

Predicción Espacio-Temporal de datos climáticos en Chile con el uso de EOFs y Deep Learning

POR. LUIS ALEJANDRO ZAVALA ZAVALA
 ANDRÉS E. PÉREZ CARVAJAL

Proyecto Capstone 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, 2023
Santiago

Agradecimientos

Luis Zavala:

Con agradecimiento, dedico estas palabras a mi familia. Su confianza depositada en mis sueños y su incansable aliento en los momentos más desafiantes, han sido el sostén que me permitió alcanzar esta meta. Gracias por ser mi inspiración, por brindarme el espacio y la tranquilidad para crecer y aprender.

También quiero expresar mi sincero reconocimiento a mi compañero Andrés Pérez, con quien compartí este reto. Juntos, hemos demostrado que la colaboración y el compromiso mutuo pueden superar cualquier obstáculo académico, aprendiendo el valor del trabajo en equipo.

Andrés Pérez Carvajal:

Partir agradeciendo a mi familia, a mi amor Marcia, Magdalena y Amanda, nunca se rindan, mis amadas hijas maravillosas, gracias por aceptarme, apoyarme y quererme. A mi mamá y a mi papá, quienes siempre han estado a mi lado y también a mis hermanas.

También mis agradecimientos a Luis, un extraordinario compañero, como equipo pudimos superar varios retos en el transcurso del Magister.

Agradecer al profesor Dr. Mauricio Herrera por darnos esta oportunidad de abordar este innovador tema, sus comentarios fueron precisos y muy motivantes.

Y por último, la colosal compañía de mis queridos amigos Tony y Peter.

Índice

Agradecimientos	2
Resumen	4
1. Introducción	5
2. Hipótesis y Objetivos	6
3. Relevancia del trabajo	7
4. Metodología	8
4.1 Contextualización de la Metodología	8
4.2 Datos y Metodología	9
5. Resultados	19
5.1 Otros Resultados	26
6. Conclusión	42
7. Bibliografía	44

Resumen

El objetivo de esta investigación es evaluar la efectividad del uso conjunto de Funciones Ortogonales Empíricas (EOF) y Deep Learning en la predicción de patrones espacio-temporales climáticos en Chile. Se propone un enfoque basado en el modelo de Federico Amato (Amato F. 2020) , adaptando y aplicando esta metodología a datos climáticos específicos de Chile.

Los resultados obtenidos demuestran una alta precisión en la predicción de patrones climáticos, especialmente en el pronóstico de temperaturas máximas. Se observó una eficacia notable en la representación de la variabilidad climática y la reducción de la dimensionalidad, así como un Error Absoluto Medio (MAE) bajo en la predicción de las temperaturas máximas, junto a una representación efectiva de patrones climáticos complejos.

En síntesis, en este proyecto se observó que es factible la aplicabilidad de esta metodología para analizar y predecir con precisión el clima en Chile, destacando su enfoque innovador en la modelización climática.

1. Introducción

La creciente complejidad y la urgencia de los desafíos presentados por el cambio climático exigen herramientas de predicción más avanzadas y precisas, especialmente en regiones geográficas con alta variabilidad climática como Chile. La gestión eficaz de recursos y la planificación ambiental en estas áreas dependen críticamente de nuestra capacidad para interpretar y anticipar patrones climáticos. Con este fin, este estudio se centra en explorar y adaptar metodologías innovadoras, que hasta ahora han demostrado ser prometedoras en otros contextos, para analizar datos climáticos y proyectar sus tendencias con una precisión sin precedentes.

Este trabajo se inspira en la posibilidad de que, mediante la sinergia entre técnicas computacionales avanzadas y el conocimiento específico del clima regional, lograr pronósticos que no solo reflejen la complejidad del clima, sino que también proporcionen la claridad necesaria para decisiones informadas. El propósito es dilucidar si la integración de métodos avanzados de análisis y modelización puede traducirse en un beneficio tangible para la comprensión y pronóstico del clima a nivel local y regional, todo dentro del marco de una sostenibilidad ambiental y la adaptación al cambio climático.

2. Hipótesis y Objetivos

Hipótesis:

La aplicación de un enfoque de predicción espacio-temporal que integra descomposición en funciones empíricas ortogonales (EOFs) y Deep Learning, ¿Es una herramienta precisa para el análisis de tendencias climáticas en Chile?

Objetivo General:

- Evaluar la efectividad de Deep Learning en conjunto con EOFs en la mejora de las predicciones climáticas espacio-temporal en Chile.

Objetivos Específicos:

- Analizar la distribución espacial y temporal de las variables climáticas en Chile para identificar patrones climáticos distintivos.
- Preparar y adaptar los datos climáticos históricos para su análisis mediante técnicas de Deep Learning y EOFs.
- Implementar y validar un modelo de Deep Learning que incorpore EOFs para una predicción climática precisa.
- Comparar el rendimiento predictivo de un modelo de Deep Learning completo con uno simplificado basado en las principales componentes que explican la mayor variabilidad climática.

3. Relevancia del trabajo

Esta tesis incorpora el innovador marco de trabajo propuesto por Federico Amato en (Amato F. 2020) adaptándolo para predecir patrones climáticos en Chile.

En la aplicación de este método se observó la flexibilidad, la eficacia en el análisis y la predicción espacio-temporal del clima en Chile, respondiendo a sus peculiaridades geográficas y climáticas. De manera similar, proyectos destacados como "GraphCast Learning skillful medium-range global weather forecasting" (2022), también han reconocido la importancia y el potencial del aprendizaje profundo en el pronóstico de patrones climáticos. Estas iniciativas reflejan un esfuerzo común y creciente en la comunidad científica para integrar y adaptar las técnicas de aprendizaje profundo en la predicción climática, tanto a escalas globales como regionales.

La relevancia de este trabajo radica en su contribución al entendimiento y manejo de los desafíos climáticos en Chile. Proporciona un modelo práctico para adaptar y aplicar metodologías avanzadas en contextos geográficos y climáticos específicos. Además, los resultados pueden fortalecer las políticas públicas, especialmente en la planificación de recursos hídricos y estrategias de adaptación al cambio climático, potenciando decisiones más informadas y efectivas tanto en la gestión ambiental como en el desarrollo sustentable del país.

Por lo tanto, no solo se aborda eficazmente los desafíos climáticos específicos de Chile, sino que también se alinea con las tendencias globales en la modelización climática, subrayando su papel en la evolución de la ciencia del clima y el pronóstico meteorológico mediante técnicas de aprendizaje profundo (DL)

4. Metodología

4.1 Contextualización de la Metodología

La singularidad geográfica y climática de Chile, que abarca desde desiertos áridos hasta extensos glaciares, proporciona un escenario ideal para probar y depurar el método propuesto por Federico Amato y su equipo. Al aplicar el modelo a estos datos, no solo se obtiene una visión más detallada de los patrones climáticos complejos de la región, sino que también se demuestra cómo el aprendizaje profundo puede ser adaptado eficazmente para capturar las peculiaridades de diferentes microclimas.

La implementación de este modelo tendrá un impacto significativo en la gestión de recursos naturales y la planificación ambiental, especialmente en Chile. Por ejemplo, una comprensión más precisa de los patrones de precipitación y temperatura puede influir directamente en la agricultura y la gestión del agua. Además, los hallazgos de este estudio tienen el potencial de influir en las políticas relacionadas con el cambio climático y la protección del medio ambiente, proporcionando un análisis basado en datos, permitiendo orientar a los responsables de la toma de decisiones hacia estrategias efectivas de mitigación y adaptación al cambio climático.

Siguiendo el enfoque propuesto por Amato, se descompuso los datos climáticos de Chile utilizando Funciones Empíricas Ortogonales (EOFs), lo que permite una división clara de la variabilidad espaciotemporal en componentes espaciales (coeficientes espaciales) y temporales (funciones bases temporales) separados, más manejables. Luego, estos componentes son modelados mediante redes neuronales profundas, que permiten “aprender” las particularidades geográficas de Chile, minimizando la discrepancia entre los valores predichos y los observados.

Tras entrenar el modelo, se procede a la recombinación de estos coeficientes espaciales con bases temporales, lo que permite reconstruir el campo espaciotemporal completo y realizar predicciones en áreas sin mediciones directas. La efectividad del modelo se mide con indicadores como el error absoluto medio (MAE) y se valida mediante comparaciones con datos reales y simulados. Esta validación confirma la robustez del modelo y su capacidad para generalizar a nuevos datos.

La aplicación de esta metodología en un contexto específico como el de Chile no solo contribuye a la comprensión de los desafíos climáticos y ambientales, sino que también establece un precedente para futuras investigaciones en análisis y modelización climática basados en datos

locales. Este estudio demuestra la adaptabilidad y relevancia global del enfoque de Amato, ofreciendo insights valiosos para la ciencia climática.

Para abordar la predicción de patrones climáticos, se realiza mediante el uso focalizado de la temperatura máxima diaria como variable principal. Esta elección se fundamenta en la representatividad de la temperatura máxima para evaluar las condiciones climáticas y su influencia directa en la variabilidad ambiental y los procesos de gestión de recursos. Además de la temperatura máxima, se dispone de una variedad de datasets que abarcan variables como la temperatura media, la temperatura mínima, la evaporación y la precipitación. Aunque se ha evaluado cada uno de estos conjuntos de datos, para los fines de esta tesis solo se presentan los resultados obtenidos a partir de la temperatura máxima. Los resultados derivados de las demás variables ambientales se incluirán en las secciones correspondientes, proporcionando una visión integral de las condiciones climáticas estudiadas sin detallar extensivamente el proceso analítico, el cual sigue una metodología consistente a través de todas las variables.

Al aplicar los métodos de Descomposición en Funciones Empíricas Ortogonales (EOFs) combinados con Deep Learning, se constituyen en una estrategia prometedora para la mejora de la precisión en la predicción de fenómenos climáticos. La utilización de esta única variable no solo simplifica el proceso de análisis, sino que también asegura la coherencia y claridad de la interpretación de los modelos predictivos desarrollados en el marco de este trabajo.

4.2 Datos y Metodología

Se utiliza el conjunto de datos "ERA5_tmax_1950_2022" el cual se origina del proyecto ERA5, que es parte del servicio Copernicus Climate Change Service (C3S) de la Unión Europea y producidos por el Centro Europeo de Previsiones Meteorológicas a Mediano Plazo (ECMWF). ERA5 proporciona datos climáticos y meteorológicos reanalizados, cubriendo el período desde 1950 hasta el presente, ofreciendo información detallada sobre una amplia gama de variables atmosféricas, terrestres y oceánicas, siendo además conocido por su alta resolución espacial (0.25×0.25 grados) y temporal. Es ampliamente utilizado en investigaciones climáticas, estudios ambientales y aplicaciones de modelización meteorológica.

El identificador específico "tmax" en "ERA5_tmax_1950_2022" sugiere que este conjunto de datos se centra en la temperatura máxima diaria a nivel global o regional, proporcionando registros de temperaturas máximas desde 1950 hasta 2022.

Los datos que se utilizaron se encuentran en una malla de 6355 puntos ubicados entre las coordenadas geográficas (latitud -17.5° a -56.0° y longitud -76.0° a -66.0°). Cada uno de los puntos tiene asociado datos históricos con series temporales de temperatura máxima (tmax), además de su posición, caracterizada por las coordenadas geográficas – longitud y latitud.

El primer paso en el manejo de este dataset fue extraer las variables esenciales: longitud, latitud, tiempo y tmax. Estos datos proporcionan la base para el análisis, ofreciendo información sobre las ubicaciones geográficas y los registros temporales, junto con las temperaturas máximas correspondientes.

Una vez obtenidos estos datos, se procede a aplanar la matriz de temperatura máxima, transformándola desde una estructura tridimensional a una serie unidimensional. Esta transformación es crítica ya que simplificó los datos, facilitando su manejo y análisis posterior.

Luego, para asegurar que cada temperatura aplanada mantuviera su contexto espacial y temporal, se replican los valores de latitud, longitud y tiempo. Esto garantizó que cada registro de temperatura se vinculara correctamente con su ubicación y momento específicos.

El último paso fue organizar estos datos en un DataFrame de pandas. Como se observa en la tabla 1, se alinea cada registro de temperatura con su respectiva coordenada geográfica y temporal, lo que resulta en una tabla estructurada.

En síntesis, se transforma un conjunto de datos climáticos complejos en un DataFrame estructurado, estableciendo una base sólida para el análisis de los patrones climáticos en Chile.

	time	latitude	longitude	tmax
0	0	-17.5	-76.00	20.583710
1	0	-17.5	-75.75	20.675873
2	0	-17.5	-75.50	20.619415
3	0	-17.5	-75.25	20.507324
4	0	-17.5	-75.00	20.443390
...
169443360	26662	-56.0	-67.00	8.452789
169443361	26662	-56.0	-66.75	8.351440
169443362	26662	-56.0	-66.50	8.144287
169443363	26662	-56.0	-66.25	7.930939
169443364	26662	-56.0	-66.00	7.778046

169443365 rows × 4 columns

Tabla 1 – Dataframe de datos estructurados. Elaboración propia

In [10]: y

Out[10]:

	time	0	1	2	3
latitude longitude					
-56.0 -76.00	7.436432	7.325989	7.412354	6.728241	
-75.75	7.469635	7.343445	7.422302	6.739868	
-75.50	7.475433	7.335968	7.418976	6.749817	
-75.25	7.455505	7.335968	7.409027	6.755646	
-75.00	7.414825	7.347595	7.401550	6.755646	
...
-17.5 -67.00	11.052826	10.903412	11.760193	9.562592	
-66.75	13.076904	13.008820	13.992615	11.575867	
-66.50	14.845245	14.723206	15.582489	13.604919	
-66.25	17.505280	17.729431	17.734406	16.381165	
-66.00	17.814941	17.902954	17.987610	16.389465	

6355 rows × 26663 columns

Tabla 2 – Temperaturas máximas diarias. Fuente elaboración propia

Es así que se adopta un enfoque único para la división y organización de los datos, con la creación de dos tablas fundamentales a partir del DataFrame.

