-learn, buscando mejorar la estabilidad y eficiencia del entrenamiento del modelo.

Aplicación AutoEncoder

Para la construcción del AE se utilizó la librería de TensorFlow Keras, usando una capa de dependencias temporales con una red de memoria de largo plazo (LSTM) de 64 unidades y una capa densa que reduce la dimensionalidad a un espacio latente específico permitiendo una representación compacta de las series temporales.

En el entrenamiento del modelo se consideraron 50 epoch y 16 batch size permitiendo al modelo por medio del optimizador Adam y la función de pérdida de Error Cuadrático Medio (MSE) un adecuado valor de discrepancia entre las series originales y las reconstruidas. Durante el entrenamiento se monitorea la pérdida para evaluar la convergencia del modelo y detectar un posible sobreajuste.

Una vez finalizado el AE, se utilizó para reconstruir las series temporales del conjunto de train, se hizo la desnormalización de los datos reconstruidos para revertir la escala aplicada y se compraron visualmente las series originales con las reconstruidas verificando que los patrones de precipitación hayan sido capturados.



Reconstrucción de la señal de la estación del punto representante - Elaboración propia

Aplicación de la descomposición modal dinámica (DMD)

El siguiente paso fue aplicar el DMD directamente sobre el espacio latente obtenido del AE, en lugar de hacerlo sobre las series originales. Esto permitió aprovechar la reducción de dimensionalidad para un análisis más eficiente.

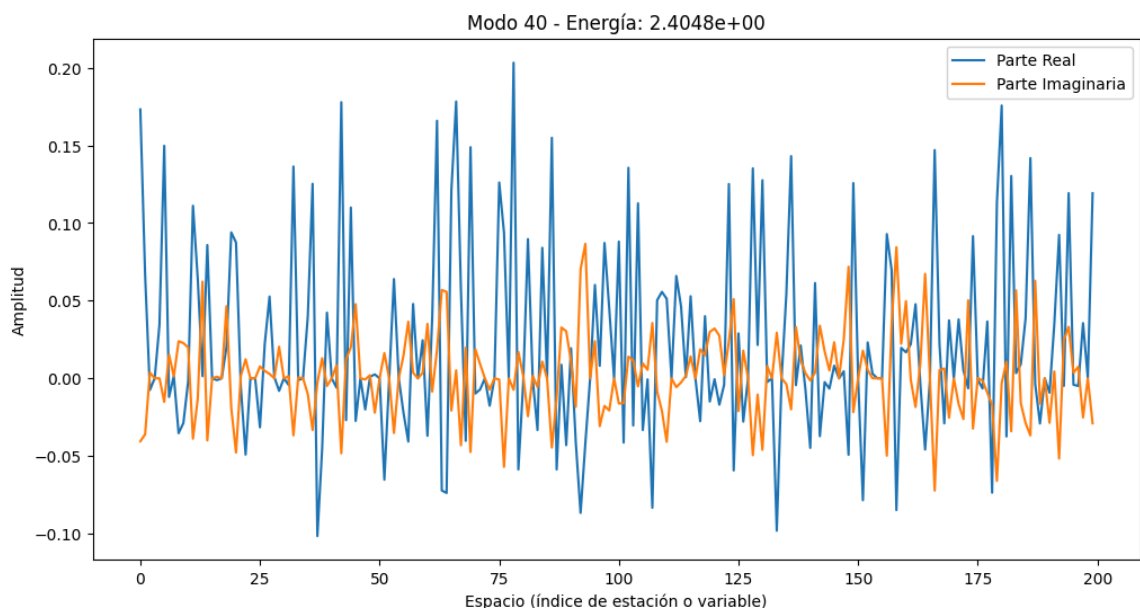
- **Proyección al espacio latente:** Las series temporales normalizadas se codificaron usando el modelo encoder del AE, transformándose a un espacio latente de menor dimensión.
- **Ajuste del modelo DMD:** Se utilizó la biblioteca Py DMD para descomponer las series latentes en modos dinámicos, identificando patrones espaciales y temporales recurrentes. Para esto se calculó un rango reducido ($\text{svd_rank} = 100$) para optimizar la resolución modal y se ajustó el modelo DMD para capturar dinámicas dominantes.
- **Predicción en el espacio latente:** A partir de las últimas ventanas temporales, se realizaron predicciones extendidas (365 días hacia adelante), agregando un pequeño ruido para simular incertidumbre.
- **Reconstrucción de las predicciones:** Las predicciones en el espacio latente se decodificaron al espacio original usando el decodificador del AE. Luego, las series reconstruidas fueron desnormalizadas para interpretar los resultados en su escala original.

Interpretación del DMD

Tras aplicar DMD a las series temporales, se obtienen un conjunto de modos espaciales y sus correspondientes eigenvalores. Cada modo DMD tiene una amplitud asociada, que surge de la proyección inicial de los datos sobre ese modo, la cual mide qué tan

“presente” está un modo en la reconstrucción de los datos originales, es decir, a mayor amplitud contribuye significativamente más a explicar la variabilidad inicial de la señal. A partir de la amplitud se calcula la energía que tiene el modo. Considerando que esta amplitud puede ser un número complejo el valor será el cuadrado de los componentes que luego son sumados. Los valores energéticos de cada modo, son ordenados para verificar los de mayor importancia.

Por otro lado, está la frecuencia de oscilaciones del modo que es representada por la parte imaginaria de los eigenvalores, mientras mayor es esta indica ciclos u oscilaciones más rápidas, lo cual es útil en el posterior análisis debido a que puede corresponder a patrones estacionales o fenómenos periódicos.

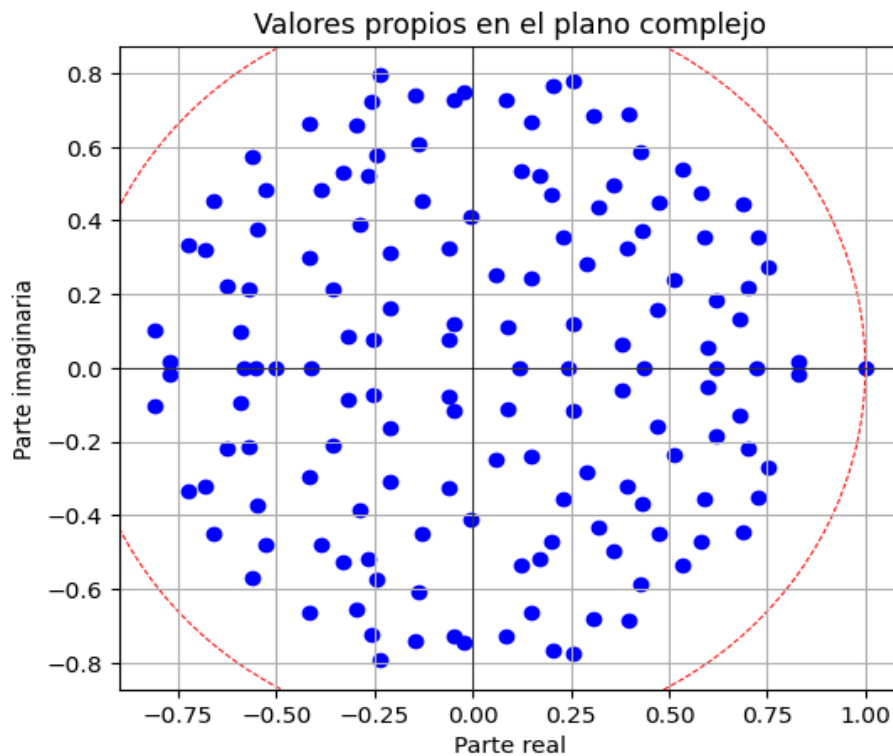


Descomposición de los modos generados - Elaboración propia.

Valores propios en el plano complejo

El análisis de los valores propios en el plano complejo permite evaluar la estabilidad y la dinámica de los modos DMD. Cada valor propio se representa como un punto en un plano donde el eje real indica la tasa de crecimiento o amortiguamiento del modo, mientras que el eje imaginario está relacionado con la frecuencia de oscilación de dicho modo.

En la gráfica siguiente, los valores propios están mayoritariamente distribuidos dentro del círculo unitario (radio = 1), lo que indica que los modos asociados son estables y no presentan crecimiento exponencial. Este comportamiento es fundamental en el análisis de series temporales, ya que asegura que las predicciones y reconstrucciones permanecen acotadas dentro de rangos razonables.



