

**APLICACIÓN DE DOWNSCALING ESTADÍSTICO A DATOS DE
TEMPERATURA Y PRECIPITACIÓN UTILIZANDO TÉCNICAS DE
INTELIGENCIA ARTIFICIAL. CASO DE ESTUDIO: LA CUENCA DEL RÍO
DAGUA, VALLE DEL CAUCA**

ARTURO ALEXIS GRANADA GIRÓN

**UNIVERSIDAD DEL VALLE
ESCUELA DE INGENIERÍA CIVIL Y GEOMÁTICA
2025**

**APLICACIÓN DE DOWNSCALING ESTADÍSTICO A DATOS DE
TEMPERATURA Y PRECIPITACIÓN UTILIZANDO TÉCNICAS DE
INTELIGENCIA ARTIFICIAL. CASO DE ESTUDIO: LA CUENCA DEL RÍO
DAGUA, VALLE DEL CAUCA**

ARTURO ALEXIS GRANADA GIRÓN

MAURICIO EDILBERTO RINCÓN ROMERO. Ph.D.

**UNIVERSIDAD DEL VALLE
ESCUELA DE INGENIERÍA CIVIL Y GEOMÁTICA
INGENIERÍA TOPOGRÁFICA
SANTIAGO DE CALI
2025**

PÁGINA DE ACEPTACIÓN

Nota de aceptación

Rafael Á. Girón G.
Jurado

Guillermo E. Armenta P.
Jurado

Santiago de Cali, 2025

Tabla de contenido

Índice de figuras	VIII
Índice de tablas	XII
ABREVIATURAS	1
RESUMEN	3
INTRODUCCIÓN	1
1. PROBLEMA	5
1.1. Definición del problema	5
1.2. Árbol del problema	7
1.3. Pregunta de investigación	8
1.4. OBJETIVOS	8
1.4.1. Árbol de objetivos	8
1.4.2. Objetivo general	8
1.4.3. Objetivos específicos	9
1.5. JUSTIFICACIÓN	9
2. REVISIÓN DE LITERATURA	11
2.1. MARCO CONCEPTUAL	11
2.2. MARCO TEÓRICO	18
2.2.1. Análisis de componentes principales	19
2.2.2. Interpolación bicúbica	20
2.2.3. Regresión lineal	21
2.2.4. Random forest	21
2.2.5. Máquina de soporte de vectores	22
2.2.6. Redes neuronales artificiales	24
2.2.7. Medidas de desempeño y métricas de validación	29
2.2.7.1. Error cuadrático medio (ECM)	30
2.2.7.2. Raíz del error cuadrático medio (RECM)	30
2.2.7.3. Error absoluto medio (EAM)	30
2.2.7.4. Error porcentual absoluto medio (EPAM)	31

2.2.7.5.	Coeficiente de correlación de Pearson (r)	31
2.2.7.6.	Coeficiente de determinación (R^2)	32
2.2.7.7.	Coeficiente de correlación de Spearman (S)	32
2.2.7.8.	Diagrama de Taylor	33
2.3.	MARCO REFERENCIAL	33
3.	METODOLOGÍA	38
3.1.	Localización	39
3.2.	Datos utilizados	39
3.2.1.	GFDL Earth System Model version 4.1 (GFDL-ESM4)	40
3.2.2.	Fifth Generation European ReAnalysis (ERA5-Land)	42
3.2.3.	Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM)	43
3.2.4.	Datos de elevación	44
3.3.	Marco experimental	45
3.3.1.	Horizonte de tiempo	46
3.3.2.	Análisis exploratorio y descriptivo de datos climatológicos	46
3.3.2.1.	Distribución espacio-temporal homogénea	48
3.3.2.2.	Correlación y descripción de los datos	48
3.3.2.3.	Limpieza de datos, datos faltantes y atípicos	49
3.3.3.	Componentes principales	50
3.3.4.	Escalamiento de datos	50
3.3.4.1.	Estandarización en la temperatura	51
3.3.4.2.	Normalización en la precipitación	51
3.3.5.	<i>Downscaling</i> estadístico	51
3.3.5.1.	Interpolación bicúbica	52
3.3.5.2.	Partición de los datos para el entrenamiento y prueba de los modelos	53
3.3.5.3.	Selección de variables de entrenamiento y objetivo	54
3.3.5.4.	Regresión lineal (RL)	55
3.3.5.5.	Random forest regressor (RFR)	58
3.3.5.6.	Regresión de vectores de soporte (RVS)	59
3.3.6.	<i>Downscaling</i> estadístico con RNA	59
3.3.6.1.	Partición de los datos	61
3.3.6.2.	Validación cruzada	61
3.3.6.3.	Arquitectura del Modelo RNA	62
3.3.6.4.	Estrategia de compilación	64
3.3.6.5.	Tuneo de hiperparámetros	65
3.3.6.6.	Evaluación de los modelos	67
3.3.7.	Desempeño de los modelos de <i>downscaling</i> estadístico	68
3.4.	Plataforma y publicación	68
3.4.1.	Formato de archivo NetCDF	69
3.4.2.	Backend	70
3.4.3.	Frontend	71