La primera de ellas, tabla 2, enfocada a la variable objetivo: las temperaturas máximas diarias. Cada fila en esta tabla representa una estación climática, identificada por su latitud y longitud, pudiendo examinar en detalle las temperaturas en diversas ubicaciones geográficas

In [11]: X

Out[11]:

	latitude	longitude
0	-56.0	-76.00
1	-56.0	-75.75
2	-56.0	-75.50
3	-56.0	-75.25
4	-56.0	-75.00
...
6350	-17.5	-67.00
6351	-17.5	-66.75
6352	-17.5	-66.50
6353	-17.5	-66.25
6354	-17.5	-66.00

6355 rows × 2 columns

Tabla 3 – Coordenadas geográficas (grados) de Estaciones. Elaboración propia

La tabla 3 es muy importante para entender cómo las características geográficas impactan en las temperaturas observadas.

Una vez establecidas estas dos tablas, se procede a dividir las en una proporción de 70% para testeo y 30% de entrenamiento. Esta división inicial fue diseñada con el objetivo de preparar los datos para técnicas de reducción de dimensionalidad, adaptándose a la complejidad de los datos climáticos.

Además de esta división inicial, se adoptó un enfoque distinto al convencional para la asignación de los datos en las fases de entrenamiento, validación y testeo del modelo. En vez de dividir los datos en un 60% para entrenamiento, 20% para validación, 20% para testeo y luego de varias iteraciones, se determina que un 70% del total de los datos al conjunto de testeo, es lo óptimo. Esta proporción inusualmente alta se eligió intencionalmente para someter el modelo a un escenario de prueba más riguroso, abarcando un amplio espectro de condiciones climáticas. Posteriormente, el 21% de los datos se utilizó para entrenamiento, y el 9% restante para validación. Esta distribución refleja el propósito de equilibrar un entrenamiento adecuado del modelo con la necesidad de una validación y testeo exhaustivos.

Al adoptar este enfoque de dos etapas en la estructuración y preparación de los datos, se establece una sólida base para el análisis en profundidad de los patrones climáticos en Chile, teniendo en cuenta tanto la variabilidad geográfica como las complejidades inherentes a los datos climáticos. De esta forma, no solo se logra organizar los datos de manera que faciliten el análisis posterior, sino que también asegurar que las técnicas de modelado avanzadas que se aplicarán más adelante operen sobre un conjunto de datos óptimo y representativo.

Una parte esencial y muy motivante, es el proceso de Descomposición en Valores Singulares (SVD, por sus siglas en inglés) es una técnica matemática poderosa utilizada en el análisis de datos, particularmente útil en el campo de la climatología para identificar patrones significativos en conjuntos de datos complejos y multidimensionales. Al aplicar SVD a los datos climáticos, se desglosa la matriz original de temperaturas en componentes fundamentales que revelan las estructuras subyacentes y la variabilidad inherente al sistema que se está estudiando.

El resultado de este proceso se presenta en tres elementos principales: la matriz U , que contiene los modos temporales; el vector S , que consta de los valores propios singulares y mide la importancia (varianza explicada) de cada componente; y la matriz V^T , que representa los

componentes espaciales. Estos componentes son clave para comprender cómo se distribuye la variabilidad climática a través del tiempo y el espacio (Figura 1).

En la práctica, con SVD es posible observar que una pequeña cantidad de componentes (EOFs) puede explicar una gran proporción de la variabilidad total en los datos climáticos. Esto no es sorprendente, ya que es característico de datos donde existen patrones dominantes, como ciclos estacionales o tendencias climáticas a gran escala. Por lo tanto, al centrar la atención en los primeros componentes principales, se puede capturar la mayoría de la información relevante, mientras que los modos subsiguientes a menudo reflejan patrones más detallados o ruido aleatorio.

Este fenómeno de reducción de dimensionalidad es lo que hace que SVD sea tan valioso en el análisis de datos climáticos. Permite simplificar los conjuntos de datos complejos, facilitando así la interpretación de los fenómenos climáticos y la implementación de modelos predictivos más eficientes. La Figura 1 muestra las matrices y vectores resultantes que encapsulan las dimensiones más significativas de la variabilidad climática observada.

```
Diagonal S:
[8.8825903e+02 3.8340625e+02 1.7660229e+02 ... 4.3522973e-02 3.6939200e-02
 0.0000000e+00]

Matrix U:
[[ 4.5448628e-01 -1.6810211e+00 9.9422622e-01 ... -1.4100815e-03
  -8.0901338e-03 9.9969488e-01]
 [ 1.4112687e+00 -1.5739706e+00 2.2627880e-01 ... 4.2971517e-03
  6.5954947e-03 9.9992633e-01]
 [-1.8694229e+00 -2.7992901e-01 -1.5389957e-01 ... 1.2297867e-02
  -3.1963617e-03 9.9987566e-01]
 ...
 [ 4.9407038e-01 9.0699887e-01 3.8473544e-01 ... -1.4758648e-01
  -2.5166281e-02 1.0003505e+00]
 [ 5.8235818e-01 9.0518892e-01 4.5298150e-01 ... 1.2700078e-01
  2.4029516e-02 1.0011379e+00]
 [-9.6862644e-01 4.2683238e-01 -4.5061342e-02 ... -4.8826870e-02
  7.7507216e-03 9.9955553e-01]]

Matrix V^T:
[[ 0.00615898 -0.00481248 0.00067966 ... -0.00246945 0.00215276
  0.01082531]
 [ 0.00620776 -0.0082435 0.00597052 ... 0.00117118 0.00145505
  0.01187403]
 [ 0.00638902 -0.0081687 0.00454752 ... -0.0057449 0.00445604
  0.00728741]
 ...
 [ 0.00740309 -0.00923126 -0.00102735 ... 0.00249095 -0.00318978
  0.12278331]
 [ 0.00736942 -0.00822524 -0.00431568 ... -0.00217014 -0.00319243
  -0.0076703 ]
 [ 0.00644768 -0.00956585 0.0051911 ... 0.00750519 0.00365213
  -0.02387552]]
```

Fig.1 Resultados de aplicar SVD al conjunto de datos. Elaboración propia

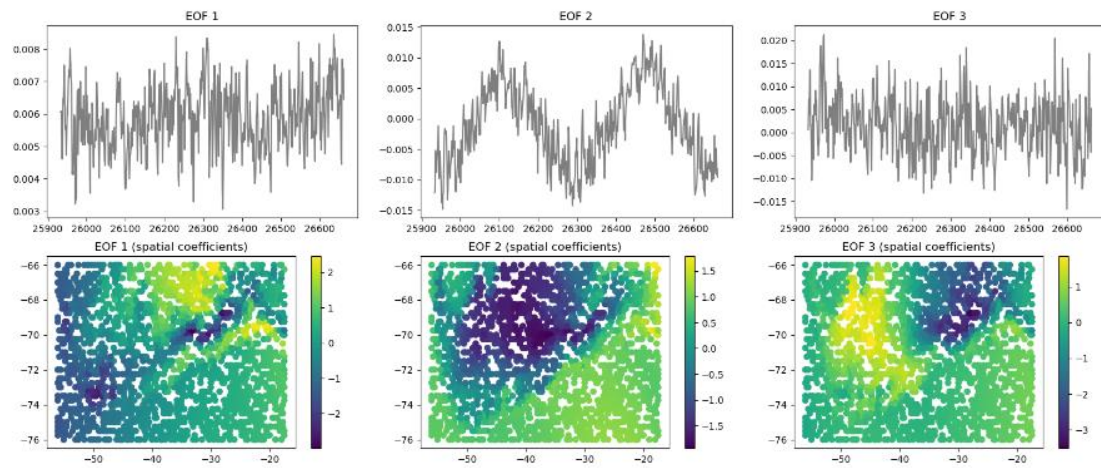


Fig. 2 Combinación de análisis temporales y espaciales de los tres primeros Modos Empíricos Ortogonales (EOFs) del conjunto de datos. *Elaboración propia*

La Figura 2 captura la esencia de las variaciones en un período de dos años. Los gráficos superiores representan las fluctuaciones temporales de cada uno de los EOFs. En el EOF 1, se observa una serie de fluctuaciones consistentes a lo largo del tiempo, lo cual puede estar reflejando un patrón climático estable y recurrente. El EOF 2 muestra un comportamiento más dinámico, con máximos y mínimos más pronunciados, sugiriendo una respuesta a ciclos estacionales o la presencia de fenómenos periódicos que impactan en el clima. El EOF 3, con su variabilidad más errática, podría estar captando influencias menos coherentes o el 'ruido' dentro del sistema climático. Los gráficos inferiores ilustran la distribución espacial de los coeficientes correspondientes a cada EOF, proporcionando una visualización de cómo estos patrones temporales se manifiestan a través del espacio geográfico. En el EOF 1, se destaca una región con valores particularmente altos, lo que puede ser indicativo de un fenómeno climático con un impacto localizado. Los EOFs 2 y 3 muestran patrones espaciales más complejos, lo que sugiere que los efectos climáticos varían significativamente de una región a otra.

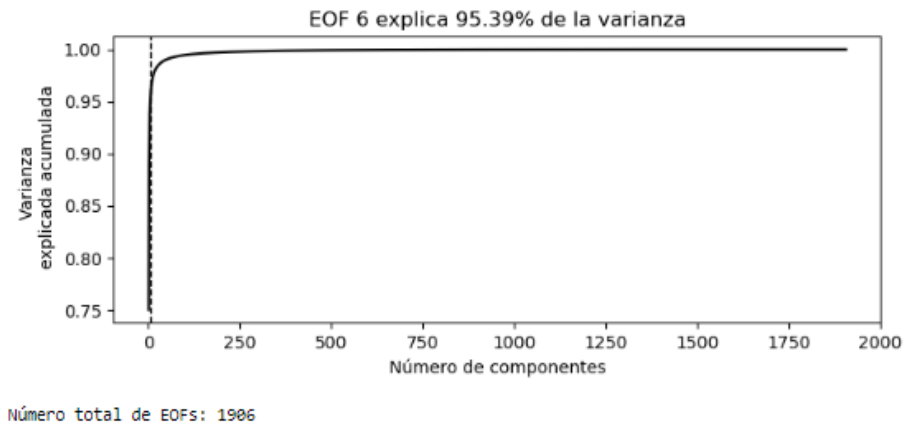


Fig. 3 Gráfico de la varianza explicada acumulada de los Componentes Principales. Elaboración propia

En la Figura 3, se observa que un conjunto muy reducido de componentes principales es suficiente para capturar la esencia de los datos climáticos analizados, en la cual los primeros 6 componentes explican más del 95% de la varianza total de la temperatura máxima, lo que es un indicativo de fuertes patrones subyacentes en los datos. Esto sugiere, además, que los fenómenos climáticos que se están estudiando tienen unos pocos modos dominantes de variabilidad, lo que simplifica significativamente el análisis y la interpretación de grandes conjuntos de datos. La presencia de 1906 componentes posibles y la concentración de la varianza en solo 6 EOFs resalta la eficacia de utilizar estas técnicas para identificar los aspectos más significativos del sistema climático en cuestión.

```
True - reconstructed data:
[[-4.1961670e-05  2.7465820e-04 -4.9591064e-05 ... -6.6757202e-05
  -1.3351440e-05  3.4332275e-05]
 [-1.2969971e-04  4.9209595e-04 -4.3487549e-04 ... -2.8610229e-04
  -4.5013428e-04  4.5776367e-05]
 [-3.7479401e-04 -1.2626648e-03 -2.0599365e-04 ... -3.2997131e-04
  -4.9018860e-04  3.4332275e-05]
 ...
 [ 8.5830688e-05  5.9509277e-04 -8.7738037e-05 ... -5.7220459e-05
  1.1444092e-05  3.8146973e-06]
 [ 6.8664551e-05  5.9127808e-04 -7.8201294e-05 ... -5.7220459e-05
  5.7220459e-06  1.3351440e-05]
 [ 1.7452240e-04 -1.2569427e-03  7.3432922e-05 ... -8.5830688e-05
  -3.2424927e-05  3.8146973e-06]]
```

Fig. 4 –Matriz reconstruida (EOFs). Elaboración propia

El último paso en el análisis SVD, consistió en verificar la capacidad de los EOFs para reconstruir la matriz original de datos climáticos, Z . La matriz reconstruida, ajustada para la media temporal, exhibió diferencias mínimas que se pueden observar en la Figura 4, al compararla con la matriz original, evidenciando la eficacia de los EOFs seleccionados para capturar la variabilidad

climática esencial con una menor cantidad de datos. Este resultado confirma la precisión de la descomposición y su aplicabilidad para simplificar el análisis climático manteniendo la integridad de la información.

La integración de los análisis temporales y espaciales, junto con el estudio de la varianza explicada acumulada, ha proporcionado una perspectiva integral de la variabilidad climática, al concentrar la varianza en solo 6 componentes, destacando patrones climáticos consistentes y significativos, lo cual es esencial para comprender cómo la variabilidad climática interactúa con sus impactos regionales.

Ahora bien, la reconstrucción de los datos climáticos, que resultó en mínimas discrepancias con los datos originales, valida aún más la representatividad de estos componentes, lo que demuestra que la información climática clave puede ser efectivamente capturada y representada por un número reducido de factores, facilitando la optimización de modelos predictivos y la condensación del análisis para aplicaciones climáticas y meteorológicas.

En la etapa final de preparación de los datos para el entrenamiento del modelo, se llevó a cabo una división y normalización estratégica de los datos derivados de la SVD. Esta preparación ha proporcionado los siguientes conjuntos, con sus respectivos porcentajes:

1. X_{train} y X_{valid} : Los datos de entrenamiento y validación respectivamente. X_{train} consta de aproximadamente el 70% de los datos originales, mientras que X_{valid} representa el 30% restante, asegurando así una distribución equilibrada para un entrenamiento y validación eficaces.
2. u_{train} y u_{valid} : Estos subconjuntos contienen los EOFs para entrenamiento y validación, siguiendo la misma proporción de división que X_{train} y X_{valid} .
3. s : Este vector incluye los valores singulares ajustados al número de EOFs seleccionados, lo que representa la esencia de la variabilidad en los datos.
4. v : Representa los modos espaciales, también ajustados en número para coincidir con los EOFs seleccionados.