Valores propios en el plano complejo - Elaboración propia

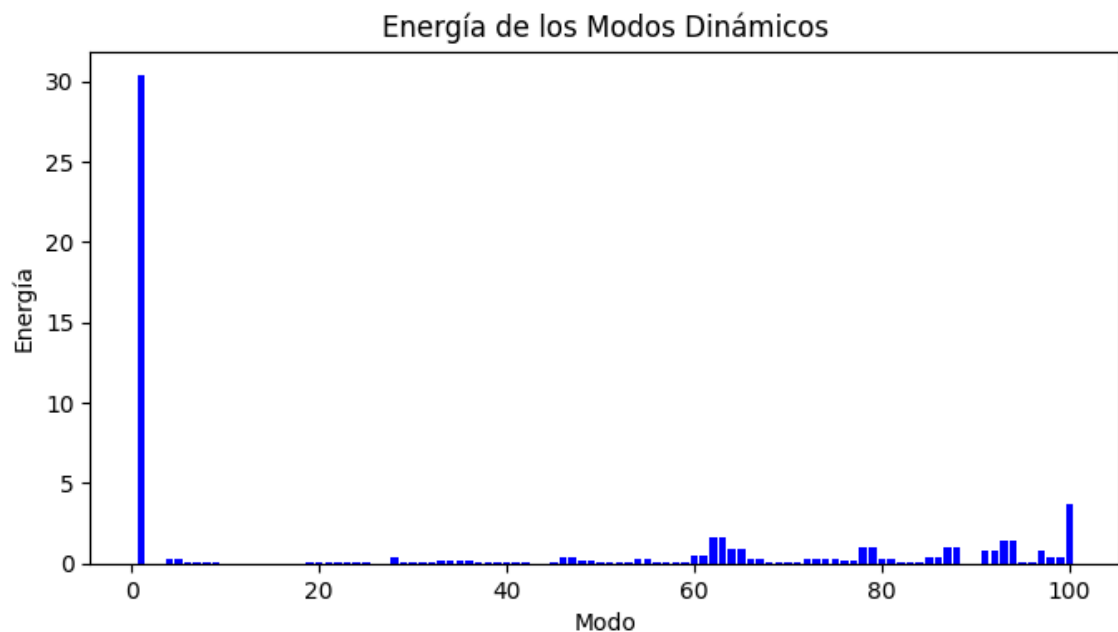
Los valores propios cercanos al borde del círculo sugieren dinámicas con menor amortiguamiento, mientras que los más alejados del borde indican modos altamente atenuados. Aquellos ubicados cerca del eje real reflejan componentes de baja frecuencia o estacionarias, mientras que los valores con una parte imaginaria significativa corresponden a oscilaciones periódicas de mayor frecuencia.

Este análisis es clave para identificar las dinámicas dominantes de la señal y descartar modos irrelevantes o inestables que podrían introducir ruido o errores en las predicciones. Además, la distribución uniforme de los valores propios indica una

diversidad de frecuencias capturadas, lo que complementa el análisis espectral del sistema.

Evaluación energética de los modos DMD

La evaluación energética de los modos dinámicos es fundamental para identificar cuáles contribuyen de manera significativa a la reconstrucción y predicción de las series temporales. La energía de cada modo se calcula como el cuadrado de la amplitud normalizada respecto al total de amplitudes, lo que permite determinar la importancia relativa de cada uno en la descomposición modal.



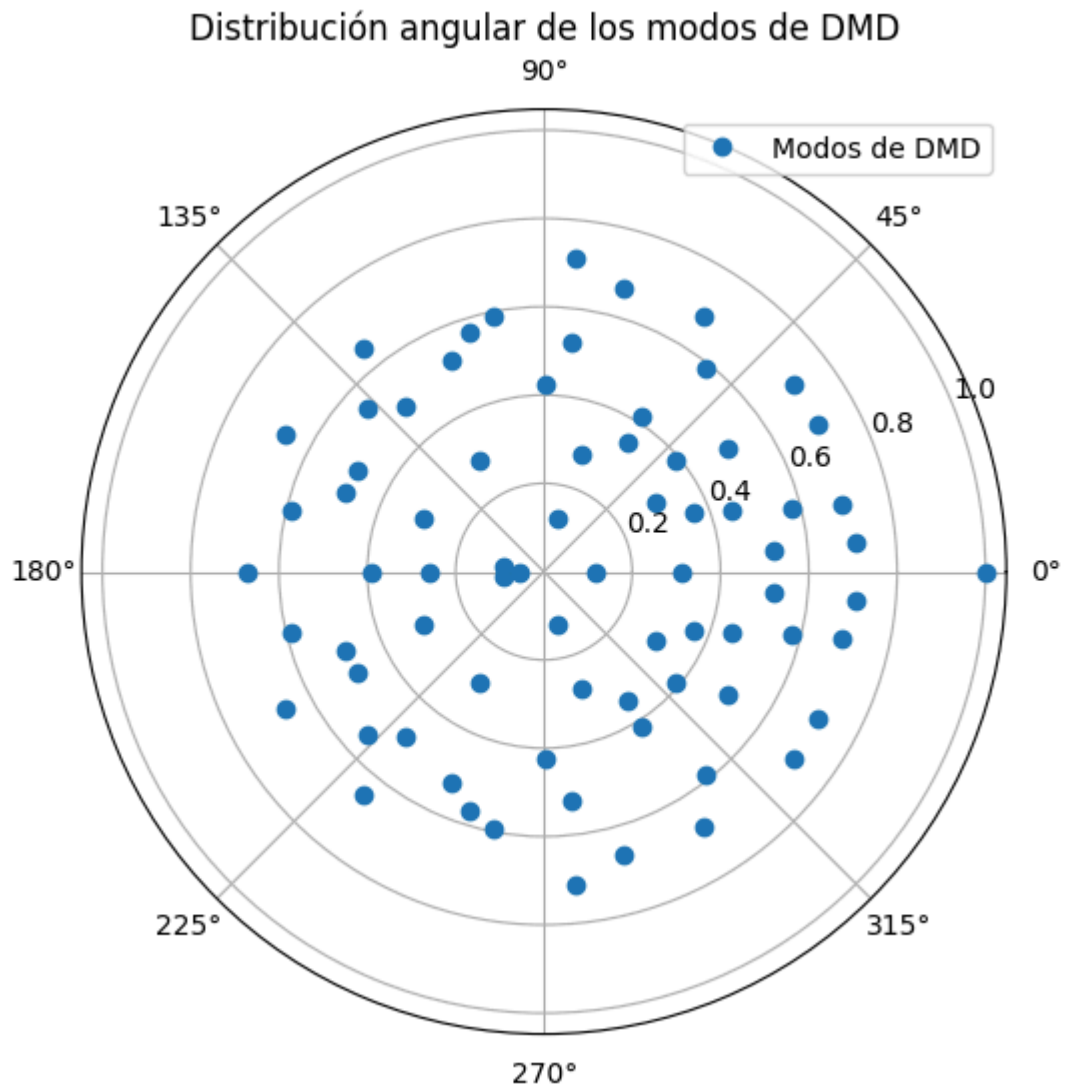
Energía de los modos dinámicos - Elaboración propia.

En el histograma presentado, se observa que los primeros modos concentran la mayor parte de la energía, confirmando su relevancia en la representación de las dinámicas

principales del sistema. Este comportamiento es esperable, ya que los modos de mayor energía suelen capturar patrones dominantes o recurrentes en los datos, mientras que los modos con energía significativamente menor tienden a representar ruido o componentes de baja relevancia para el modelo.

El análisis de la distribución energética también es útil para determinar un umbral de corte, permitiendo seleccionar únicamente los modos más significativos y descartar aquellos que no aportan valor añadido. En este caso, se evidencia que el modo más energético destaca de manera notable sobre el resto, mientras que otros modos con energías menores, aunque no despreciables, podrían ser útiles para capturar variaciones secundarias en las series temporales.

Distribución angular y frecuencia de los modos DMD



Distribución angular de los modos de DMD - Elaboración propia.

Para analizar la distribución angular de los modos DMD, se calculan los ángulos asociados a los eigenvalores, los cuales representan la dirección de las oscilaciones en el

plano complejo. La magnitud del eigenvalor indica la estabilidad o crecimiento del modo, mientras que el ángulo refleja la frecuencia relativa de oscilación.

En la gráfica polar se observa que los modos están distribuidos de manera uniforme a lo largo del plano angular, lo que sugiere la presencia de dinámicas recurrentes en diferentes escalas temporales. Modos con ángulos cercanos a 0° o 360° están más alineados con componentes estacionarios o de baja frecuencia, mientras que aquellos en ángulos intermedios (90° y 270°) están asociados a oscilaciones más rápidas y patrones periódicos más intensos.