4. RESULTADOS Y DISCUSIONES	73
4.1. Análisis exploratorio de los datos	74
4.1.1. Resultados y discusiones para la Temperatura	74
4.1.1.1. Correlación entre la temperatura de ERA5-Land y las mediciones locales	75
4.1.1.2. Descripción estadística de las fuentes de temperatura	77
4.1.1.3. Correlación entre la temperatura de GFDL-ESM4, ERA5-Land y las observaciones locales	80
4.1.2. Resultados y discusiones para la precipitación	81
4.1.2.1. Descripción estadística de las fuentes de precipitación	82
4.1.2.2. Correlación entre la precipitación de ERA5-Land y las mediciones locales	88
4.1.2.3. Comparativo entre la precipitación de GFDL-ESM4, ERA5-Land y las observaciones locales	89
4.1.3. Variabilidad climática por hora	90
4.1.4. Componentes principales	93
4.2. <i>Downscaling</i> estadístico para datos de temperatura y precipitación	95
4.2.1. Regresión Lineal	96
4.2.2. Random Forest	103
4.2.2.1. Resultados Random forest para la temperatura	103
4.2.2.2. Resultados Random forest para la precipitación	105
4.2.3. Regresión de vectores de soporte	107
4.2.3.1. Resultados Regresión de vectores de soporte para temperatura	107
4.2.3.2. Resultados Regresión de vectores de soporte para precipitación	108
4.2.4. Modelamiento con redes neuronales ConvLSTM2D	110
4.2.4.1. Entrenamiento del modelo ConvLSTM2D para la temperatura	110
4.2.4.2. Evaluación del modelo ConvLSTM2D para la temperatura	111
4.2.4.3. Entrenamiento del modelo ConvLSTM2D para la precipitación	115
4.2.4.4. Evaluación del modelo ConvLSTM2D para la precipitación	116
4.2.4.5. Comparación de la modelación ConvLSTM2D de los datos de temperatura y precipitación	119
4.3. Validación del modelo ConvLSTM2D para el <i>Downscaling</i> estadístico	123
4.3.1. Validación estación "La Cumbre"	125
4.3.2. Validación estación "Farallones"	126
4.3.3. Validación estación "Unipacífico"	128
4.3.4. Análisis de proyecciones climáticas mediante el <i>downscaling</i> estadístico con redes ConvLSTM2D	130
4.3.4.1. Proyecciones climáticas para el escenario SSP2-4.5	130
4.4. Plataforma web	135
4.4.1. Visor geográfico	135
4.4.1.1. Control de proyecciones	135
4.4.1.2. Control temporal	136
4.4.2. Descarga de archivos NetCDF	136
4.4.3. API	137