Estos conjuntos y sus proporciones son esenciales para un modelado efectivo, ya que permiten que el modelo aprenda de una amplia gama de datos y se valide contra un conjunto independiente, garantizando así su capacidad de generalizar y predecir patrones climáticos de manera precisa.

Modelo de Deep Learning

Basado en la metodología del trabajo de Federico Amato, se ha implementado un modelo de aprendizaje profundo en Keras, orientado a la predicción y reconstrucción de patrones climáticos a partir de datos geográficos y temporales. Este modelo refleja con precisión la estructura y técnicas propuestas en el estudio de Amato.

La arquitectura del modelo consta de capas de entrada para las variables geográficas, seguidas de múltiples capas densas intercaladas con capas de normalización por lotes (BatchNormalization). Estas capas densas, con 100 unidades neuronales cada una, emplean la función de activación ReLU y una inicialización 'he_normal', optimizada para activaciones ReLU. Este diseño es esencial para extraer patrones complejos y aprender eficientemente las dinámicas de los datos climáticos.

El modelo también incluye una salida auxiliar destinada a predecir los coeficientes espaciales u , y una salida principal que reconstruye los datos climáticos. Esta última se logra a través de una función de recomposición que utiliza estos coeficientes junto con los valores singulares s y los modos espaciales v , derivados de la SVD, para reconstruir la matriz original de datos climáticos.

En términos de entrenamiento, el modelo aplica el enfoque 'One Cycle Scheduler' para optimizar la tasa de aprendizaje, una estrategia eficaz para acelerar la convergencia y evitar el sobreajuste. Se incorporó un mecanismo de 'Early Stopping' para detener el entrenamiento si no hay mejoras en un determinado número de épocas, lo cual es necesario para preservar la capacidad de generalización del modelo.

Este modelo, basado fielmente en la metodología de Amato, constituye una aplicación directa y técnica de aprendizaje profundo para el análisis avanzado de datos climáticos. Su capacidad para procesar y analizar datos de alta dimensión y complejidad lo convierte en una herramienta poderosa para desentrañar patrones climáticos y avanzar en la comprensión de fenómenos meteorológicos.

```

model = keras.Model(inputs=[input_A], outputs=[output, aux_output])
model.compile(loss=['mae', 'mae'], loss_weights=[1.0,0.0], optimizer="nadam")

history = model.fit([X_train], [y_train.values[:1334,:], u_train], epochs=n_epochs,
                    validation_data=([X_valid], [y_train.values[1334:,:], u_valid]),
                    callbacks=[onecycle, early_stopping_cb], verbose = 0)

y_hat_test, y_pred_test = model.predict([X_test])

140/140 [=====] - 86s 517ms/step

```

Fig. 5 – Modelo Deep Learning. Elaboración propia

La Figura 5 muestra la ejecución del modelo de aprendizaje profundo, detallando la compilación, el entrenamiento y la evaluación de este. El modelo se compila con una función de pérdida basada en el error absoluto medio y se optimiza utilizando el algoritmo Nadam. Durante el entrenamiento, se aplican técnicas de regulación como el One Cycle Scheduler para ajustar la tasa de aprendizaje y Early Stopping para prevenir el sobreajuste, mejorando así la generalización del modelo. Finalmente, el modelo entrenado se evalúa utilizando un conjunto de datos de prueba para predecir y verificar su rendimiento predictivo. Este proceso integral garantiza que el modelo esté bien afinado para capturar los patrones climáticos complejos presentes en los datos.

5. Resultados

Los resultados obtenidos del modelo de aprendizaje profundo muestran un rendimiento sobresaliente, en la Figura 6 se observa un MAE de 1.6952 lo que refleja una alta precisión en las predicciones climáticas. Es notable que el modelo mantenga una eficacia comparable incluso cuando la distribución de los datos se haya ajustado a un 21% para entrenamiento y aproximadamente un 9% para validación (Figura 6), con una diferencia mínima en comparación con la convención de 60% de datos de entrenamiento y 20% de validación, observado en la Figura 7 con un valor de MAE de 1.6396. Esta mejora en la eficiencia del uso de datos no solo optimiza los recursos computacionales, sino que también evidencia la robustez del modelo frente a variaciones en la asignación de datos. La habilidad del modelo para lograr resultados similares con un conjunto de entrenamiento y validación más pequeño subraya la calidad del aprendizaje y la generalización que el modelo ha logrado, lo cual es fundamental para aplicaciones prácticas en el campo de la meteorología y la climatología.

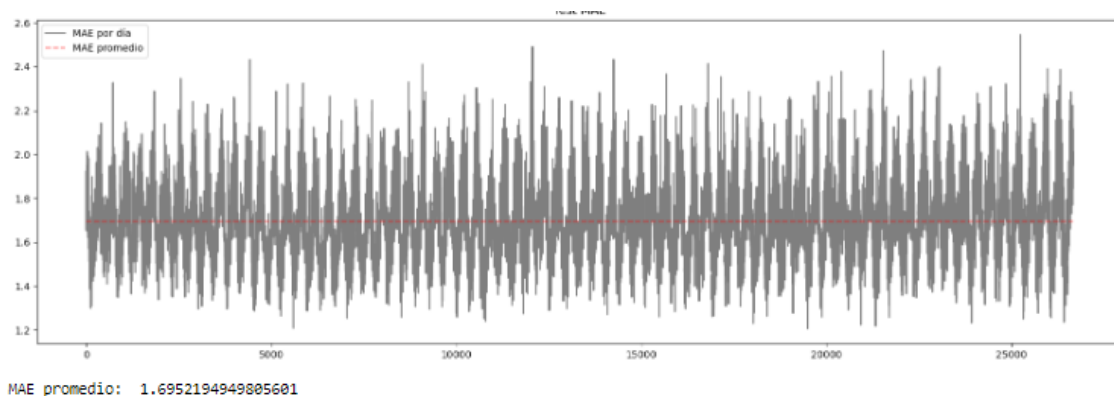


Figura 6 – Gráfico MAE por día y promedio. Elaboración propia

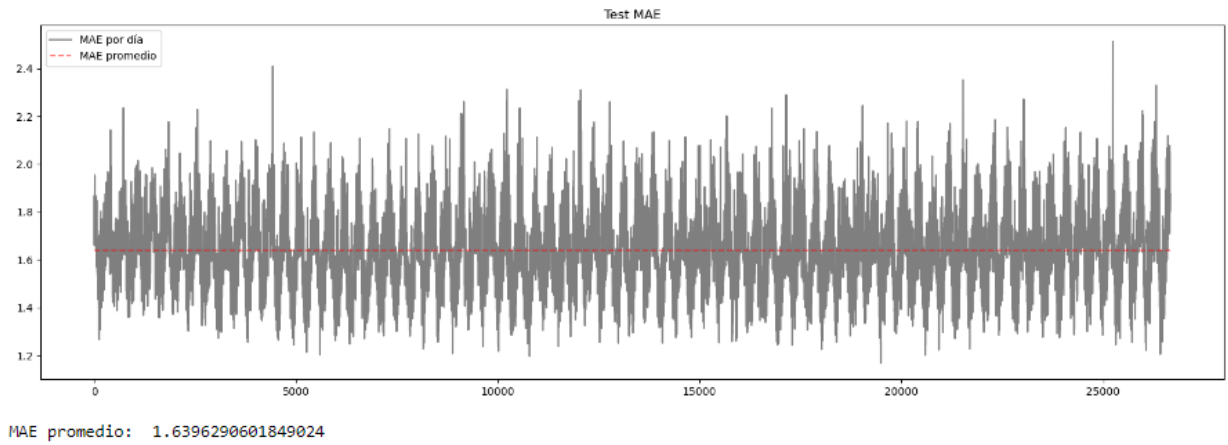


Fig. 7 – Gráfico MAE por día y promedio. Elaboración propia

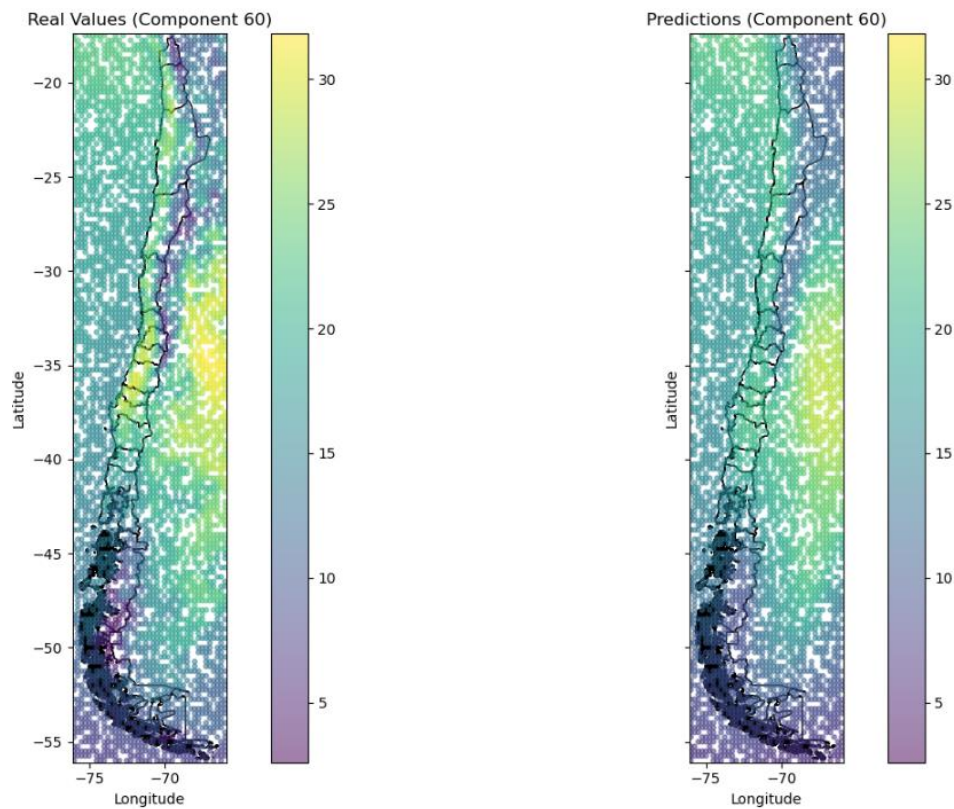


Fig. 8 – Mapa comparativo de valores reales y predicción (60). Elaboración propia

La Figura 8 compara los valores reales observados y las predicciones realizadas por el modelo para la componente de tiempo 60 (2 de marzo de 1950) del conjunto de datos. En el primer mapa (izquierda), que exhibe los valores reales, se observa una distribución espacial de los datos a lo largo de Chile, con una gradación de colores que varía desde tonos morados, indicando valores inferiores, hasta tonos amarillos que representan valores superiores.

El segundo mapa (derecha), que muestra las predicciones del modelo, refleja una distribución de colores similar, lo que sugiere una alta correspondencia con los valores reales. Esta similitud es un indicativo del buen desempeño del modelo en la captura de la variabilidad espacial de la componente analizada.

Cabe destacar la presencia de áreas sin datos en ambos mapas, lo que corresponde a los puntos que fueron retenidos para el entrenamiento del modelo y, por ende, no están presentes en el conjunto de prueba. La ausencia de predicciones en estas zonas es una consecuencia directa de la metodología de separación de datos, lo cual es una práctica estándar en la validación de modelos predictivos.

El análisis general de las Fig. 7 y Fig. 8, sugiere que el modelo ha logrado una muy buena precisión en predecir la componente climática en estudio, lo que es evidente a partir de la coherencia visual entre los valores reales y predichos. Este nivel de precisión es prometedor para futuras aplicaciones en el modelado climático y en la toma de decisiones basada en datos. La efectividad del modelo es aún más significativa dado que se ha logrado con un conjunto de entrenamiento reducido, lo que subraya la eficiencia del enfoque metodológico adoptado en este trabajo de investigación.

Para profundizar en el análisis se realizó una predicción utilizando el modelo previamente entrenado sobre una malla completa que incluye todas las estaciones en estudio. Este proceso se realiza para obtener una visualización más clara y detallada de los resultados del modelo en el contexto geográfico completo. Al aplicar el modelo a este conjunto de datos, se buscaba verificar la capacidad del modelo para predecir patrones climáticos en un área extensa, facilitando así la interpretación y el análisis de su comportamiento general y su precisión en diferentes localizaciones.