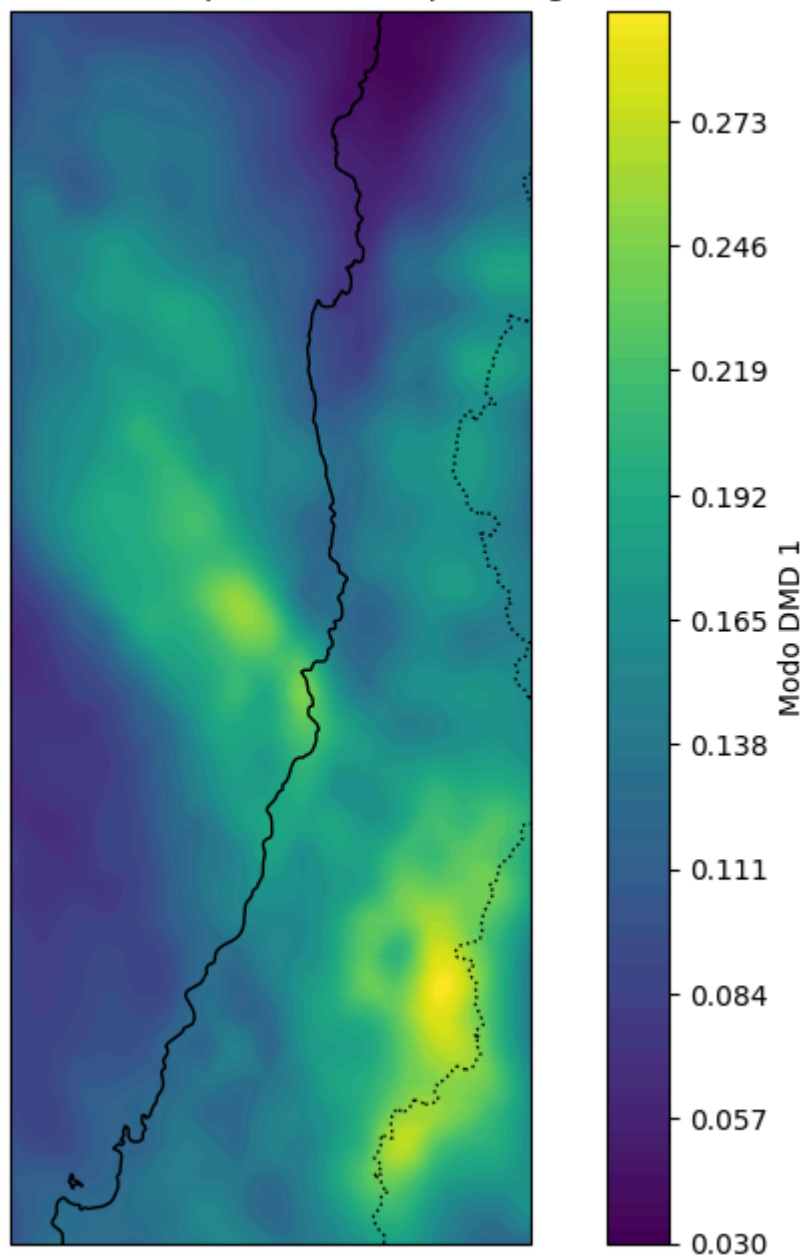
Esta representación angular facilita la identificación de patrones dominantes en la señal original y la segmentación de dinámicas según su contribución específica. Además, los valores en el eje radial de la gráfica reflejan la magnitud de los eigenvalores, permitiendo identificar qué modos tienen una influencia mayor o menor en la señal reconstruida. Este análisis complementa la evaluación energética, proporcionando una visión más detallada de las frecuencias y contribuciones de cada modo en el comportamiento dinámico de los datos.

Retorno al espacio físico

Se crea una malla definiendo valores mínimos y máximos de latitud y longitud, de tal forma que las estaciones, por medio de la interpolación, representan el valor del modo reconstruido en cada punto de esta malla. Para lograr la reconstrucción al espacio físico

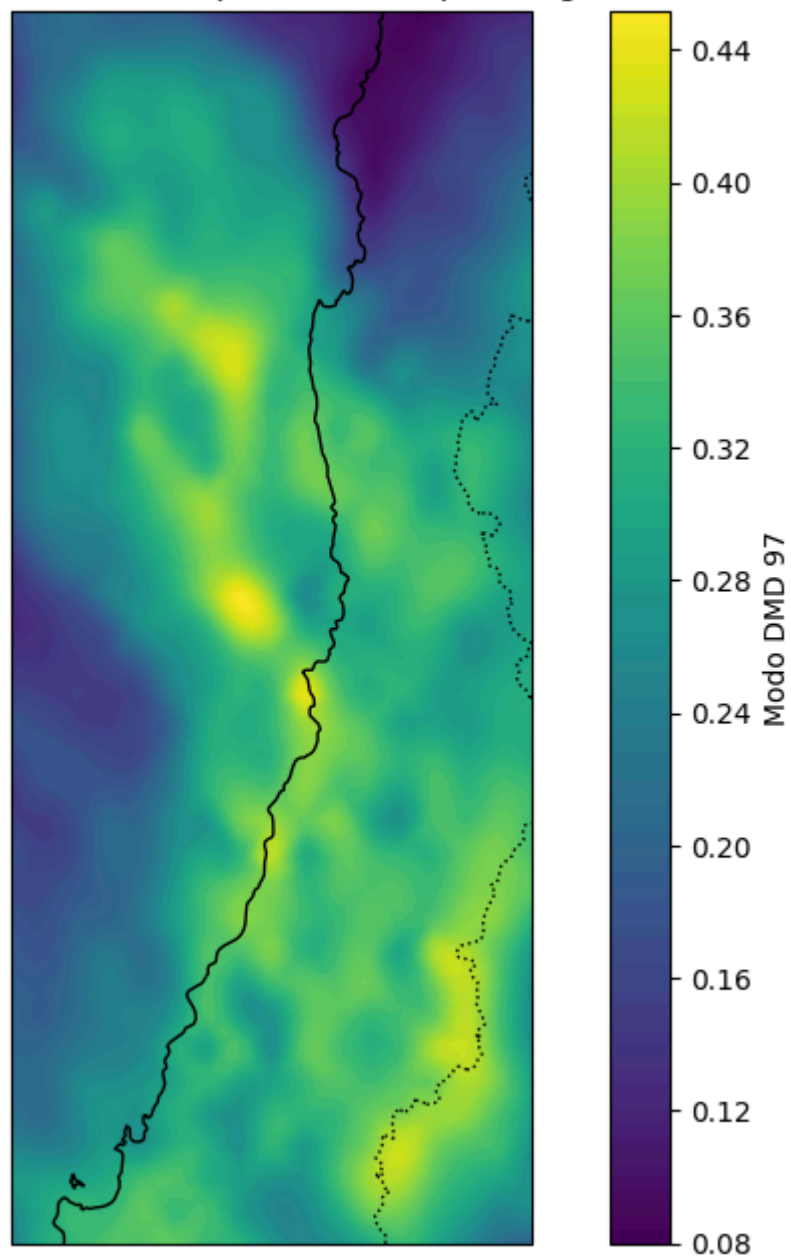
se selecciona un modo para obtener la representación latente, el cual es un vector que describe la estructura espacial del modo en esta representación comprimida del AE. Finalmente para la interpolación, este vector es descomprimido devolviendo una representación de las magnitudes del modo en el espacio original y obteniendo la precipitación en la red de estaciones.

Modo DMD 1 - Interpolado en Mapa Geográfico



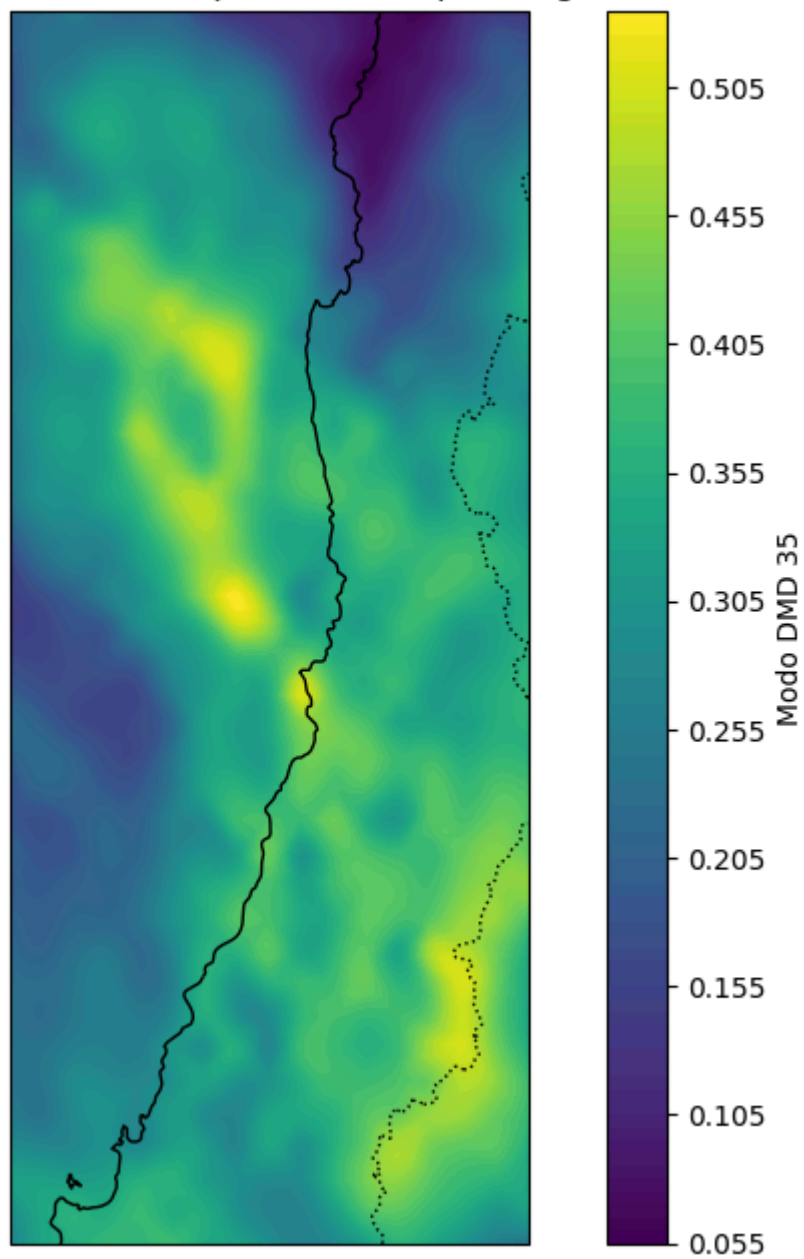
Modo DMD 1 con mayor energía - Elaboración propia.