5. Conclusiones	140
BIBLIOGRAFÍA	143

Índice de figuras

2.1.	Temperatura promedio diaria cerca a la superficie a nivel global del Modelo GFDL-ESM4 de la NOAA. Datos históricos con una resolución espacial por píxel de 1°.	16
2.2.	Precipitación promedio diaria a nivel mundial del modelo GFDL- ESM4 con una resolución espacial por píxel de 1°.	17
2.3.	Diagrama simple de un Random Forest	22
2.4.	Gráfico del comportamiento de los Kernel por modelo de SVR	23
2.5.	Representación del modelo perceptrón	25
2.6.	Modelo de una red neuronal artificial	26
2.7.	Función lineal	27
2.8.	Función ReLU	27
2.9.	Función Sigmoide bipolar	27
2.10.	Función Sigmoide	27
2.11.	Arquitectura de una red neuronal LSTM	28
2.12.	Diagrama de Taylor	34
3.1.	Diagrama de flujo de la metodología desarrollada.	38
3.2.	Región Pacífica del departamento del Valle del Cauca, ubicación de la cuenca del río Dagua.	40
3.3.	Temperatura y precipitación del GCM GFDL-ESM4 para la región Pacífica del departamento del Valle del Cauca. Resolución espacial de 1° y resolución temporal de 3 horas.	41
3.4.	Temperatura promedio y precipitación acumulada de ERA5-Land, para el mes de mayo de 2015 en el Pacífico del departamento del Valle del Cauca. Datos con una resolución espacial de 0.1° y temporal de 1 hora.	43
3.5.	Mapa de distribución espacial de las estaciones meteorológicas automáticas en la zona de estudio, superpuesto sobre el MDE de Alos-Palsar. La escala de colores representa la altitud del terreno, desde 0 metros, hasta más de 3500 metros. Las estaciones, distribuidas capturar la variabilidad climática altitudinal de la región. Los datos climatológicos registrados por estas estaciones son con frecuencia horaria.	45
3.6.	Diagrama del procedimiento desarrollado en el análisis exploratorio y descriptivo de los datos y diferentes variables utilizadas.	47
3.7.	Diagrama metodológico del proceso del <i>Downscaling</i> estadístico	52

3.8. Armonización de la resolución espacial del modelo GFDL-ESM4 a resolución espacial de datos ERA5-Land utilizando el método de interpolación bicúbica. . . .	53
3.9. Visualización de la partición de datos en conjuntos de entrenamiento y prueba para los modelos RL, RF y MSV	53
3.10. Visualización de las variables de entrada, modelos aplicados y variables respuesta. .	55
3.11. Vector de entrada con 8 secuencias de modelo GCM a la red ConvLSTM2D para realizar predicción de 3 pasos en el tiempo con los valores correspondiente a ERA5-Land.	60
3.12. División tripartita del set completo de datos.	61
3.13. Particiones de 5-Fold para la validación cruzada en el set de datos.	62
3.14. Arquitectura del modelo de Downscaling estadístico utilizando diferentes capas de RNA, y bloques residuales.	64
3.15. Metodología y herramientas utilizadas para la creación de la plataforma digital. . .	69
4.1. (A) Serie temporal horaria que muestra el patrón de variabilidad de la temperatura observada (rojo) entre la estación "Farallones", "La Cumbre", "Unipacífico" y el valor del píxel correspondiente del producto ERA5-Land (azul). (B) Diagrama de dispersión que ilustra la correlación entre las observaciones in situ y los valores del producto ERA5-Land.	76
4.2. Distribución y diagrama de cajas de la temperatura entre los valores observados por la estación 'Farallones', ERA5-Land y GFDL-ESM4.	79
4.3. Distribución y diagrama de cajas de la temperatura entre los valores observados por la estación 'La Cumbre', ERA5-Land y GFDL-ESM4.	79
4.4. Distribución y diagrama de cajas de la temperatura entre los valores observados por la estación 'Unipacífico', ERA5-Land y GFDL-ESM4.	80
4.5. Distribución estadística mensual multianual de la precipitación según escenarios SSP del modelo GFDL-ESM4 (2015-2023)	84
4.6. Distribución mensual de precipitación total (mm/mes) según ERA5-Land para los píxeles correspondientes a las estaciones meteorológicas (2015-2023).	85
4.7. Comparación de la distribución total mensual (mm/mes) de precipitación entre datos observados y ERA5-Land para la estación "La Cumbre".	86
4.8. Comparación de la distribución total mensual (mm/mes) de precipitación entre datos observados y ERA5-Land para la estación "Farallones".	86
4.9. Comparación de la distribución total mensual (mm/mes) de precipitación entre datos observados y ERA5-Land para la estación "Unifacífico".	87
4.10. Precipitación total y temperatura promedio por hora entre el año 2015 y 2023. Datos estimados por ERA5-Land y las mediciones realizadas por la estación "La Cumbre"	91
4.11. Precipitación total y temperatura promedio por hora entre el año 2015 y 2023. Datos estimados por ERA5-Land y las mediciones realizadas por la estación "Farallones"	92
4.12. Precipitación total y temperatura promedio por hora entre el año 2015 y 2023. Datos estimados por ERA5-Land y las mediciones realizadas por la estación "Unipacífico".	93