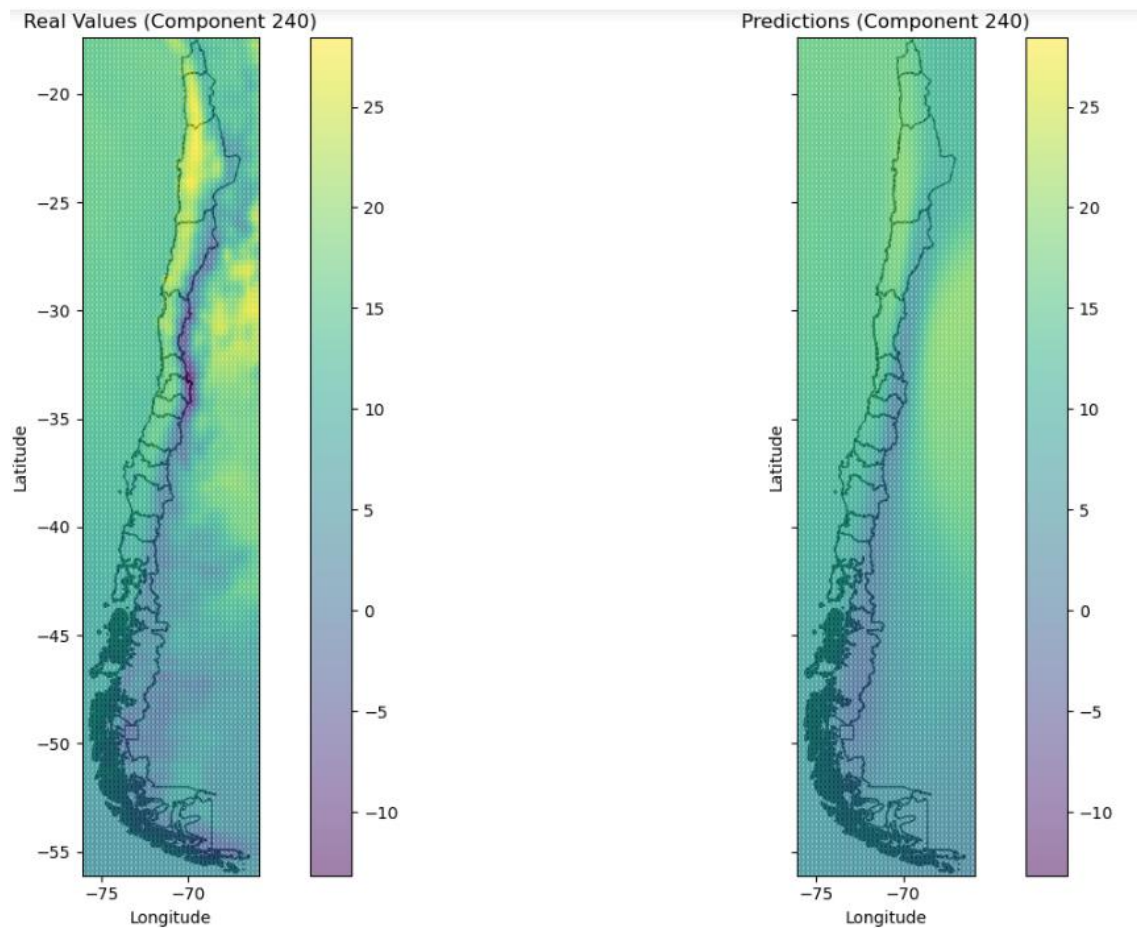
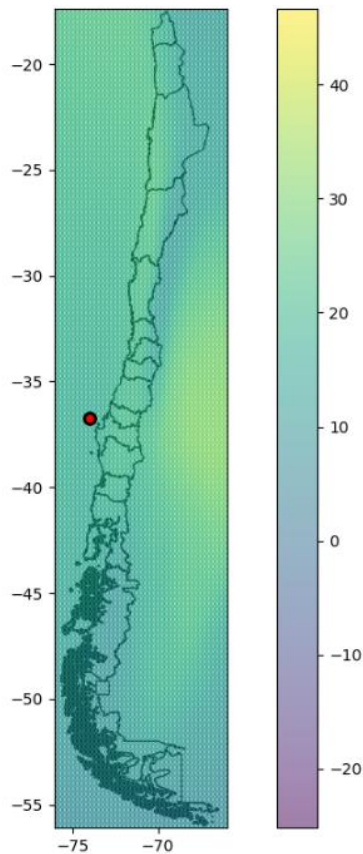


Fig. 9 – Mapa comparativo de valores reales y predicción (29-08-1950). Elaboración propia

La Figura 9 proporciona una visualización directa del rendimiento del modelo entrenado, a través de la comparación entre los valores reales y predichos en una malla completa. Los resultados muestran una cobertura exhaustiva de la región en estudio y revelan la capacidad del modelo para predecir consistentemente a lo largo de diversas ubicaciones geográficas. La alineación general de las predicciones con los valores reales demuestra la efectividad del modelo en capturar las tendencias climáticas esenciales. Aunque la precisión puede variar en puntos específicos, en general, el modelo refleja con éxito los patrones climáticos observados, validando su aplicación para análisis climáticos y potencialmente para la toma de decisiones en contextos relacionados con el clima. Las diferencias notadas en la densidad de las predicciones enfatizan áreas para futuras mejoras y ajustes del modelo.

Mapa Predicho (Todos los Componentes)



Mapa Predicho (6 Componentes)

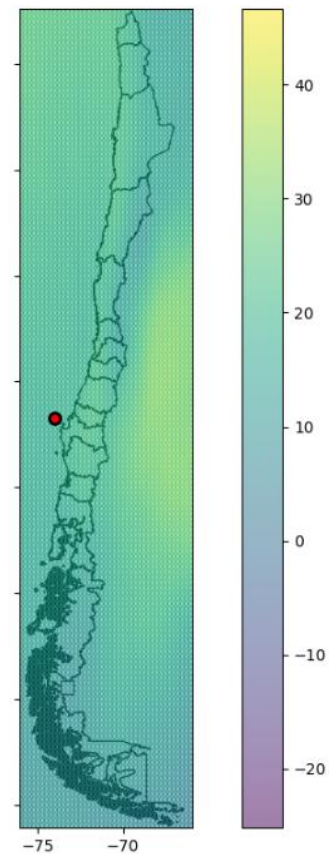


Fig. 10 – Mapa comparación de modelos. Elaboración propia

Otro análisis que se incorporó para profundizar aún más en este estudio y realizar una comparación eficaz, es un entrenamiento adicional del modelo utilizando únicamente las primeras 6 componentes, las cuales explican más del 95% de la varianza acumulada de los datos. Este enfoque selectivo permite reducir significativamente la dimensionalidad de los datos, lo que puede tener múltiples beneficios: desde una mayor eficiencia computacional hasta una posible mejora en la capacidad de generalización del modelo debido a la reducción del ruido y la complejidad. Este segundo modelo, que opera con una representación simplificada del conjunto de datos (Figura 10), será evaluado en contraste con el modelo original que utiliza todas las componentes disponibles. La comparación entre ambos modelos no solo ofrece insights sobre la importancia de las componentes seleccionadas en la predicción de patrones climáticos, sino que también pone a prueba la hipótesis de que un modelo más parsimonioso puede ser igualmente efectivo, al manejar grandes volúmenes de datos.

O sea, implementar esta estrategia de reducción de dimensionalidad se alinea con prácticas de modelado eficientes y proporciona una valiosa oportunidad para entender mejor la relevancia de las componentes dominantes en los datos climáticos. Además, permite explorar cómo la simplificación de los datos afecta la interpretación de los resultados y la toma de decisiones en aplicaciones climáticas.

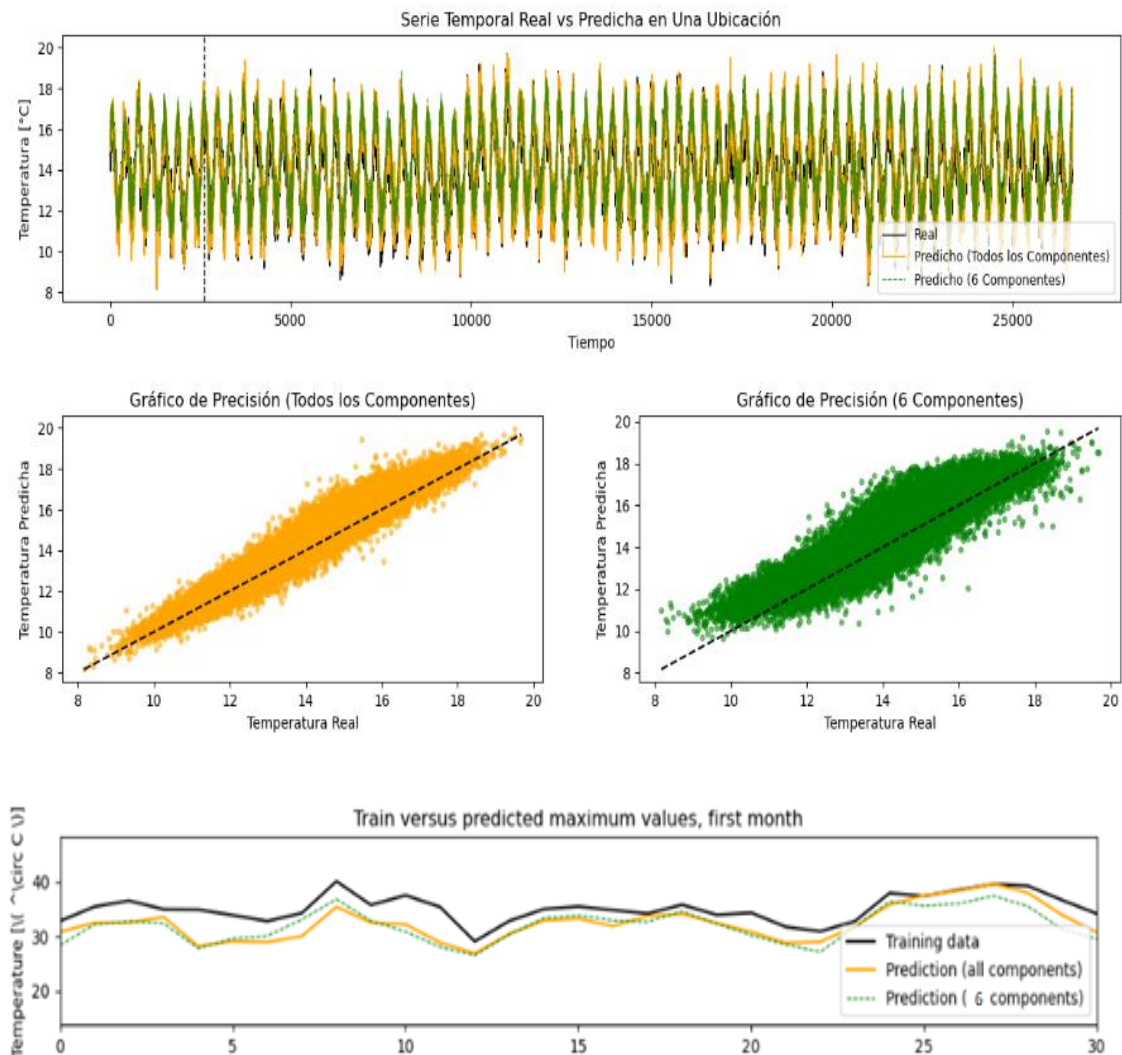


Fig. 11 – Gráficos predicción de temperatura para una ubicación específica. Elaboración propia

La Figura 11 proporciona una evaluación detallada de las predicciones de temperatura para una ubicación específica, marcada con un punto rojo en el mapa de la comparación de modelos previamente presentado en la Figura 10. En la parte superior, un gráfico de series temporales muestra las temperaturas reales en comparación con las predicciones del modelo utilizando todos los componentes y un modelo simplificado con 6 componentes. La consistencia entre las líneas de las series temporales real y predicha demuestra una sincronización notable entre las fluctuaciones reales y las estimadas por ambos modelos en el tiempo indicado por la línea vertical, correspondiente al índice de tiempo seleccionado (2590).

En la parte media, dos gráficos de dispersión evalúan la precisión de los modelos en la estación específica indicada (índice 3165). El gráfico de la izquierda muestra la relación entre las temperaturas reales y las predicciones hechas por el modelo que utiliza todos los componentes, y el de la derecha corresponde al modelo que utiliza solo 6 componentes. Ambos gráficos revelan una correlación fuerte, como lo indica la proximidad de los puntos a la línea de identidad (la línea diagonal que representa una coincidencia perfecta entre las temperaturas reales y predichas). Esta correlación sugiere que la reducción de componentes no ha comprometido la precisión de las predicciones en este punto específico.

En la parte inferior, se realiza un análisis de un mes, en el que se observa la proximidad entre los datos reales y las predicciones. Es notable cómo las predicciones que incorporan todos los componentes se alinean estrechamente con los datos de entrenamiento, mostrando una sinergia casi paralela a lo largo de todo el rango temporal. Esto indica una alta precisión del modelo en la captura de las tendencias máximas diarias. Por otro lado, la predicción que se basa solo en 6 componentes, aunque ligeramente más alejada de la serie de entrenamiento, sigue manteniendo un nivel de precisión admirable. La congruencia de esta con las predicciones completas y los datos reales sugiere que un modelo simplificado puede ser suficiente para capturar las tendencias climáticas significativas, ofreciendo una alternativa eficiente en términos de recursos computacionales sin comprometer en gran medida la calidad de la predicción. Esta observación es crucial, ya que subraya la capacidad del modelo de aprendizaje profundo para discernir y representar los patrones climáticos esenciales incluso con un conjunto limitado de variables, lo que es particularmente valioso en aplicaciones prácticas donde la velocidad y la eficiencia son fundamentales.

Los resultados ilustrados en estos gráficos son excepcionales, indicando que incluso con una reducción significativa en la cantidad de componentes utilizadas, el modelo simplificado mantiene una alta precisión, comparable al modelo original con todos los componentes. Esta observación resalta la eficacia del modelo en capturar los elementos esenciales de la variabilidad climática con un número reducido de variables, lo que implica una optimización de la complejidad del modelo sin sacrificar la calidad de las predicciones. La efectividad de ambos modelos en este punto específico es prometedora y subraya el potencial de aplicar estos métodos para estudios climáticos precisos y eficientes en términos de recursos computacionales.

5.1 Otros Resultados

En las visualizaciones presentadas desde la Figura 12 en adelante, se destacan los resultados conseguidos al aplicar el modelo en diferentes variables medioambientales. Aunque se requiere un análisis más profundo para entender completamente la distribución y las características específicas de cada variable, es posible observar indicadores claros del rendimiento del modelo.