Modo DMD 97 - Interpolado en Mapa Geográfico



Modo DMD 97 5to con mayor energía - Elaboración propia.

Modo DMD 35 - Interpolado en Mapa Geográfico



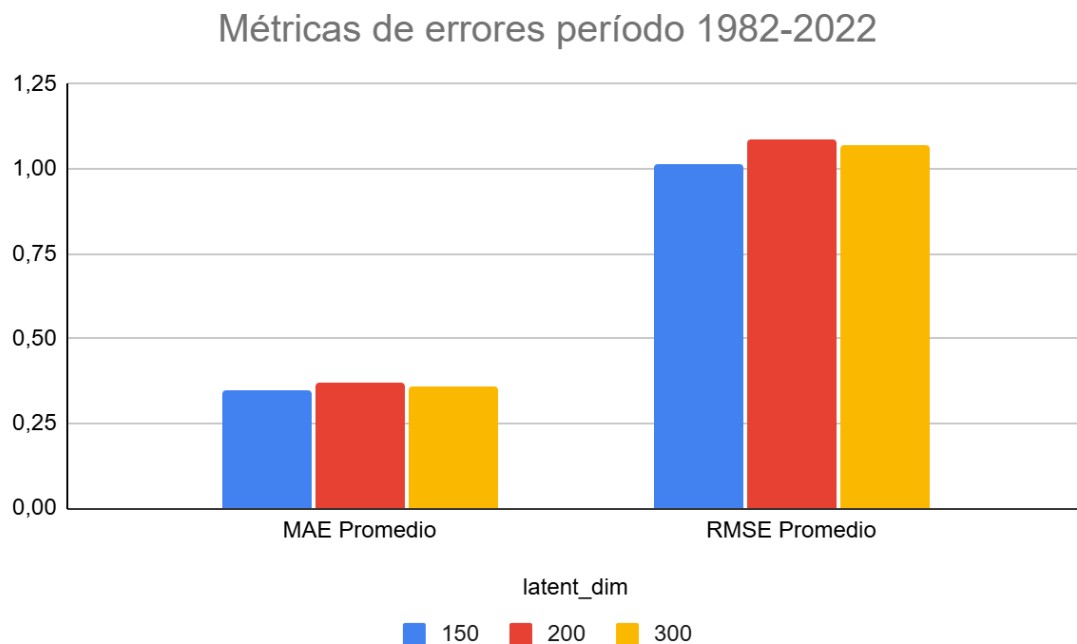
Modo DMD 35 10mo con mayor energía - Elaboración propia.

5. Resultados

Sensibilización de la dimensión latente

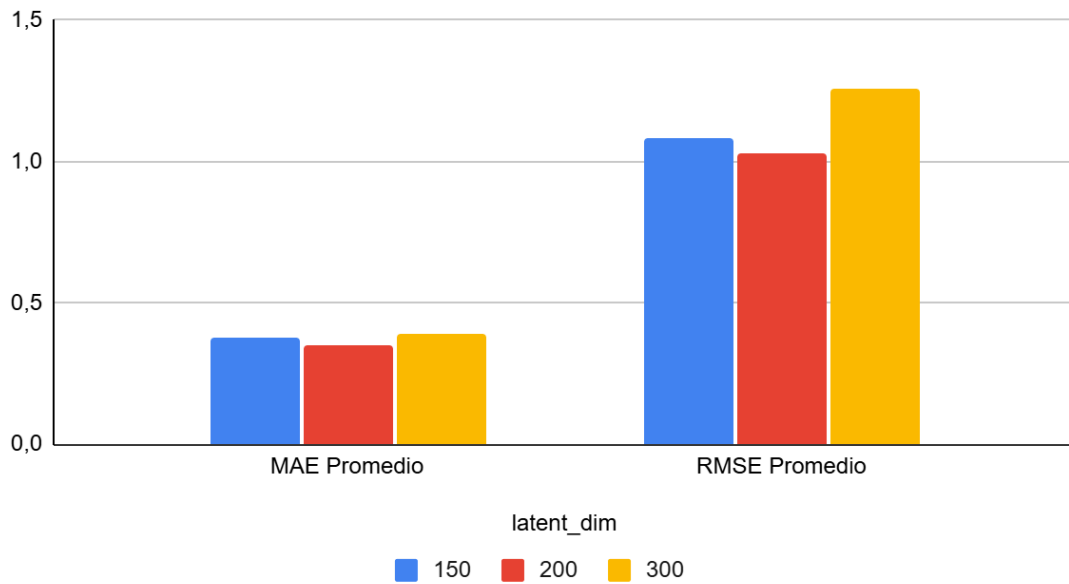
Se realizó la variación de la dimensión latente entre 150, 200 y 300, para cada intervalo de años estudiado los cuales van desde 1982, 1992, 2002 y 2012 hasta el 2022, de tal forma que permita revisar cómo afecta la cantidad de años y la variación de la dimensión latente a los resultados del modelo.

A continuación se presentan primeramente los gráficos con los valores de los errores por período de años mostrando sus distintas dimensiones latentes y en segundo lugar manteniendo las dimensiones latentes constantes variando el período de años mostrando las variaciones producidas en las métricas de errores.



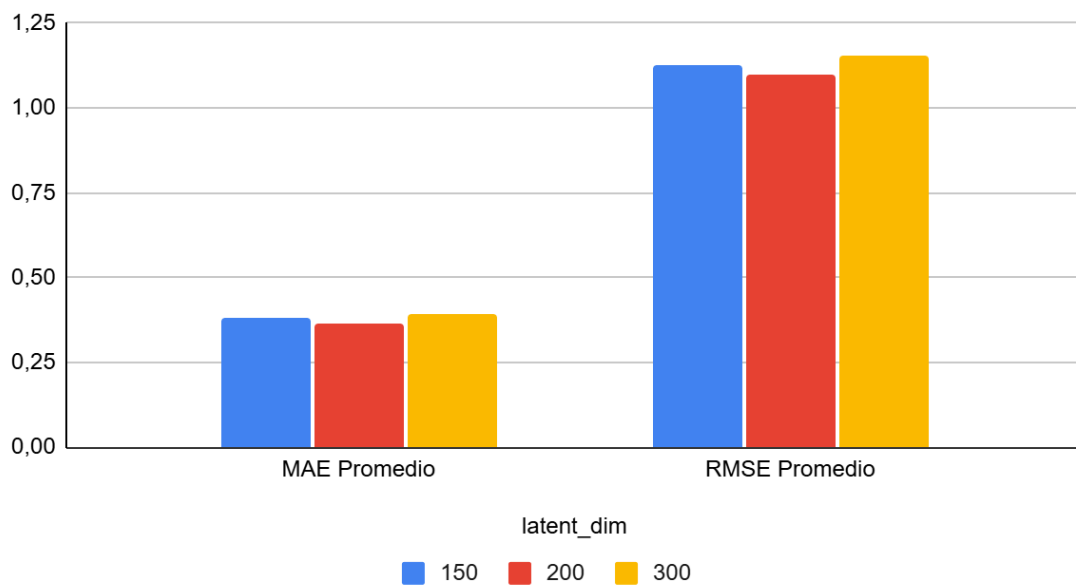
Métricas de error período 1982-2022 - Elaboración propia.