4.13. Distribución de la varianza explicada por componentes principales de las variables climáticas del GCM.	94
4.14. Resultados del modelo RLM para los valores de temperatura. El resumen detalla el ajuste del modelo, coeficientes y significancia estadística.	97
4.15. Histograma de los residuos de la temperatura.	98
4.16. Q-Q Plot de los residuos del modelo RL para la temperatura.	99
4.17. Resultados del modelo RLM para los valores de precipitación. El resumen detalla el ajuste del modelo, coeficientes y significancia estadística.	100
4.18. Histograma de los residuos de la precipitación.	101
4.19. Q-Q Plot de los residuos del modelo RL para la precipitación.	102
4.20. Evolución de las curvas de error durante el entrenamiento de la red ConvLSTM2D para la predicción de la temperatura. Métricas de evaluación RMSE a la izquierda y EAM a la derecha.	111
4.21. Mapa de distribución espacial del error cuadrático medio (ECM) por píxel para las predicciones del modelo ConvLSTM2D para el <i>downscaling</i> estadístico de la temperatura con la partición de prueba.	113
4.22. Mapa de distribución de correlación de Pearson por píxel para las predicciones del modelo ConvLSTM2D para el <i>downscaling</i> estadístico de la temperatura con la partición de prueba.	114
4.23. Evolución de las curvas de error durante el entrenamiento de la red ConvLSTM2D para la predicción de la precipitación. Métricas de evaluación RMSE a la izquierda y EAM a la derecha.	116
4.24. Mapa de la distribución del error cuadrático medio de la precipitación con el set de prueba.	118
4.25. Mapa de la distribución de correlación de Pearson para la precipitación con el set de prueba.	119
4.26. Patrón de la temperatura entre las observaciones de la estación "La cumbre", la estimación de ERA5-Land y las predicciones de <i>Downscaling</i> estadístico desarrollado.120	
4.27. Patrón de la temperatura entre las observaciones de la estación "Farallones", la estimación de ERA5-Land y las predicciones de <i>Downscaling</i> estadístico desarrollado.120	
4.28. Patrón de la temperatura entre las observaciones de la estación "Unipacífico", la estimación de ERA5-Land y las predicciones de <i>Downscaling</i> estadístico desarrollado.121	
4.29. Patrón de la precipitación acumulada por mes entre las observaciones de la estación "La Cumbre", la estimación de ERA5-Land y las predicciones de <i>Downscaling</i> estadístico desarrollado.	121
4.30. Patrón de la precipitación acumulada por mes entre las observaciones de la estación "Farallones", la estimación de ERA5-Land y las predicciones de <i>Downscaling</i> estadístico desarrollado.	122
4.31. Patrón de la precipitación acumulada por mes entre las observaciones de la estación "Unipacífico", la estimación de ERA5-Land y las predicciones de <i>Downscaling</i> estadístico desarrollado.	122
4.32. Diagrama de Taylor de los modelos de temperatura, evaluado con los datos medidos por la estación meteorológica "La Cumbre".	126

4.33. Diagrama de Taylor de los modelos de temperatura, evaluado con los datos medidos por la estación meteorológica "Farallones".	128
4.34. Diagrama de Taylor de los modelos de temperatura, evaluado con los datos medidos por la estación meteorológica "Unipacífico".	129
4.35. Distribución Temporal de la Precipitación Mensual Multianual Proyectada (2024-2035) bajo Escenario SSP2-4.5.	132
4.36. Lienzo de visualización de los datos sobre el Pacífico del departamento de Valle del Cauca.	135
4.37. Control de cada una de las variables en los diferentes escenarios presentados. . . .	136
4.38. Control que permite avanzar y retroceder en el tiempo.	136
4.39. Respuesta de la url de la API para temperatura en la hora actual de consulta. . . .	138