A través de un análisis visual que compara los resultados reales con los pronosticados, se evidencia la capacidad del modelo para replicar con precisión los patrones observados en diversas variables. Esta observación se corrobora en el caso de la temperatura máxima, donde los resultados obtenidos no solo son positivos, sino que están firmemente respaldados por las visualizaciones y por un Error Absoluto Medio (MAE) bajo. Estos resultados confirman la eficacia del modelo, no solo para la variable específica de la temperatura máxima, sino también para un espectro más amplio de variables medioambientales, subrayando su robustez y precisión en el ámbito de la modelización climática

Data set de temperatura Media

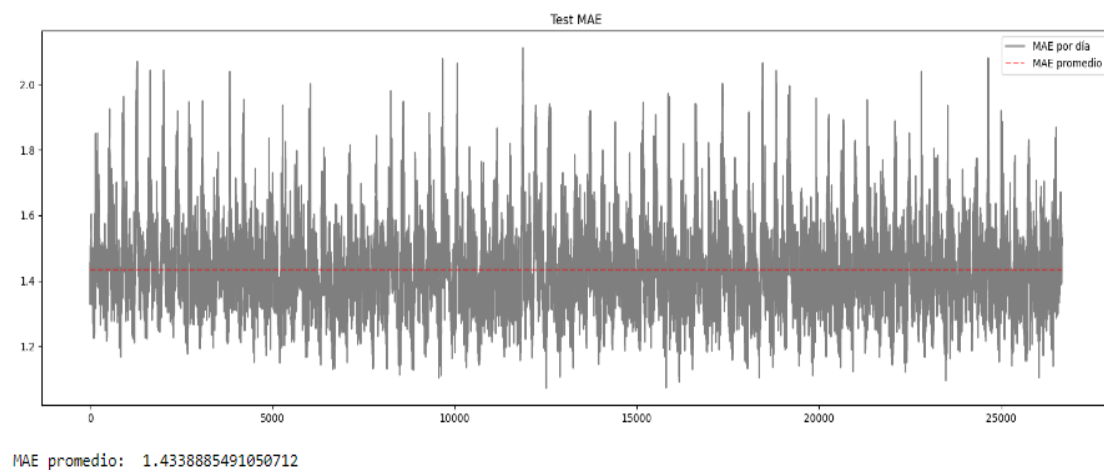


Fig. 12 – Gráfico Temperatura Media. Elaboración propia

El modelo ha arrojado resultados significativos al evaluarse con otra variable, la temperatura media, reflejados en las figuras. La consistencia del modelo se manifiesta en el gráfico de MAE de la Figura 12, donde se observa que el modelo mantiene una línea de error promedio estable a pesar de las fluctuaciones naturales de los datos diarios. Esta estabilidad en el MAE es indicativa de la fiabilidad del modelo a lo largo del período de estudio.

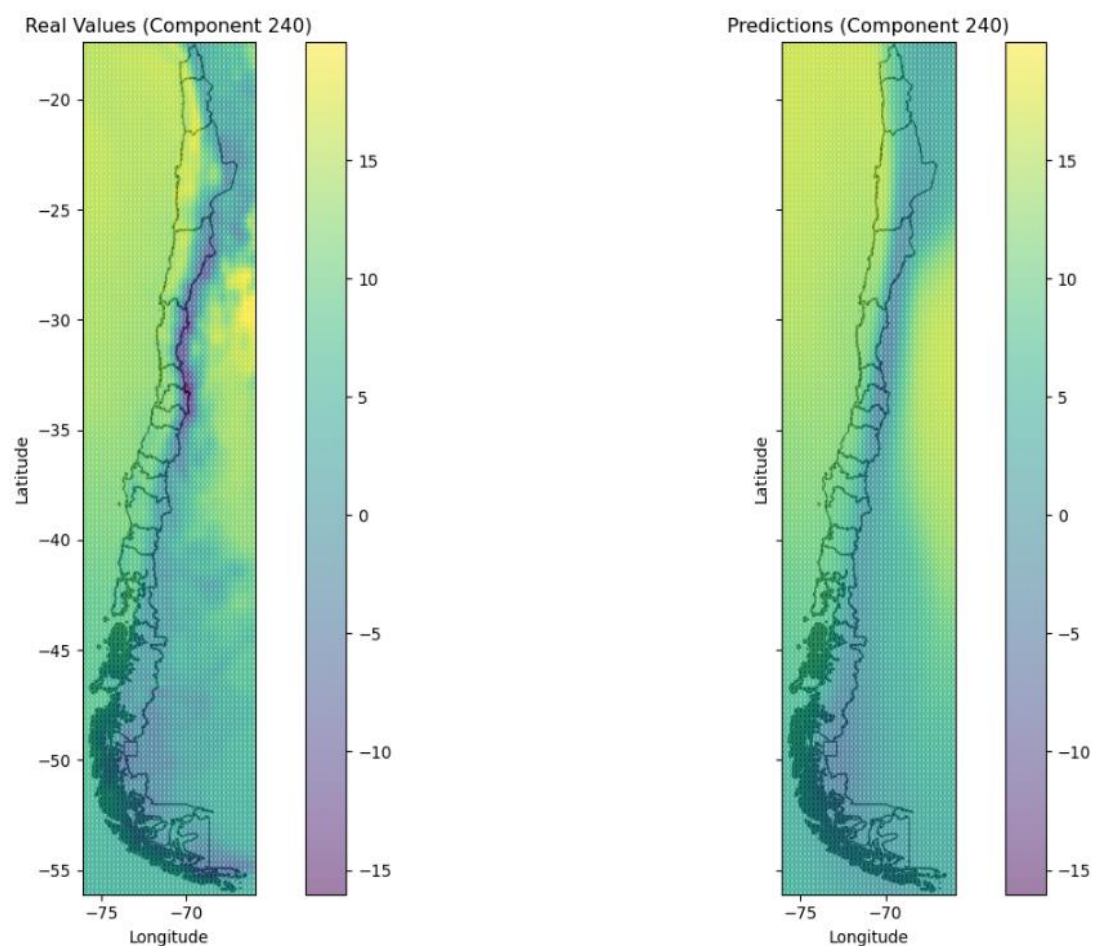
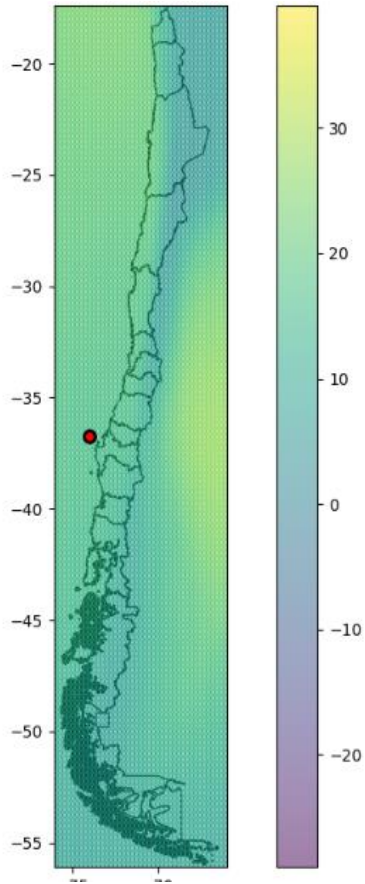


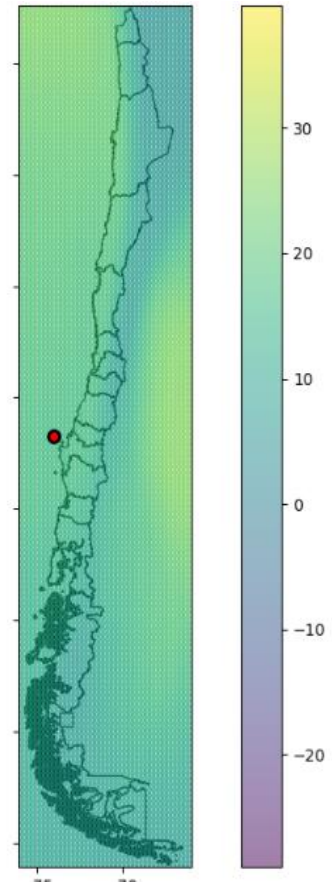
Fig. 13 – Mapa comparativo valores reales y predicción temperatura media (240). Elaboración propia

Al observar la Figura 13, los mapas de calor que representan valores reales y predicciones de temperatura media muestran una notable similitud, lo cual es un testimonio de la habilidad del modelo para recrear con precisión la distribución geográfica de la temperatura. Si bien existen algunas discrepancias entre los valores reales y las predicciones, estas no son lo suficientemente significativas como para socavar la utilidad general del modelo.

Mapa Predicho (Todos los Componentes)



Mapa Predicho (6 Componentes)



*Fig. 14 – Mapa comparativo valores predicho (todos) y predicho (6) temperatura media.
Elaboración propia*

La Figura 14, que contrasta las predicciones utilizando todos los componentes contra un modelo reducido a 6 componentes, sugiere que la simplificación del modelo mediante la reducción de componentes es viable y no impacta drásticamente en la exactitud de las predicciones. Este hallazgo es particularmente relevante, ya que indica que el modelo puede ser optimizado para eficiencia computacional sin sacrificar la precisión.

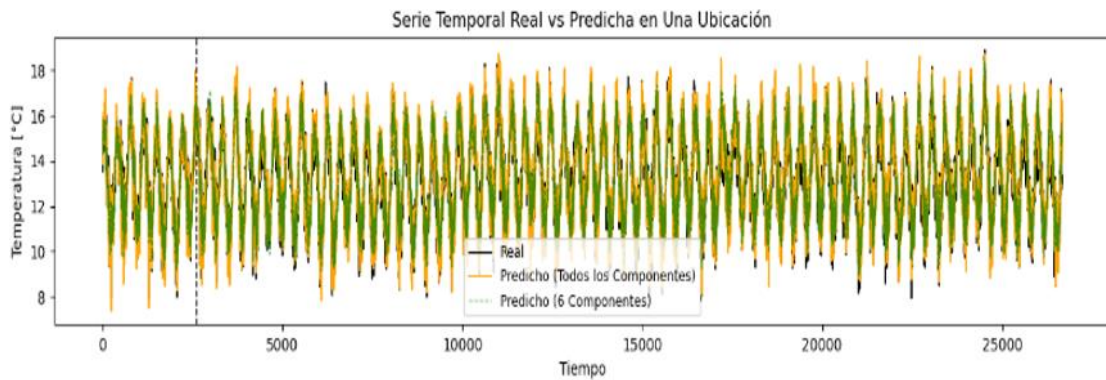


Fig. 15 – Gráfico Serie Temporal real vs Predicha en una ubicación. Elaboración propia

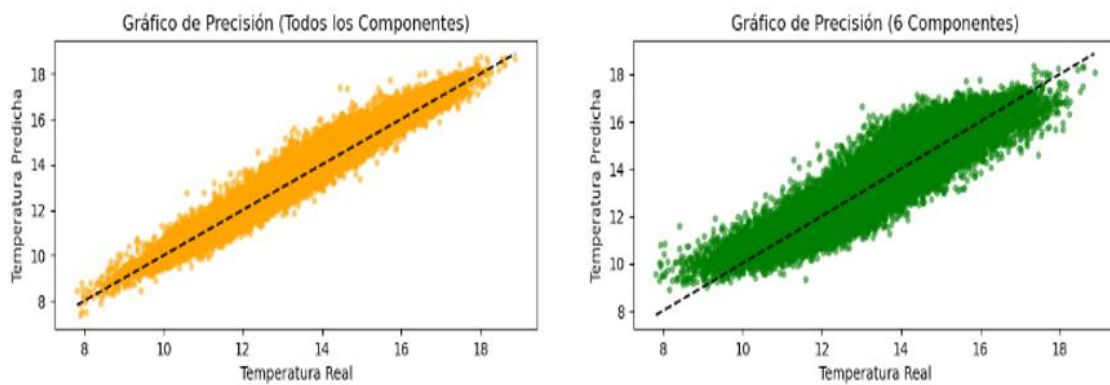


Fig. 16 – Gráficos de Precisión con distinto número de componentes en una ubicación. Elaboración propia

En conjunto, estos resultados siguen respaldando la adaptabilidad y precisión del modelo de predicción espacial y temporal para el análisis de la temperatura media en Chile (Figuras 15 y 16). El modelo no solo mantiene su precisión al ser aplicado a una nueva variable climática

Data set de temperatura Mínima

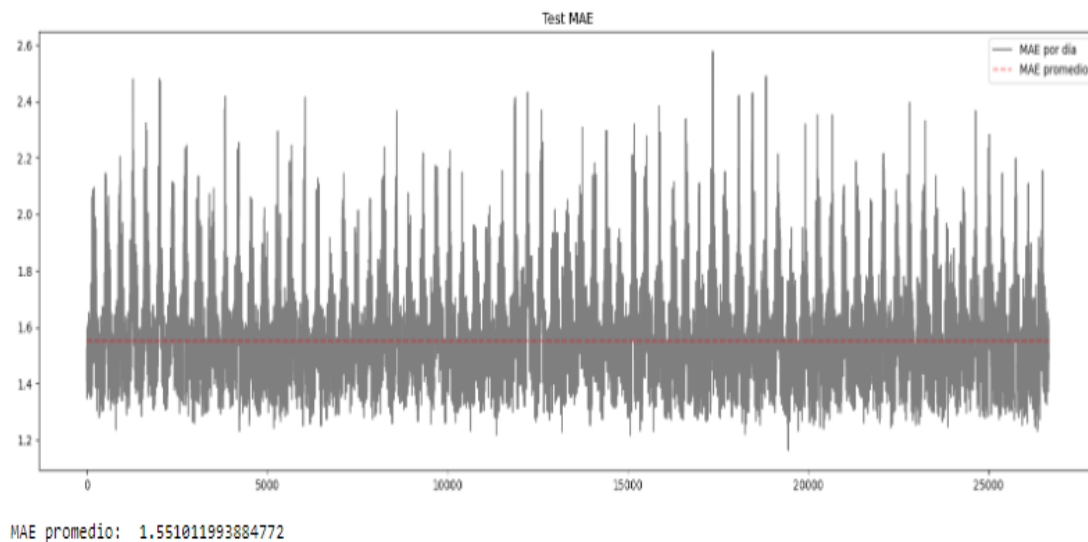
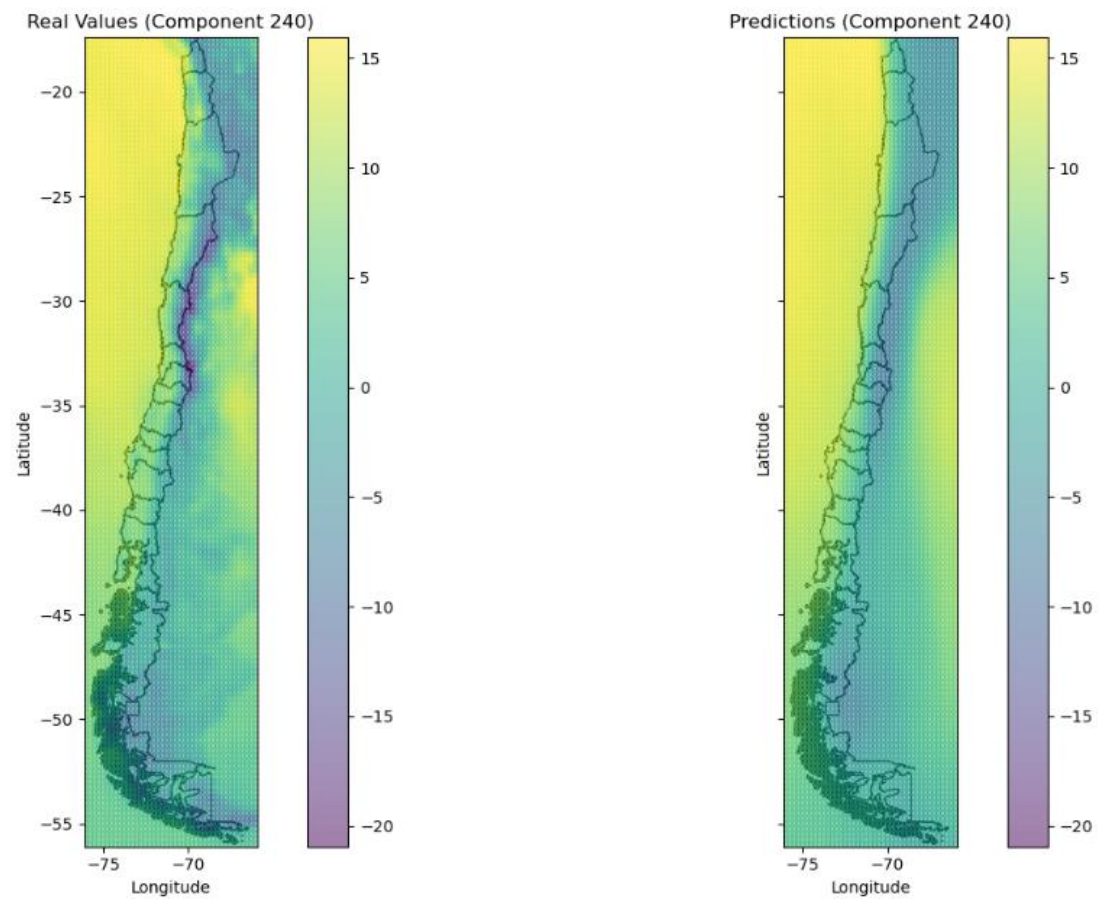