Métricas de errores período 1992-2022

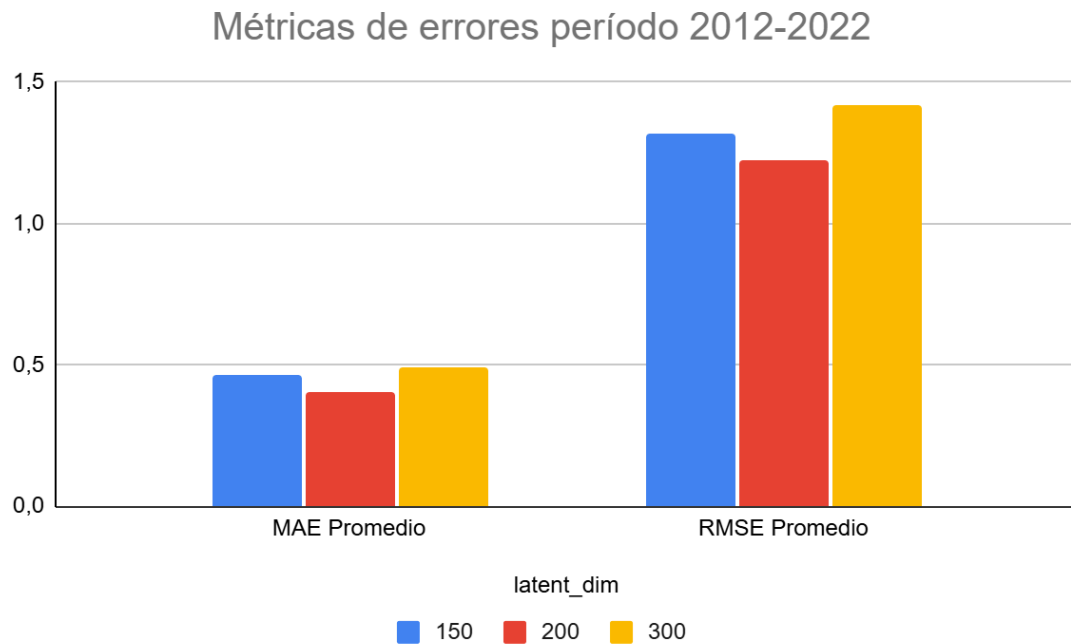


Métricas de error período 1992-2022 - Elaboración propia.

Métricas de errores período 2002-2022



Métricas de error período 2002-2022 - Elaboración propia.

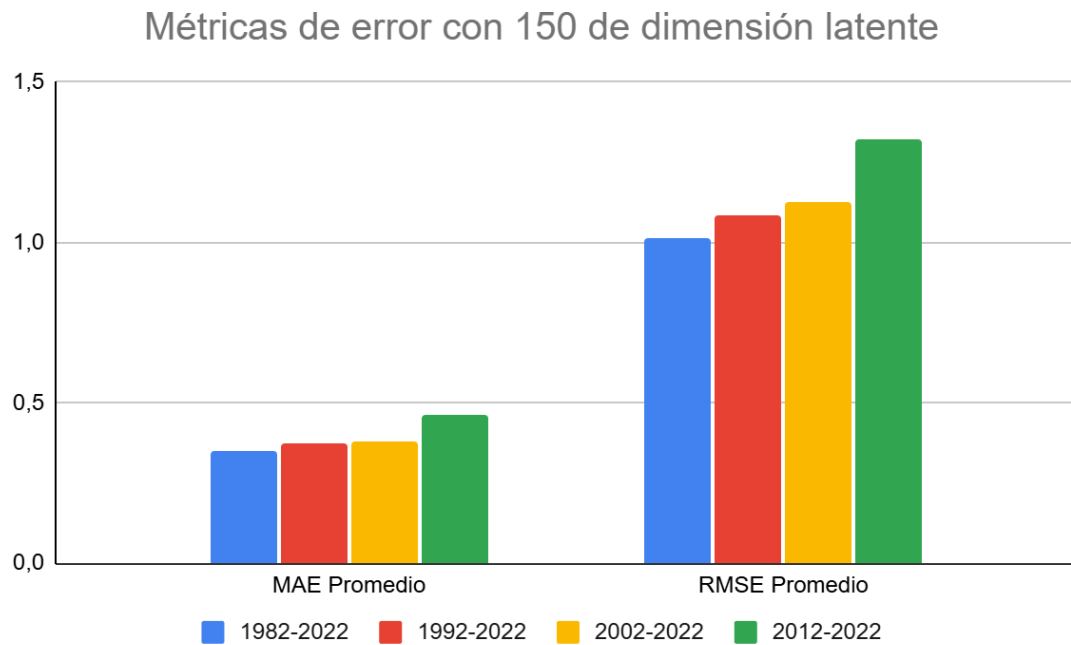


Métricas de error período 2012-2022 - Elaboración propia.

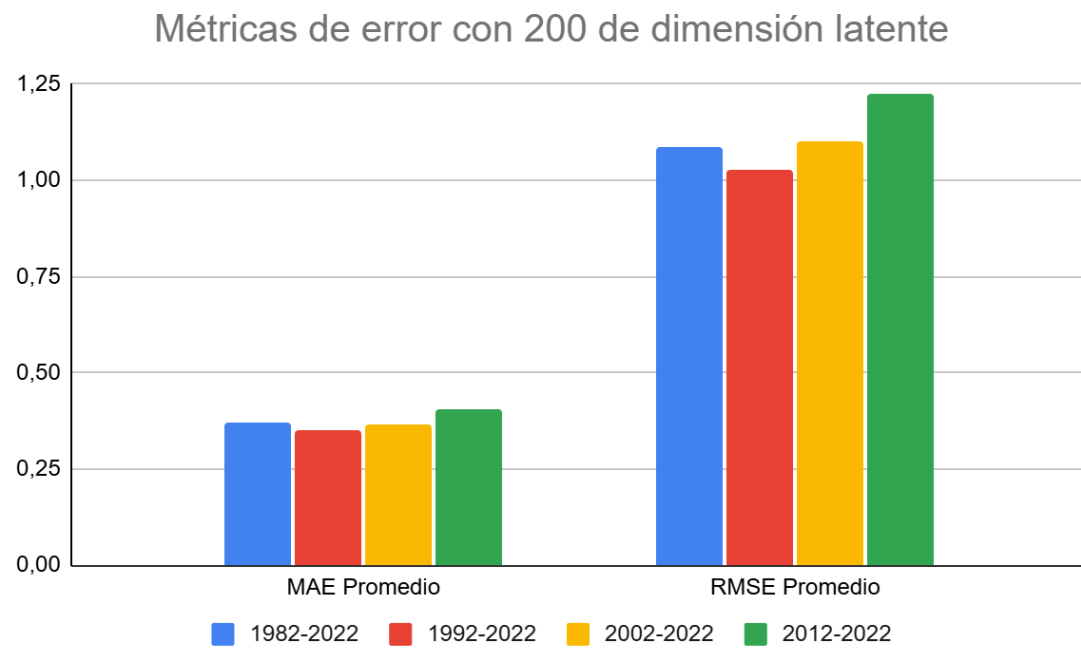
Los principales hallazgos de la información expuesta son:

- A medida que se disminuye la cantidad de años para el entrenamiento el modelo aumentaba su promedio de errores tanto en MAE como en RMSE.
- El valor de MAE más bajo lo obtuvo la combinación de 1982-2022 con 150 de dimensión latente, con un 0,3485 seguido del período 1992-2022 con 200 de dimensión latente, con un 0,349.
- De igual forma ambas combinatorias también obtuvieron en igual orden los valores más bajos de RMSE con 1,0157 y 1,0294 respectivamente.

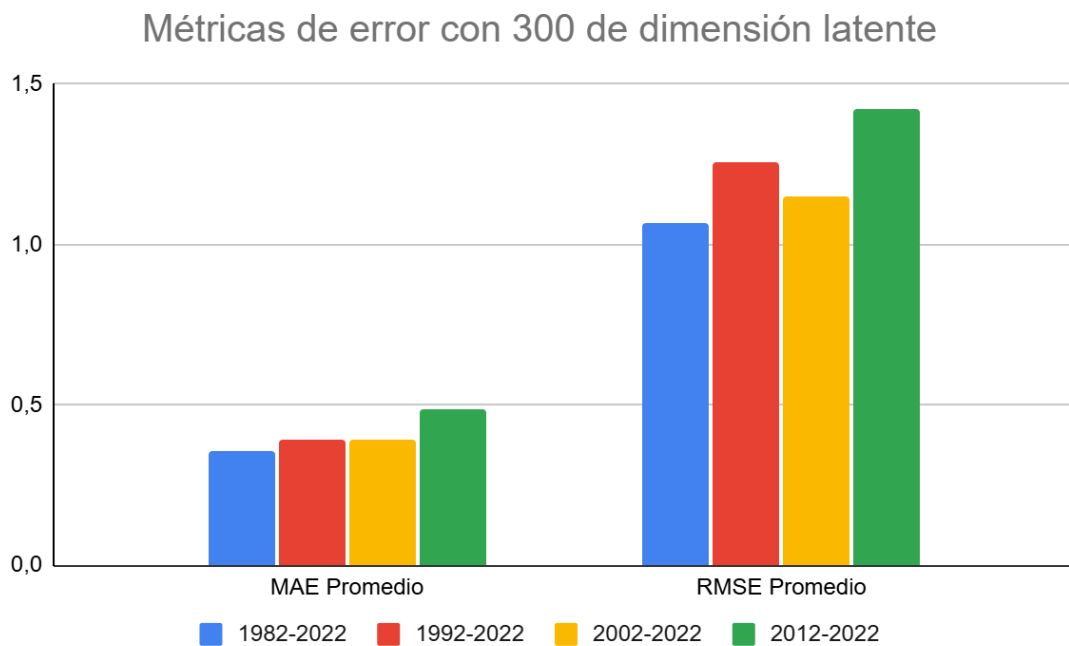
Continuando con los gráficos como se mencionó anteriormente, se muestran los valores de las métricas de error con la dimensión latente constante y variando los períodos de años utilizados:



Métricas de error con 150 de dimensión latente - Elaboración propia.