Índice de tablas

1.1. Árbol del problema	7
1.2. Árbol de objetivos	8
2.1. Cuadro de referencias utilizadas en la investigación, con los modelos, variables y resoluciones abordadas por cada autor.	37
3.1. Lista de variables climáticas del GCM GFDL-ESM4 utilizadas en esta investigación. Todas las variables son estimaciones cercanas a la superficie con una resolución espacial de 1° y resolución temporal de 3 horas.	42
3.2. Descripción de las variables temperatura y precipitación de ERA5-Land utilizadas en esta investigación.	43
3.3. Lista de estaciones climatológicas del IDEAM disponibles. Toma de registros de datos con frecuencia horaria.	44
3.4. Variables utilizadas como predictoras y objetivo. Las variables latitud, longitud y altura son utilizadas como la principales variables geográficas para el entrenamiento de los modelos, las componentes principales se fueron añadiendo progresivamente al entrenamiento de los modelos.	54
4.1. Tabla de coeficiente de determinación, correlación de Pearson y Spearman de la variable temperatura entre los datos observados por las estaciones y las estimaciones de ERA5-Land. Se realiza la correlación Punto de cada estación a Píxel correspondiente en ERA5-Land	75
4.2. Caracterización estadística de la temperatura (°C) para diferentes fuentes de datos en la zona de estudio. Primero se presenta los estadísticos descriptivos del Modelo Climático Global GFDL-ESM4, luego el análisis comparativo entre datos observados y ERA5-Land por estación meteorológica.	78
4.3. Se seleccionan los mejores resultados entre las correlaciones de los datos de temperatura Observados, ERA5-Land y los diferentes escenarios generados por el modelo GFDL-ESM4, lo cuales se resaltan de color verde. De este modo se selecciona el escenario que mejor se correlacione para realizar el entrenamiento del modelo.	81
4.4. Estadísticas descriptivas de precipitación mensual (mm) del modelo GFDL-ESM4, ERA5-Land y estaciones meteorológicas para el período de estudio.	83

4.5. Métricas de correlación entre precipitación horaria observada en las estaciones y los valores de precipitación total/hora (mm/h) ERA5-Land para pares estación-píxel correspondientes.	88
4.6. Análisis de métricas de correlación para la precipitación: comparación entre escenarios de GFDL-ESM4 con datos de precipitación de ERA5-Land y observaciones meteorológicas locales. Se seleccionan los mejores resultados de color verde, donde el escenario SSP 2-45 para la precipitación (pr 245) es el que mejor se correlaciona con las fuentes locales.	89
4.7. Porcentaje de varianza explicada por cada componente principal establecida.	94
4.8. Métricas de evaluación del modelo de regresión lineal para temperatura, utilizando la componente geoespacial "Geo" y las componentes principales generadas.	96
4.9. Prueba de supuestos del modelo de RL para la temperatura.	98
4.10. Métricas de evaluación del modelo de regresión lineal para precipitación utilizando datos geoespaciales "Geo" y las componentes principales generadas.	100
4.11. Prueba de supuestos del modelo de RL para la precipitación.	101
4.12. Métricas de validación del modelo Random Forest para temperatura con 10 estimadores, evaluando el impacto progresivo de las componentes principales.	104
4.13. Métricas de validación del modelo Random Forest para temperatura con configuraciones de 25, 50 y 100 estimadores utilizando 6 componentes principales.	104
4.14. Métricas de evaluación modelo Random Forest para precipitación con 10 estimadores, evaluando la incorporación progresiva de componentes principales.	106
4.15. Métricas de evaluación del modelo Random Forest para precipitación con 25, 50 y 100 estimadores utilizando 6 componentes principales.	106
4.16. Métricas de evaluación del modelo de Regresión de Vectores de Soporte para temperatura utilizando un año de datos (545720 registros)	107
4.17. Métricas de evaluación del modelo de Regresión de Vectores de Soporte para temperatura utilizando dos año de datos (1093440 registros)	108
4.18. Métricas de evaluación del modelo de Regresión de vectores de soporte con el número de registros de un año.	109
4.19. Métricas de evaluación del modelo de Regresión de vectores de soporte con el número de registros de dos años	109
4.20. Métricas de evaluación de la pérdida (loss) de entrenamiento del modelo de ConvLSTM2D para la variable temperatura, aplicado a la partición de los datos.	111
4.21. Métricas de evaluación en el desempeño del modelo ConvLSTM2D en la predicción de la temperatura: métricas de evaluación por partición del set de datos.	112
4.22. Métricas de evaluación de la pérdida (loss) de entrenamiento del modelo ConvLSTM2D para la variable precipitación, aplicado a la partición de los datos.	115
4.23. Métricas de evaluación de desempeño del modelo de ConvLSTM2D con cada partición del set de datos.	117
4.24. Mejores resultados de los modelos evaluados para el <i>downscaling</i> estadístico para la variable temperatura.	123
4.25. Mejores resultados de los modelos evaluados para el <i>downscaling</i> estadístico para la variable precipitación.	124