Fig. 17 – Gráfico Error Absoluto Medio (MAE). Elaboración propia

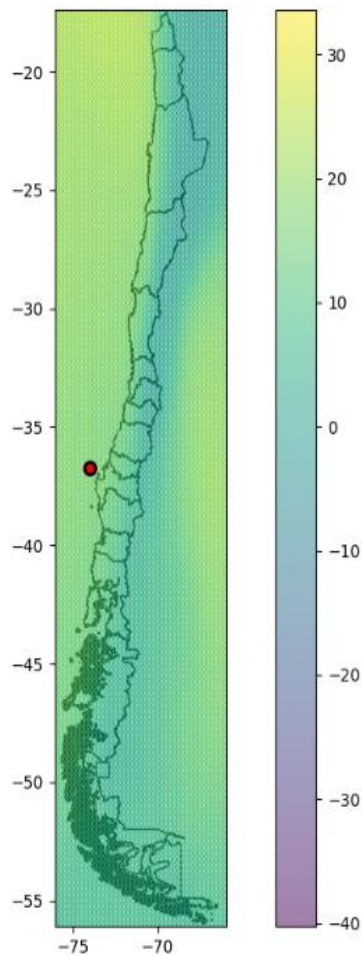
La Figura 17 muestra el error absoluto medio (MAE) de las predicciones de temperatura mínima diarias. A pesar de la variabilidad, la línea de MAE promedio subraya una consistencia notable en la precisión del modelo a lo largo del tiempo, reflejando la solidez del método aplicado incluso cuando se enfrenta a la volatilidad de las temperaturas mínimas.



*Fig. 18 – Mapa comparativo valores reales y predicción (240) para las temperaturas mínimas.
Elaboración propia*

En la Figura 18, los mapas de calor paralelos de los valores reales y las predicciones evidencian una alta correspondencia espacial, indicando que el modelo captura con eficacia la distribución geográfica de las temperaturas mínimas. La coherencia entre los patrones reales y predichos es una prueba clara de la habilidad del modelo para generalizar y aplicarse a diferentes aspectos de la variabilidad climática.

Mapa Predicho (Todos los Componentes)



Mapa Predicho (6 Componentes)

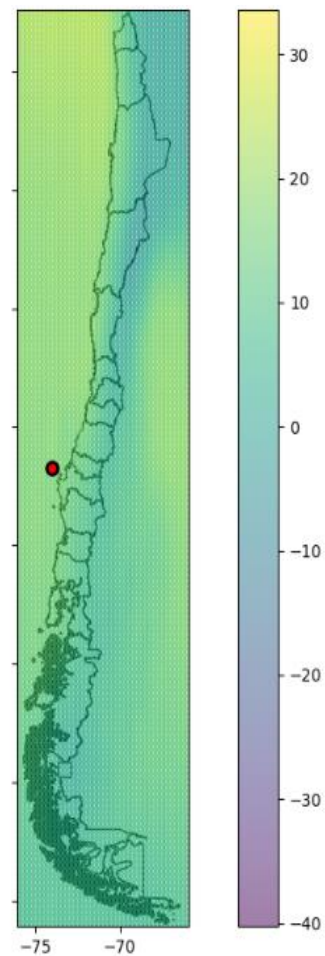


Fig. 19 – Mapa comparativo valores predichos (todos) y predicho (6). Elaboración propia

La Figura 19, se compara las predicciones basadas en todos los componentes con las que usan una selección reducida de 6 componentes, sugiere que el modelo mantiene su efectividad incluso cuando se simplifica la complejidad de los datos. Esto indica que los componentes más influyentes en la variabilidad de la temperatura mínima han sido capturados y que el modelo puede operar con eficiencia sin comprometer la precisión.

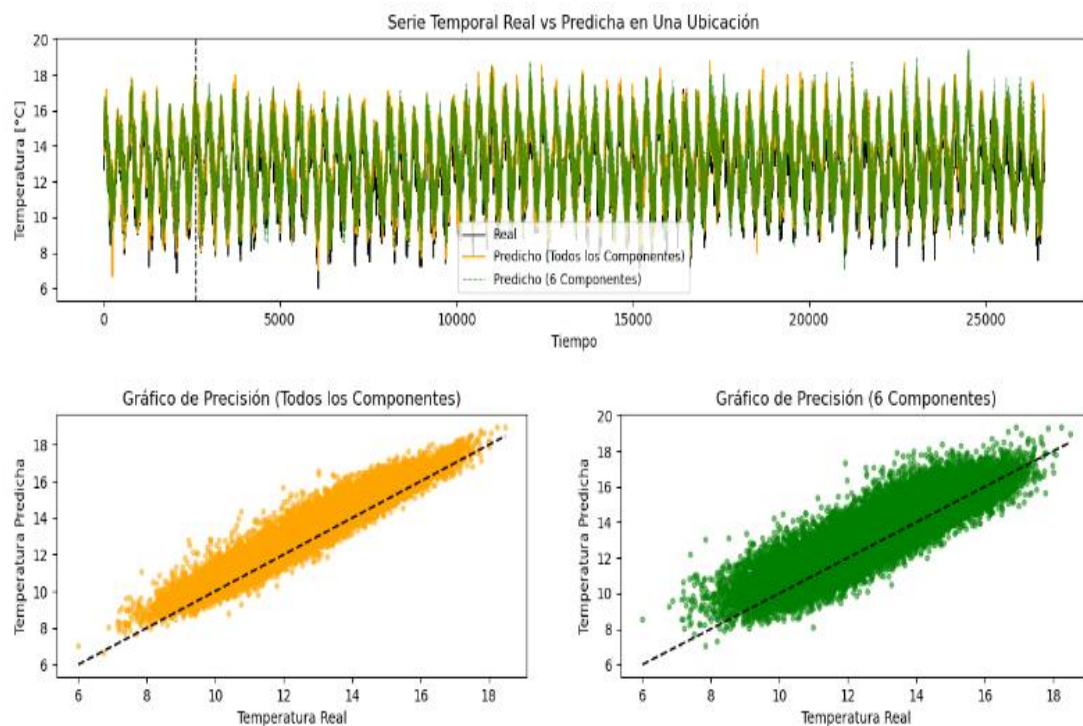


Fig. 20 – Gráficos de Serie Temporal (real y predicha) y de Precisión. Elaboración propia

Finalmente, la Figura 20 presenta una serie temporal detallada junto con gráficos de dispersión para ambas configuraciones del modelo. Las series temporales demuestran que el modelo sigue de cerca las fluctuaciones reales de la temperatura mínima, mientras que los gráficos de dispersión muestran una fuerte correlación entre las temperaturas reales y las predichas, confirmando la precisión del modelo en una dimensión cuantitativa.

Datos de Precipitación

El modelo que demostró ser eficaz para la temperatura máxima, media y mínima se ha ajustado para analizar la variable de precipitación acumulada diaria en Chile. Los resultados son indicativos de la habilidad del modelo para manejar y predecir esta variable ambiental compleja y heterogénea. A continuación, se presenta un análisis de los resultados obtenidos:

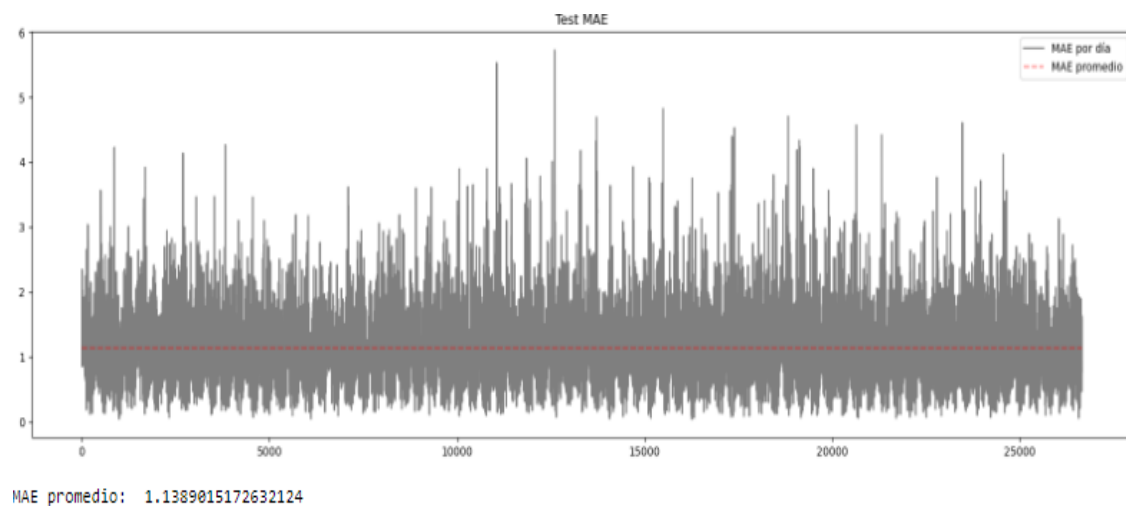


Fig. 21 – Gráfico MAE – Precipitaciones. Elaboración propia

La Figura 21 muestra el gráfico de MAE para las predicciones diarias de precipitaciones acumulada. Se observa una dispersión de errores a lo largo del tiempo, pero la línea de MAE promedio indica que el modelo ha logrado una cierta consistencia en sus predicciones, a pesar de las fluctuaciones inherentes a las precipitaciones diarias.

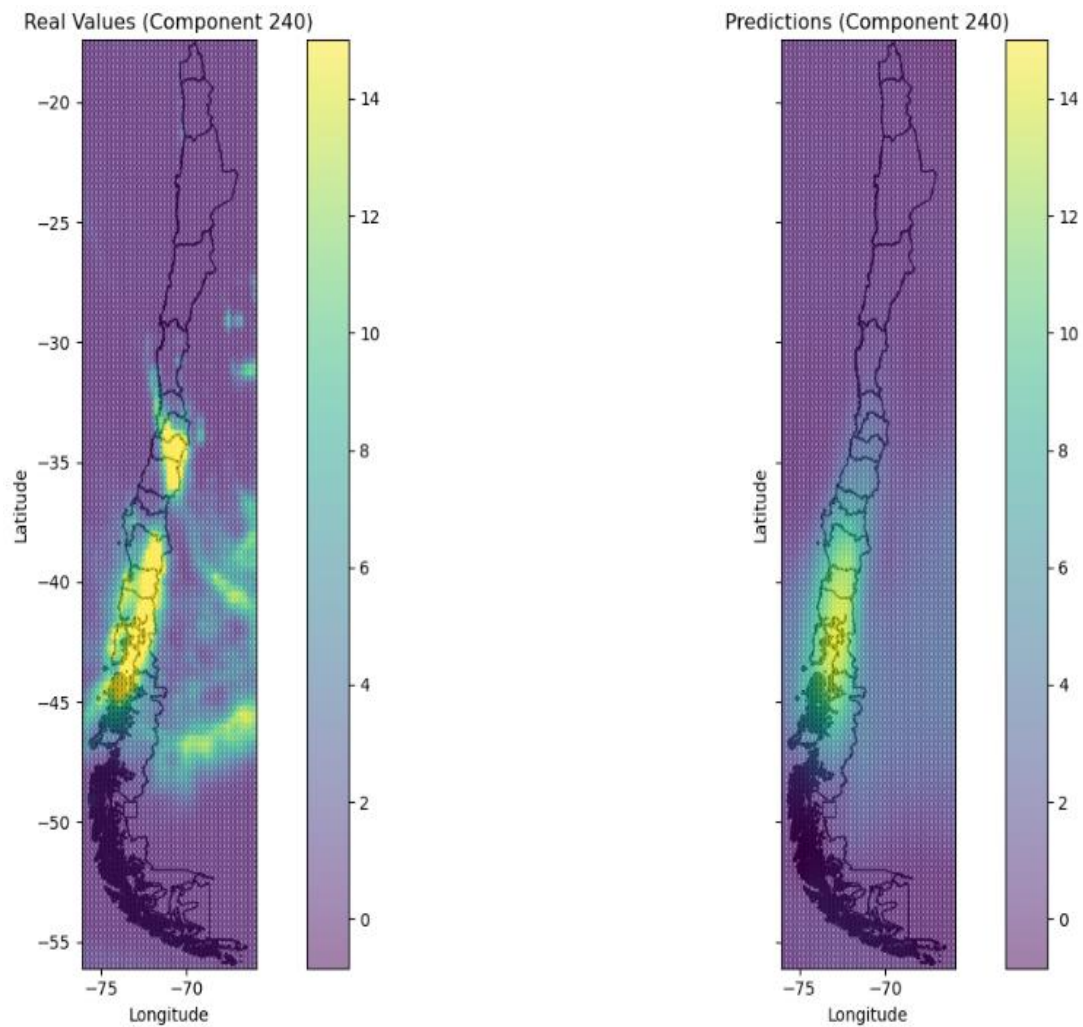


Fig. 22 – Mapas con los valores reales (izquierda) y las predicciones (derecha) de precipitación acumulada (29-08-1950). Elaboración propia

La Figura 22 presenta mapas de calor de los valores reales y las predicciones de precipitación acumulada, ofreciendo una visualización comparativa de la precisión espacial del modelo. Aunque existen variaciones entre los datos reales y las predicciones, el modelo parece capturar los patrones generales de precipitación, lo que sugiere una interpretación adecuada de la dinámica espacial de las precipitaciones. Sin embargo, aparecen mayores discrepancias en este caso en relación a las otras variables (t_{max} , t_{min} , t_{mean}). Una forma de mejorar la precisión podría ser incorporando la variable “altura” relacionada a las coordenadas geográficas.