Métricas de error con 200 de dimensión latente - Elaboración propia.



Métricas de error con 300 de dimensión latente - Elaboración propia.

Como se puede visualizar en las gráficas anteriores, es posible concluir que mientras más años de entrenamiento tiene el modelo mejor es su desempeño respecto a las métricas de error. Consistentemente, el período de 10 años tiene un mayor error que los demás, en contraste con el modelo entrenado con 40 años que tiene menores niveles de errores al momento de hacer las predicciones, no así con los niveles de dimensión latente, ya que estos tuvieron niveles de errores menores cuando se usaron 150 y 200.

Para los siguientes análisis se utilizará la combinación con la menor métrica de error que fue utilizando el período de 1982-2022 y con 150 de dimensión latente.

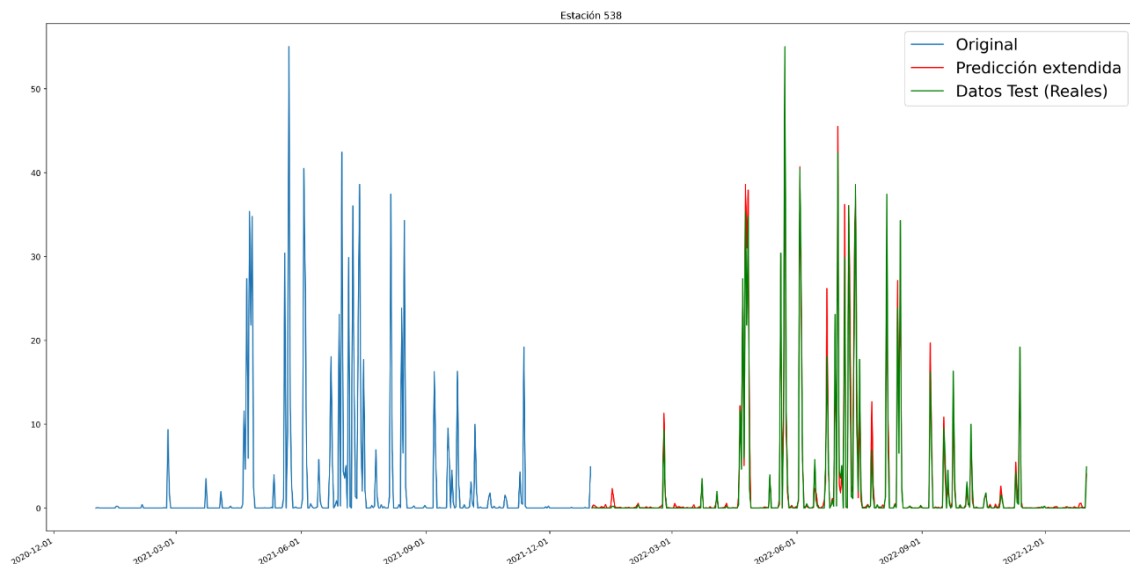
Puntos representantes

Los resultados de los puntos representantes respecto a sus métricas de error contrastando lo que es el modelo del baseline con el modelo AE y DMD son:

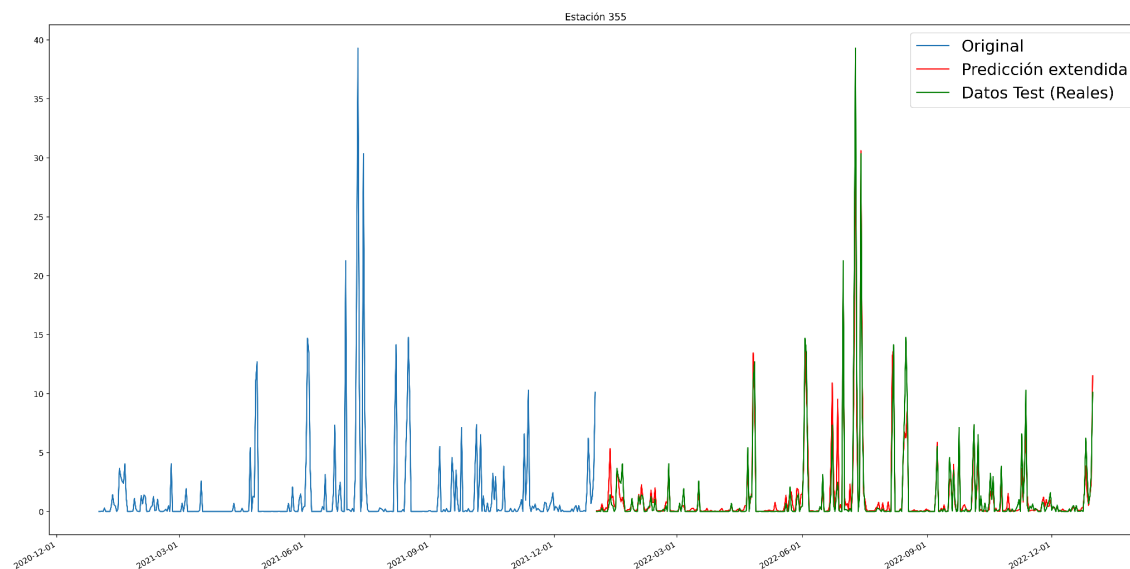
Punto representante (-35° 75' S, -71°50' W)			
Modelo	MAE Punto representante	RMSE Punto representante	Tiempo ejecución
Baseline	2,5561	6,2405	1416 segundos
AE + DMD	0,5211	1,4898	423 segundos

Punto representante (-33° S, -70°50' W)			
Modelo	MAE Punto representante	RMSE Punto representante	Tiempo ejecución
Baseline	1,6833	3,7219	1416 segundos
AE + DMD	0,4727	1,1270	423 segundos

Los valores obtenidos con el modelo desarrollado mejoran al baseline propuesto respecto a sus métricas de error al momento de hacer las evaluaciones de predicción. Por otro lado, se presenta el gráfico que muestra el valor predicho de precipitaciones con los datos reales:

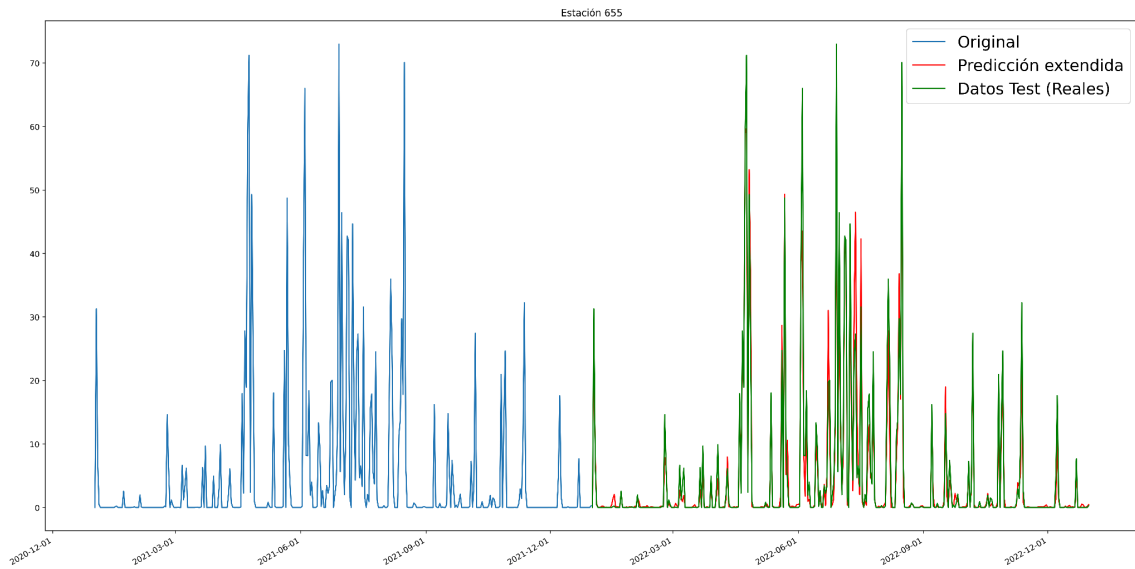


Predicción del punto representante (-35° 75' S, -71°50' W), ilustrando 2 años a modo de ejemplificación -
Elaboración propia.