4.26. Métricas de validación de los modelos de Downscaling estadístico para la variable temperatura sobre la estación "La Cumbre".	125
4.27. Métricas de validación de los modelos de DE para la variable precipitación sobre la estación "La Cumbre".	126
4.28. Métricas de validación de los modelo de DE para la variable temperatura sobre la estación "Farallones".	127
4.29. Métricas de validación de los modelos de DE para la variable precipitación sobre la estación "Farallones".	127
4.30. Métricas de validación de los modelos de DE para la variable temperatura sobre la estación "Unipacífico".	128
4.31. Métricas de validación de los modelos de DE para la variable precipitación sobre la estación "Unipacífico".	129
4.32. Estadísticos Descriptivos de Precipitación Mensual multianual (mm/mes) por Estación bajo Escenario SSP2-4.5 proyectos desde el año 2024 hasta el año 2035. .	131

Abreviaturas

ACP Análisis de componentes principales

CCG Cambio Climático Global

CMIP Proyecto de Intercomparación de Modelos Acoplados

CTeI Ciencia, Tecnología e Innovación

DE Downscaling Estadístico

EAM Error Absoluto Medio

ECM Error Cuadrático Medio

ENOS El Niño Oscilación Sur

EPAM Error Porcentual Absoluto Medio

FBR Función de base radial gaussiana

GCM Modelos Circulación General

GISMODEL Grupo de Investigación en Simulación y Modelación Dinámica Espacial

IA Inteligencia Artificial

IDEAM Instituto de Hidrología, Meteorología y Estudios Ambientales

IPCC Grupo Intergubernamental para el Cambio Climático

LSTM Long Short Term Memory

r Coeficiente de correlación de Pearson

R^2 Coeficiente de determinación

RCP Trayectorias de Concentración Representativas

RECM Raíz del error cuadrático medio

RF Random Forest

RL Regresión Lineal

RNA Red Neuronal Artificial

RNC Red Neuronal Convolucional

RVS Regresión de Vectores de Soporte

S Coeficiente de correlación de Spearman

SSP Vías Socioeconómicas Compartidas

VC Variabilidad Climática

WCRP Programa Mundial de Investigación Climática

WMO World Meteorological Organization

Resumen

Este trabajo de grado dispone datos de las variables climáticas proyectadas de temperatura y precipitación que sean de utilidad para la modelación hidrológica en el grupo de investigación GISMODEL, a partir de la integración de los escenarios climáticos de los Modelos Circulación General (GCM) definidos en el IPCC6 (Grupo Intergubernamental para el Cambio Climático, 2015) usando técnicas de Inteligencia Artificial y herramientas geomáticas, a una resolución espacio-temporal más apropiada, debido a que la resolución de los MCG es bastante gruesa para la modelación a nivel local. Con base en el estado del arte tecnológico actual, se decide implementar una metodología basada en modelamiento con Redes Neuronales Artificiales (RNA) y evaluar su capacidad con técnicas de aprendizaje automático convencionales como la Regresión Lineal (RL), Random Forest (RF) y Regresión de Vectores de Soporte (RVS), con el objeto de comparar cuál de los modelos se adapta mejor a cada una de las variables de estudio. Así mismo se evalúa la precisión de cada modelo con datos reales observados, suministrados por mediciones de estaciones meteorológicas automáticas del IDEAM ("Unipacífico", "La Cumbre" y "Farallones"). El escenario objeto de la investigación es la costa pacífica del departamento del Valle del Cauca, con caso de estudio sobre la cuenca del río Dagua. El análisis cubre el periodo de 2015-2035 y se basa en los escenarios SSP1-19, SSP1-26 y SSP2-45, para la cual se generó información espacio-temporal de temperatura y precipitación a una resolución espacial de 0.1° (~ 9 km) y resolución temporal horaria. Dado que los GCM presentan una menor resolución espacial y temporal, se emplearon los datos de ERA5-Land como referencia para el proceso de *downscaling*, permitiendo