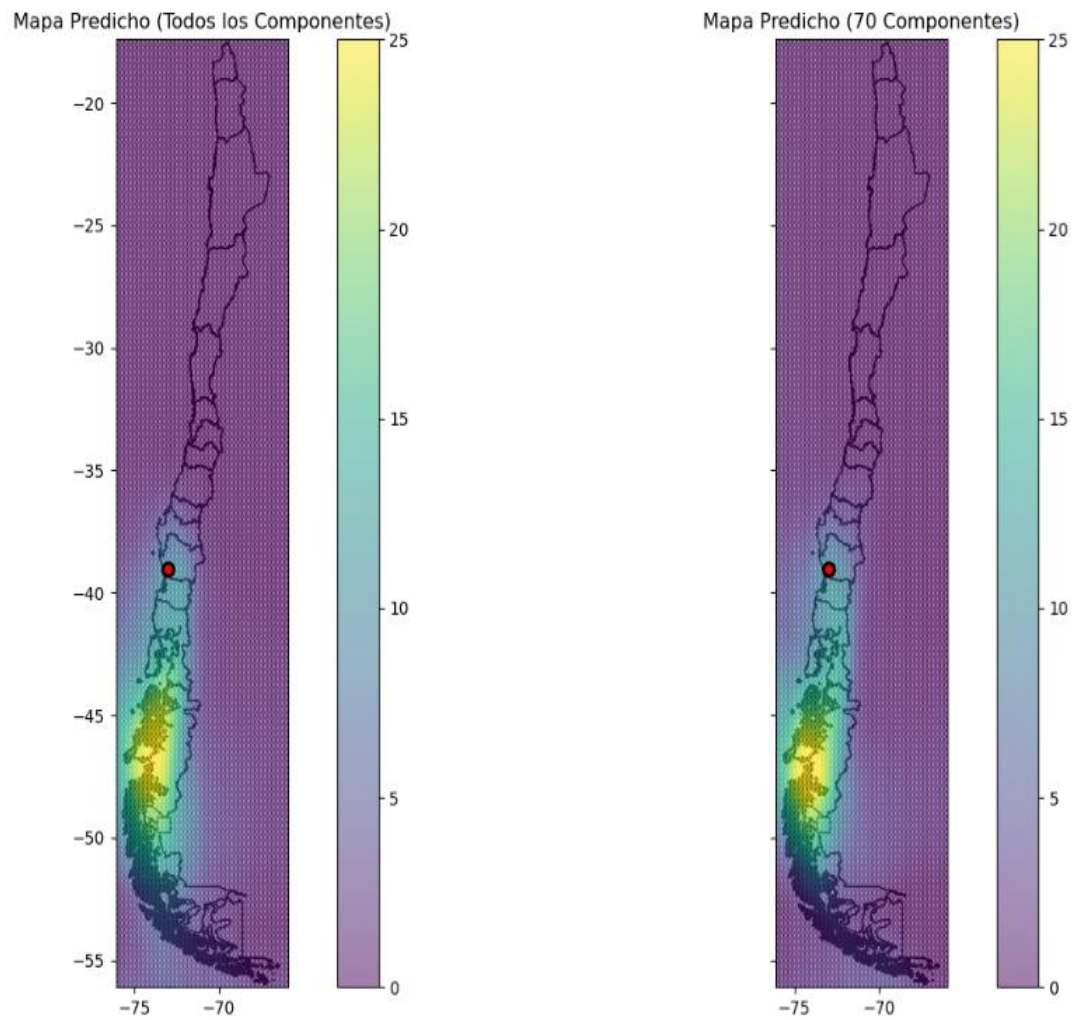


Fig. 23 – Mapas de calor de los valores predicho (todos los componentes) y los predichos (70) de precipitaciones acumulada. Elaboración propia

En la Figura 23, se compara el rendimiento del modelo utilizando todos los componentes con uno que se restringe a 70 componentes principales. La inclusión del punto rojo indica una localización geográfica específica para el análisis. Esta comparativa es crucial para entender la influencia de la dimensionalidad en la predicción de las precipitaciones acumuladas. Los resultados indican que el modelo simplificado sigue siendo efectivo, lo que destaca la robustez de las técnicas empleadas y la captura efectiva de los componentes principales que influyen las precipitaciones.

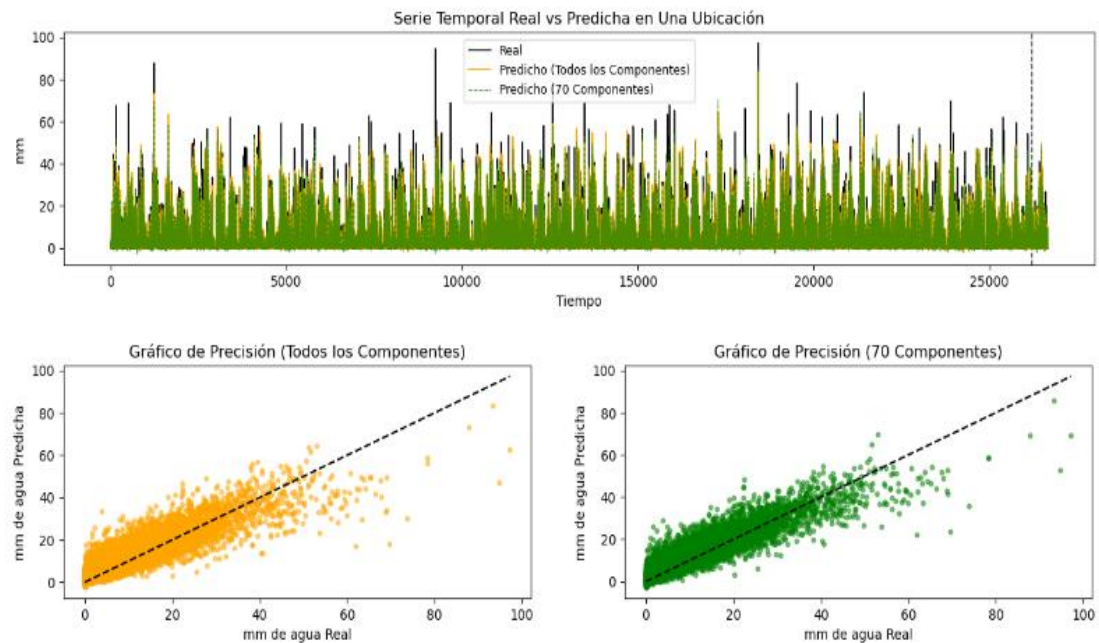


Fig. 24 – Gráficos de Serie Temporal (real y predicha) y de Precisión. Elaboración propia

Finalmente, la Figura 24, que incluye series temporales y gráficos de dispersión, muestra que las predicciones del modelo se alinean estrechamente con los valores reales a lo largo del tiempo, y existe una correlación lineal fuerte entre los datos reales y predichos. Esto demuestra la habilidad del modelo para predecir con precisión las precipitaciones acumuladas en una base diaria, validando su utilidad práctica en la interpretación de este fenómeno climático.

La adaptación del modelo para predecir la precipitación acumulada en Chile ilustra su capacidad para ser aplicado en distintas variables climáticas, manteniendo un alto nivel de precisión y ofreciendo una herramienta valiosa para la comprensión y la gestión de los recursos hídricos en el contexto del cambio climático y la variabilidad ambiental.

Data set de Evapotranspiración

Para finalizar, se analiza la evapotranspiración en Chile, adaptado de las mismas técnicas previamente aplicadas con éxito a las diferentes variables de temperatura y la precipitación, ha generado resultados significativos en cuanto a la variable de evapotranspiración diaria, medida en mm de agua equivalente.

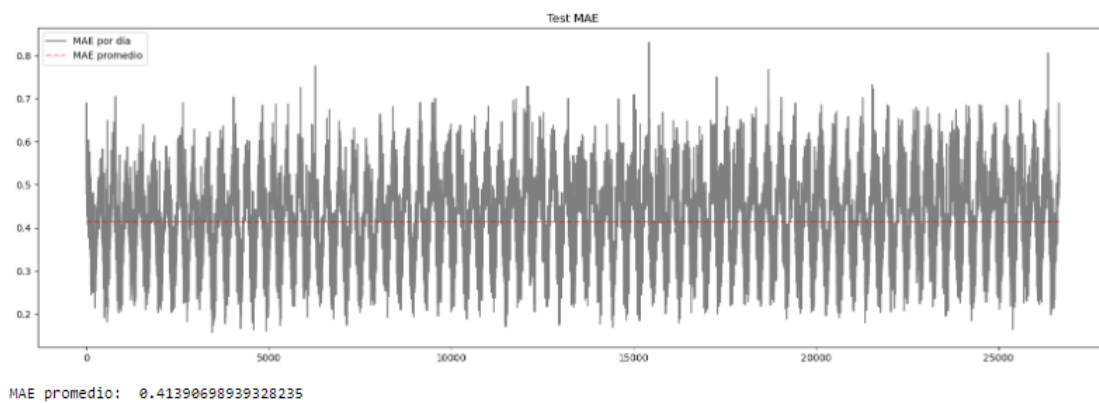
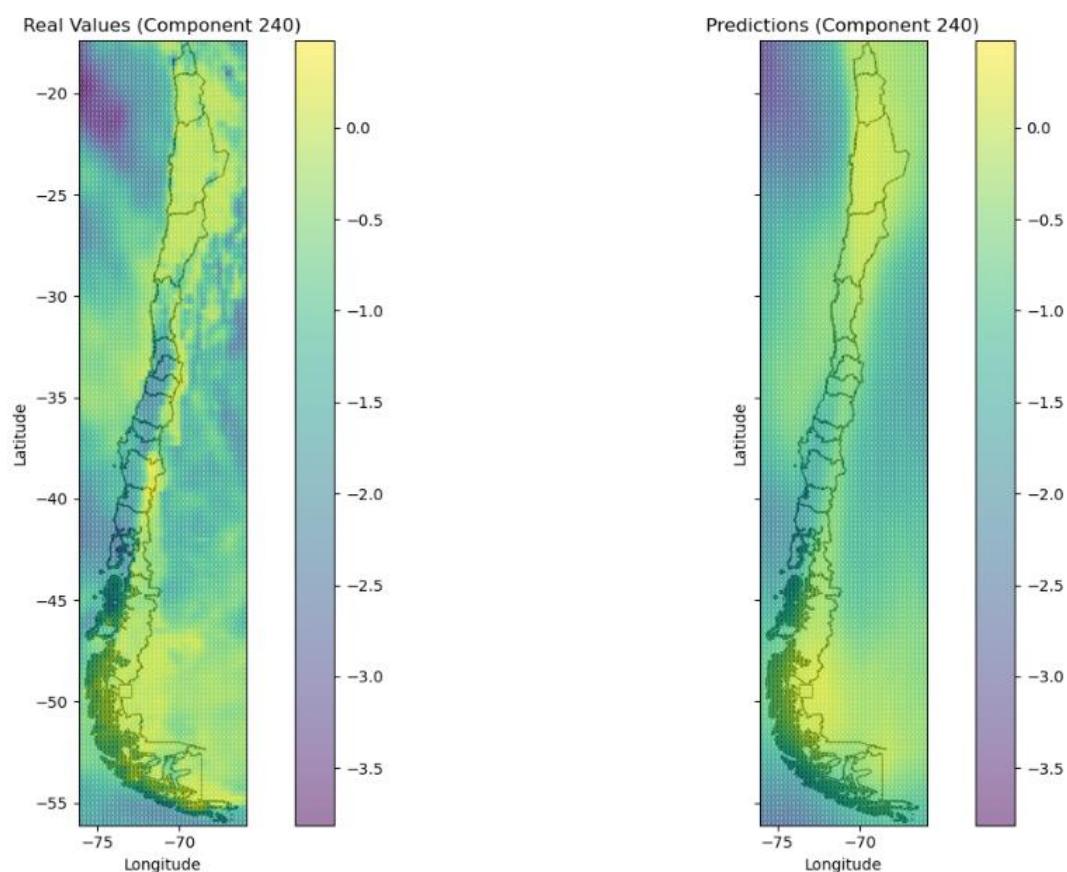


Fig. 25 – Gráfico MAE – Evapotranspiración. Elaboración propia

La Figura 25, que exhibe el error absoluto medio (MAE) diario, indica que el modelo tiene una consistencia notable en sus predicciones, con un MAE promedio que destaca su capacidad para predecir esta compleja variable ambiental con un nivel de error reducido y uniforme.



*Fig. 26 – Mapas de calor de los valores reales y las predicciones - Evapotranspiración (240).
Elaboración propia*

Los mapas de calor de la Figura 26 comparan los valores reales de evapotranspiración con las predicciones del modelo, ilustrando la destreza del modelo para simular la distribución espacial de este proceso. La similitud entre los patrones observados y los predichos señala que el modelo es capaz de capturar las tendencias clave en la dinámica de la evapotranspiración a lo largo del territorio chileno.