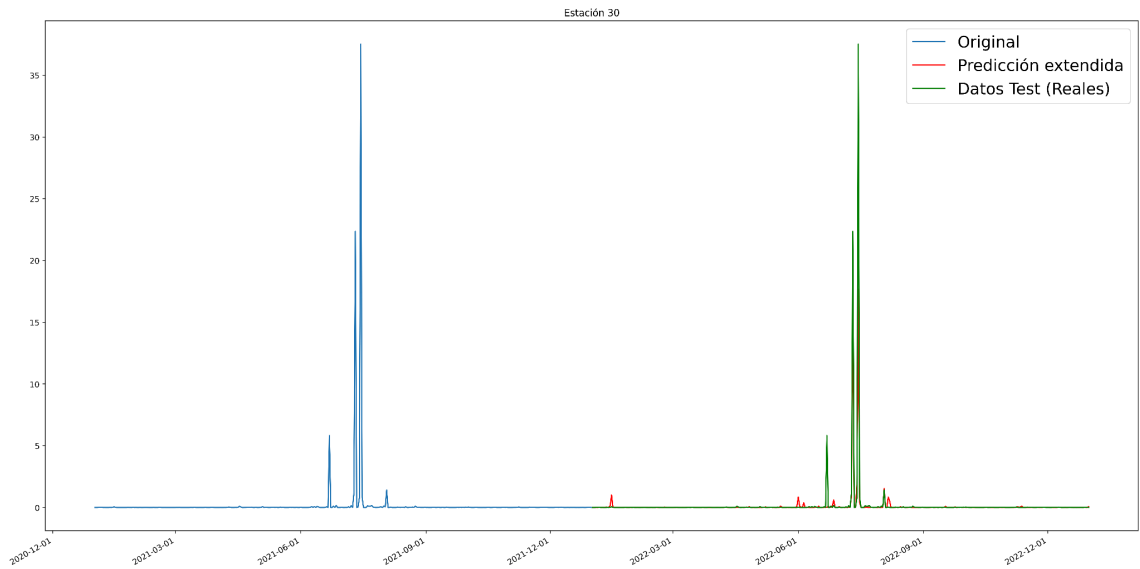


Predicción del punto representante (-33° 00' S, -70°50' W), ilustrando 2 años a modo de ejemplificación -
Elaboración propia.

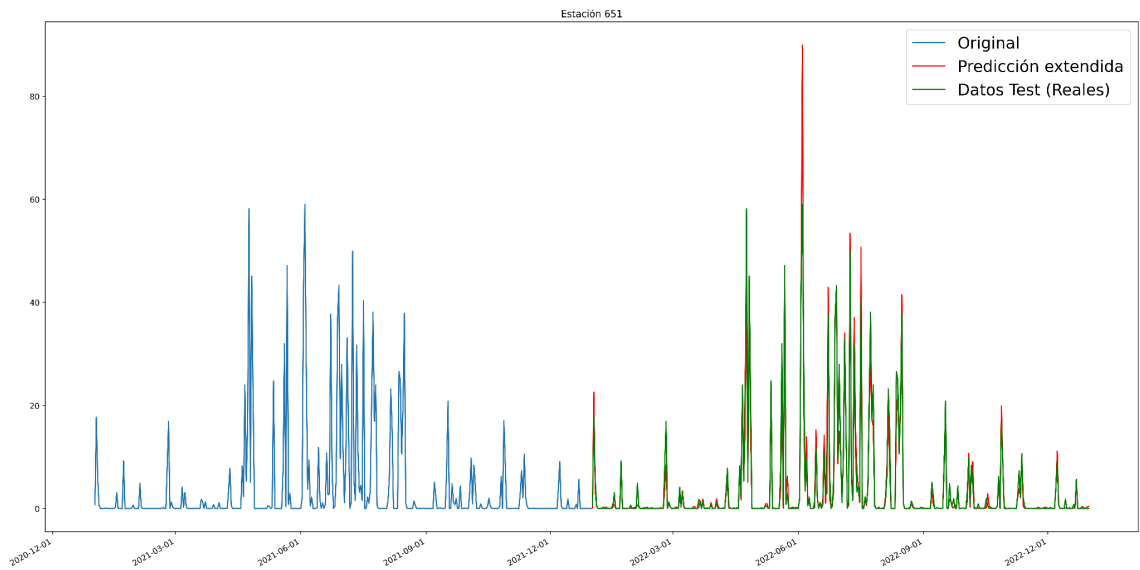
Se mostrarán a continuación otros puntos a modo de ejemplo para tener una visión más general de los resultados obtenidos, principalmente puntos que tienen errores de MAE o RMSE más altos.



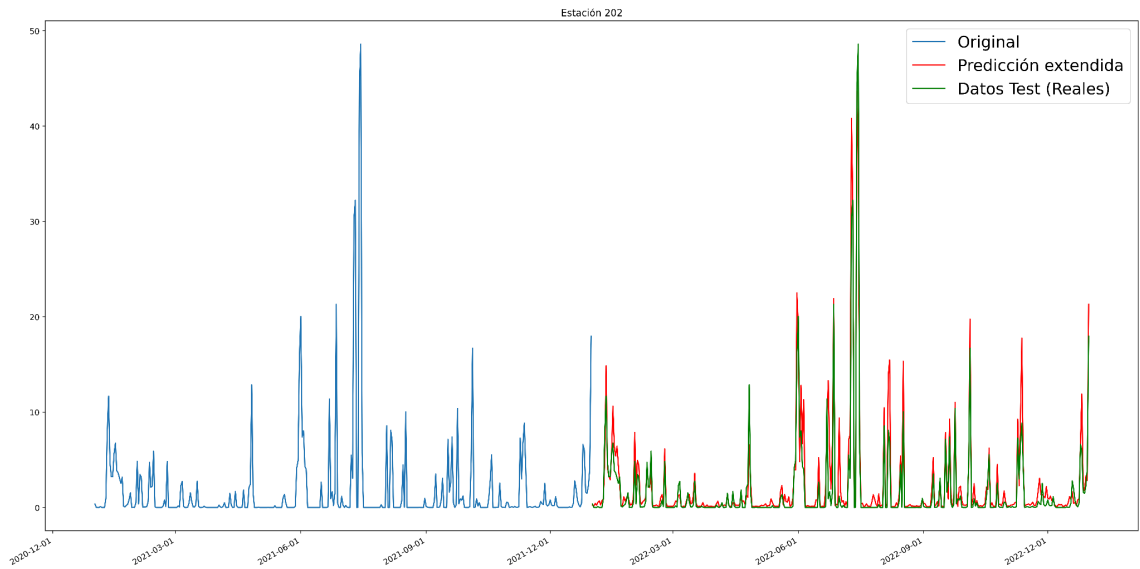
Predicción del punto (-37° 50' S, -72°00' W), MAE = 1.1669 / RMSE = 2.8621 - Elaboración propia.



Predicción del punto (-28° 25' S, -71°00' W), MAE = 0,1176 / RMSE = 1,1326 - Elaboración propia.

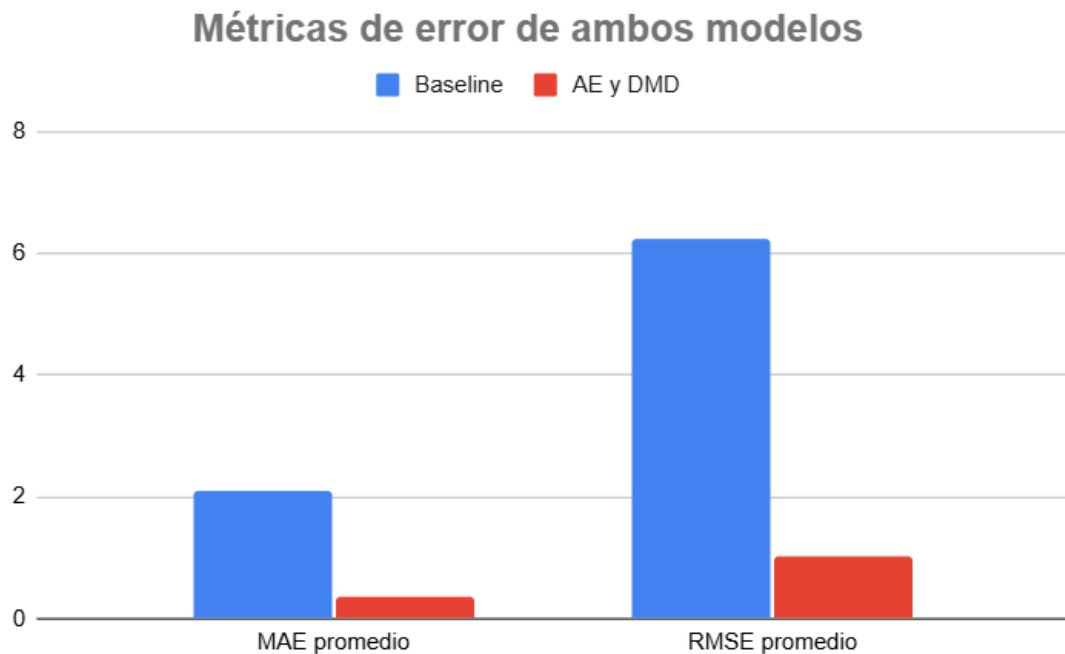


Predicción del punto (-37° 50' S, -73°00' W), MAE = 0,9880 / RMSE = 2,7608 - Elaboración propia.