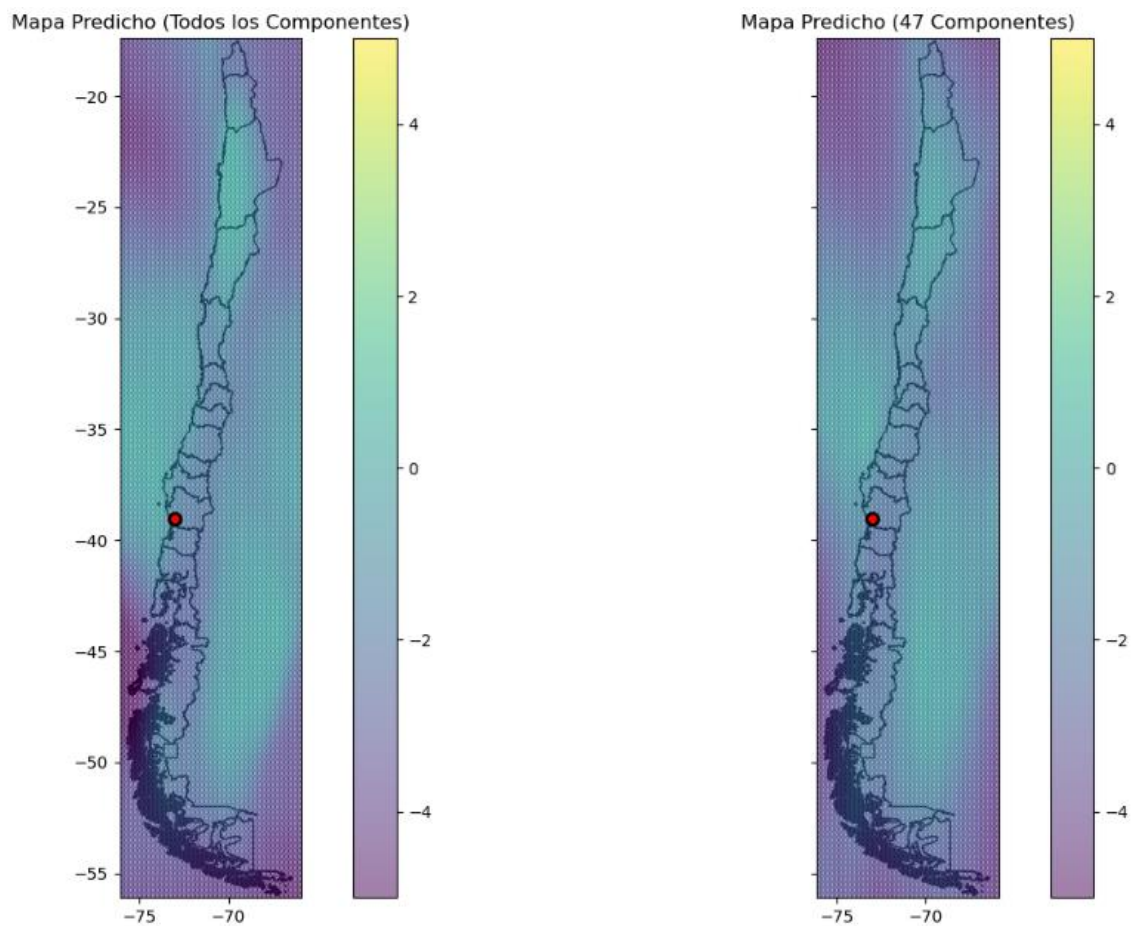


Fig. 27 – Mapas de calor de los valores predicho (todos) y los predichos (47) - Evapotranspiración. Elaboración propia

En la Figura 27, la comparación de los resultados del modelo utilizando el total de componentes y una versión simplificada con 47 componentes principales demuestra que la reducción de la complejidad de los datos no altera significativamente la exactitud de las predicciones. Esto sugiere que los componentes principales retenidos son suficientes para capturar la información esencial que influye en la evapotranspiración.

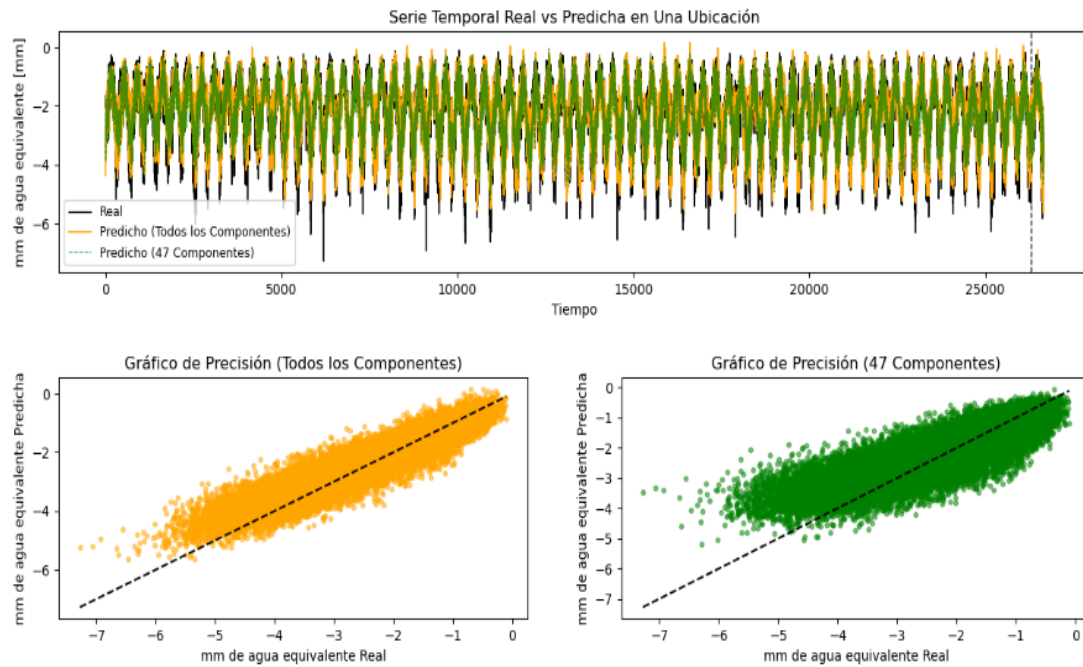


Fig. 28 – Gráficos de Serie Temporal (real y predicha) y de Precisión. Elaboración propia

La Figura 28, que incluye series temporales y gráficos de dispersión para ambas versiones del modelo, muestra una alineación estrecha entre las predicciones y los valores reales. Los gráficos de dispersión, en particular, revelan una correlación lineal entre las predicciones y las mediciones reales, confirmando la capacidad del modelo para predecir la evapotranspiración con un alto grado de precisión.

En conjunto, estos resultados indican que el modelo es efectivo para predecir la evapotranspiración en Chile, ofreciendo una herramienta valiosa para la gestión de recursos hídricos y la planificación agrícola, particularmente en el contexto del cambio climático donde la evapotranspiración juega un papel crítico en el ciclo del agua.

6. Conclusión

Con el presente estudio, siguiendo la metodología innovadora de Federico Amato, se espera marcar un hito significativo en el campo de la investigación ambiental, especialmente en la predicción climática. La diversidad geográfica y climática de Chile, con sus extremos de desiertos áridos y lluviosos y frías zonas del sur, ha proporcionado un escenario ideal para validar la eficacia de modelos avanzados en la captura y predicción de la complejidad inherente a los patrones climáticos.

La implementación y adaptación de un modelo de aprendizaje profundo para el análisis de datos climáticos chilenos ha permitido descifrar con precisión la estructura espaciotemporal de estos datos, destacando la eficacia del modelo en la representación de variaciones climáticas. Este enfoque, que ha descompuesto los procesos climáticos en funciones de base temporales y coeficientes espaciales estocásticos, ha probado ser excepcionalmente eficaz para captar las dinámicas subyacentes y realizar predicciones precisas, incluso en localizaciones sin mediciones directas.

La aplicación de técnicas como la Descomposición en Valores Singulares (SVD) ha jugado un papel crucial en la simplificación de la complejidad de los datos climáticos, resaltando patrones dominantes fundamentales para entender la variabilidad climática. El descubrimiento de que un número reducido de componentes puede explicar una gran proporción de la varianza acumulada subraya la eficiencia de la metodología empleada, permitiendo una identificación más acertada de los aspectos más significativos de los datos climáticos.

En cuanto al cumplimiento de los objetivos específicos:

Análisis de la Distribución Espacial y Temporal de Variables Climáticas: Se realizó un análisis detallado de las variables climáticas en Chile, identificando patrones climáticos distintivos. Este análisis ha permitido una comprensión más profunda de la dinámica climática regional, abordando no solo las temperaturas máximas sino también otras variables como la temperatura media, la temperatura mínima, la evaporación y la precipitación.

Preparación y Adaptación de Datos Climáticos Históricos: Los datos climáticos históricos fueron cuidadosamente preparados y adaptados para su análisis mediante técnicas avanzadas de Deep Learning y EOFs. Este proceso incluyó la transformación y estructuración de conjuntos de datos complejos en DataFrames estructurados, estableciendo una base sólida para el análisis posterior.

Implementación y Validación de un Modelo de Deep Learning con EOFs: Se implementó y validó con éxito un modelo de Deep Learning que incorpora EOFs, logrando una predicción climática precisa. Este modelo mostró un bajo error absoluto medio (MAE), lo que confirma su robustez y capacidad para generalizar a nuevos datos.

Comparación del Rendimiento Predictivo de Modelos de Deep Learning: Se comparó el rendimiento predictivo de un modelo completo de Deep Learning con un modelo simplificado basado en componentes principales. Este análisis reveló que la precisión se mantiene elevada incluso con una representación reducida de los datos, destacando la capacidad del modelo para capturar la esencia de la variabilidad climática.

Recomendaciones para futuras investigaciones:

Se sugiere la expansión del estudio para incluir un espectro más amplio de variables climáticas, el desarrollo de herramientas interactivas de visualización y análisis predictivo, y la integración de estos hallazgos con modelos climáticos globales. Estas direcciones ofrecen una visión más holística de los sistemas climáticos y podrían proporcionar una perspectiva más completa de las dinámicas climáticas a nivel mundial.

A pesar de los avances significativos, el estudio no está exento de limitaciones. Una destacable, es la notable dependencia de conjuntos de datos específicos para Chile, lo que podría restringir la aplicabilidad y generalización de los resultados a otros contextos geográficos y climáticos. Además, el modelo podría enfrentar desafíos al representar eventos climáticos extremos, dada su complejidad y variabilidad.

En conclusión, este estudio contribuye significativamente al campo de la ciencia climática, apuntando hacia el desarrollo de estrategias más resilientes y adaptativas para enfrentar los desafíos climáticos futuros. Los resultados obtenidos establecen un precedente importante para la utilización de técnicas avanzadas de Deep Learning en la modelización climática, ofreciendo una metodología valiosa que podría ser aplicada en otras regiones con características climáticas diversas. La incorporación de otras variables tales como la altura, principalmente para las precipitaciones, podría suponer una instancia que fortalezca la predicción. De igual forma, la realización de un sistema híbrido que modifique la metodología actual, en cual la parte temporal también sea tratada con una red neuronal, así como incorporar algún otro tipo de redes para el componente espacial, podría ser una instancia innovadora que no se ha estudiado actualmente.

7. Bibliografía

1. Amato, F., Guignard, F., Robert, S., & Kanevski, M. (2020). *A novel framework for spatio-temporal prediction of environmental data using deep learning*. Scientific Reports, 10, 22243. <https://doi.org/10.1038/s41598-020-79148-7>
2. Jones, N. (2017). *How machine learning could help to improve climate forecasts*. Nature, 548. <https://www.nature.com/articles/548379a>
3. Runge, J., et al. (2019). *Inferring causation from time series in earth system sciences*. Nat. Commun., 10, 1-13. <https://www.nature.com/articles/s41467-019-10105-3>
4. Rolnick, D., et al. (2019). *Tackling climate change with machine learning*. arXiv:1906.05433. <https://arxiv.org/pdf/1906.05433.pdf>
5. Reichstein, M., et al. (2019). *Deep learning and process understanding for data-driven earth system science*. Nature, 566, 195-204. <https://www.nature.com/articles/s41586-019-0912-1>
6. Shi, X., & Yeung, D.-Y. (2018). *Machine learning for spatiotemporal sequence forecasting: A survey*. arXiv preprint [arXiv:1808.06865](https://arxiv.org/pdf/1808.06865.pdf). <https://arxiv.org/pdf/1808.06865.pdf>
7. Wang, S., Cao, J., & Yu, P. S. (2019). *Deep learning for spatio-temporal data mining: A survey*. arXiv preprint arXiv:1906.04928. <https://arxiv.org/pdf/1906.04928.pdf>
8. Klemmer, K., Koshiyama, A., & Flennerhag, S. (2019). *Augmenting correlation structures in spatial data using deep generative models*. arXiv preprint arXiv:1905.09796. <https://arxiv.org/pdf/1905.09796.pdf>
9. Kadow, C., Hall, D. M., & Ulbrich, U. (2020). *Artificial intelligence reconstructs missing climate information*. Nat. Geosci., 13, 1-6. <https://www.nature.com/articles/s41561-020-0582-5>
10. McDermott, P. L., & Wikle, C. K. (2019). *Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data*. Entropy, 21, 184. <https://www.mdpi.com/1099-4300/21/2/184>
11. Fallas-Monge, J. J., Chavarría-Molina, J., & Soto-Quiros, P. (2021). *Descomposición en valores singulares de una matriz: Un repaso por los fundamentos teóricos y sus aplicaciones en el procesamiento de imágenes*. Revista Investigación Operacional, 42(2), 148-173 <https://rev-inv-ope.pantheonsorbonne.fr/sites/default/files/inline-files/42221-03.pdf>
12. Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., Ravuri, S., Ewalds, T., Alet, F., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Stott, J., Vinyals, O., Mohamed, S., & Battaglia, P. (2022). *GraphCast: Learning skillful medium-range global weather forecasting*. <https://arxiv.org/abs/2212.12794>