Predicción del punto (-37° 50' S, -73°00' W), MAE = 0,8573 / RMSE = 1,8287 - Elaboración propia.

Considerando los valores finales de las métricas de error de ambos modelos quedan graficados de la siguiente forma:



Métricas de error ambos modelos - Elaboración propia.

Obteniendo el baseline un MAE de 2,1093 y RMSE de 6,2405 y el modelo AE y DMD tiene el MAE de 0,3485 y RMSE de 1,0157 mejorando ampliamente los resultados iniciales. Por otro lado, los tiempos computacionales obtenidos para el entrenamiento del baseline es de 1416 segundos mientras que el otro modelo es de tan solo 423 segundos, es importante notar que para el baseline solo se usaron 5 años teniendo un tiempo mucho mayor de cómputos con peores métricas de error, mientras que el modelo de AE y DMD uso 40 años de entrenamiento y tiene una gran optimización en el uso de tiempo para su entrenamiento.

Modelo	MAE Promedio	RMSE Promedio	Tiempo ejecución
Baseline	2,1093	6,2405	1416 segundos
AE + DMD	0,3485	1,0157	423 segundos

6. Conclusiones

El presente trabajo aborda el desafío de predecir precipitaciones en diversas coordenadas dentro de Chile utilizando un enfoque novedoso basado en la combinación de autoencoders (AE) y descomposición modal dinámica (DMD). A partir del análisis de los resultados obtenidos, se concluye lo siguiente:

Reducción dimensional efectiva con autoencoders

El uso de autoencoders permitió reducir la dimensionalidad de los datos originales, preservando las características más relevantes para el modelo predictivo. En particular, al trabajar con un espacio latente de 150 dimensiones, se logró un equilibrio óptimo entre precisión y eficiencia computacional, seleccionando este espacio tras una sensibilización que incluyó distintos tamaños del espacio latente. Este enfoque demostró ser una alternativa eficiente para manejar conjuntos de datos climáticos de alta dimensionalidad.

Comparación con el modelo baseline

En términos de precisión, el modelo AE + DMD superó al baseline en las métricas de rendimiento evaluadas. El RMSE se redujo en 5 puntos, lo que indica que el modelo combinado captura de manera más efectiva las dinámicas temporales y espaciales de las precipitaciones. Además, al realizar las comparaciones en los gráficos mostrando los valores de precipitaciones predichos con los datos reales para el año 2022, se puede ver

que el modelo AE + DMD muestra una mejora significativa, destacando la capacidad del modelo para representar patrones climáticos complejos.

Respecto al tiempo de ejecución, el AE + DMD demostró una notable ventaja. Al operar sobre un espacio latente comprimido, el tiempo requerido para entrenar y realizar predicciones fue de 993 segundos menor en comparación con el baseline, que trabajaba sobre la totalidad de los datos originales. Esto posiciona al AE + DMD como una solución viable para aplicaciones en tiempo real o con restricciones computacionales.

Análisis de los modos dinámicos

La descomposición modal dinámica permitió identificar modos dominantes que representan estructuras temporales clave del sistema. La distribución de los eigenvalores en el plano complejo evidenció la existencia de modos con características oscilatorias y estacionarias, lo que refuerza la capacidad del DMD para capturar fenómenos periódicos y de largo plazo. Además, la evaluación de las frecuencias y energías de los modos destacó la contribución de un conjunto reducido de ellos a la reconstrucción precisa de los datos originales, confirmando la relevancia de la selección del espacio latente.

Contribuciones metodológicas

Este trabajo demuestra la utilidad de combinar técnicas de aprendizaje profundo con métodos dinámicos para abordar problemas complejos de predicción climática. La integración del AE y el DMD no solo ofrece una mejora en la precisión y eficiencia, sino

que también proporciona un marco flexible que puede adaptarse a distintos fenómenos y escalas espaciales.

En conclusión, el modelo AE + DMD constituye una alternativa eficiente y precisa para predecir precipitaciones en regiones específicas de Chile. Su capacidad de manejar datos de alta dimensionalidad, reducir costos computacionales y superar a métodos tradicionales lo posiciona como una herramienta valiosa en el ámbito de las ciencias climáticas y la toma de decisiones relacionadas con fenómenos meteorológicos. Este enfoque no solo valida las hipótesis iniciales del estudio, sino que también abre nuevas oportunidades para aplicaciones prácticas y extensiones metodológicas futuras.

Trabajo futuro

Si bien los resultados obtenidos son prometedores, existen diversas oportunidades para expandir y mejorar este trabajo:

- **Incorporación de variables adicionales:** La inclusión de variables climáticas relevantes, como temperaturas superficiales, niveles de humedad y presión atmosférica, podría enriquecer los modelos y mejorar la precisión de las predicciones.
- **Fenómeno de El Niño y La Niña:** La incorporación de datos relacionados con los eventos de El Niño y La Niña permitiría evaluar el impacto de estos

fenómenos en las precipitaciones, proporcionando una perspectiva más integral y robusta sobre la variabilidad climática.

- **Extensión del modelo a EDMD:** La aplicación de Extended Dynamic Mode Decomposition (EDMD) ofrecería una mayor capacidad para capturar dinámicas no lineales, lo cual sería particularmente relevante en sistemas climáticos complejos.
- **Aumento del número de estaciones meteorológicas:** Ampliar la red de estaciones utilizadas para la recopilación de datos permitiría cubrir zonas que actualmente no se han considerado, extendiendo la evaluación a más zonas de Chile, con sus respectivas particularidades climáticas.
- **Evaluación del impacto de la topografía:** Dado el carácter altamente montañoso de Chile, sería interesante estudiar con mayor profundidad cómo la topografía influye en los patrones de precipitación y su predicción.

Limitaciones

Este estudio enfrenta varias limitaciones que deben considerarse al interpretar los resultados y planificar investigaciones futuras:

- **Variabilidad climática no representada:** Aunque el modelo muestra buenos resultados en términos de predicción, no aborda explícitamente cambios climáticos de largo plazo, como los inducidos por el calentamiento global.

- **Validación en diferentes contextos:** Aunque los resultados son válidos para las coordenadas estudiadas, se requiere mayor validación en otras regiones del país para generalizar los hallazgos.
- **Limitaciones en la generalización del modelo:** El modelo fue diseñado y evaluado exclusivamente para la predicción de precipitaciones en coordenadas específicas de Chile. Aunque los resultados son prometedores, su generalización a otras regiones con diferentes características climáticas o hidrológicas podría requerir ajustes sustanciales en los datos de entrada, la reducción dimensional o la parametrización del DMD.

Bibliografia

1. Akbari Asanjan, A., Yang, T., Hsu, K., Sorooshian, S., Lin, J., & Peng, Q. (2018). Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks. *Journal of Geophysical Research: Atmospheres*, 123(23), 12,543–12,563. <https://doi.org/10.1029/2018JD028375>
2. Bank, D., Koenigstein, N., & Giryas, R. (2021). *Autoencoders*. En arXiv:2003.05991 [cs.LG]. <https://doi.org/10.48550/arXiv.2003.05991>
3. Kutz, J. N., Brunton, S. L., Brunton, B. W., & Proctor, J. L. (2016). *Dynamic mode decomposition: Data-driven modeling of complex systems*. Philadelphia: Society for Industrial and Applied Mathematics.
4. Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., & Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science*, 319(5863), 607–610. <https://doi.org/10.1126/science.1152339>
5. Man, J., Zhang, W., Feng, J., Li, H., & Shao, C. (2023). *W-MAE: Pre-trained weather model with masked autoencoder for multivariable weather forecasting*. arXiv. <https://doi.org/10.48550/arXiv.2304.08754>
6. Salinas, D., Flunkert, V., & Gasthaus, J. (2019). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. arXiv. <https://doi.org/10.48550/arXiv.1704.04110>

7. Zhang, C.-J., Wang, H.-Y., Zeng, J., Ma, L.-M., & Guan, L. (2020).

Tiny-RainNet: A deep convolutional neural network with bi-directional long short-term memory model for short-term rainfall prediction. *Meteorological Applications*, 27, e1956. <https://doi.org/10.1002/met.1956>