así una mayor fidelidad en la representación de las variaciones climáticas locales para el periodo 2024-2035. Entre los modelos evaluados, la arquitectura ConvLSTM2D demostró ser la que mejor representa las dependencias temporales y espaciales complejas en los datos climáticos. Los resultados obtenidos para la variable de temperatura muestran un coeficiente de determinación $R^2 = 0.972$, un error cuadrático medio (ECM): 1.001°C y un error absoluto medio (EAM): 0.733°C reflejando una alta precisión en las predicciones. Para la precipitación, el modelo alcanzó un $R^2 = 0.020$, un RECM: 0.264 mm/h y un EAM: 0.150 mm/h, lo que indica desafíos adicionales en la representación a resolución horaria de esta variable, debido a la alta variabilidad en las precipitaciones en la región Pacífica del departamento. La información ambiental generada está a disposición del público,

desplegada en la nube, para la consulta y descarga de datos de temperatura y precipitación en los diferentes futuros climáticos establecidos por el IPCC6, por medio de una plataforma digital de uso libre.

Palabras clave: downscaling estadístico, temperatura, precipitación, escenarios climáticos, ERA5-Land, aprendizaje automático, inteligencia artificial.

INTRODUCCIÓN

Este proyecto aporta la iniciativa presentada del Grupo de Investigación en Simulación y Modelación Dinámica Espacial - GISMODEL a la convocatoria 21 del Ministerio de Ciencia, Tecnología e Innovación "Convocatoria de la asignación para la CTel del Sistema General de Regalías", articulada con lo establecido en el Plan Nacional de Desarrollo 2018 - 2022 (MinCiencias, 2022).

El enfoque general de la iniciativa busca mejorar la comprensión del sistema hidroclimático de la cuenca del río Dagua a través de la representación de los datos de temperatura y precipitación proyectados hasta el año 2035. El proyecto genera información para estas dos variables a partir de la integración de los datos de Modelos de Circulación Global (del inglés GCM) y estimaciones climáticas con reanálisis ERA5-Land, donde los resultados son validados con datos de las observaciones medidas por las estaciones climatológicas de las instituciones ambientales nacionales y regionales. La propuesta explora diferentes técnicas de modelamiento climático comunes como la Regresión Lineal (RL) y modelos más robustos de Inteligencia Artificial (IA) que permitan comparar y mejorar tanto las resoluciones espaciales como temporales, para así lograr disponer de escenarios climáticos futuros aptos para poder ser usados en investigaciones con resoluciones más finas que las originadas por los GCM.

De este modo se plantea la pregunta ¿cómo se modela el clima?. El clima es un patrón promedio de variables meteorológicas en un área determinada durante un periodo prolongado. Entre las variables ambientales más significativas que determinan el clima se encuentran la temperatura y la precipitación, que a su vez caracterizan los distintos ecosistemas que encontramos en la superficie terrestre. Los GCM describen la dinámica del sistema climático global y son herramientas ampliamente utilizadas para estudiar la evolución del clima, en el pasado y a futuro a través de simulaciones que abarcan hasta el año 2100 o 2200 (Hannah, 2022).

Los GCM son diseñados por diversas instituciones, como servicios meteorológicos nacionales, centros de investigación, laboratorios climáticos y universidades alrededor del mundo, utilizando registros históricos desde el año 1850 hasta el año 2014 (Eyring y cols., 2016). Los GCM modelan múltiples variables climáticas clave para la investigación a gran escala y presentan resoluciones espaciales de 100 km o más por lado de píxel. La resolución temporal de los datos generados

varía según la variable, y puede abarcar desde datos diarios hasta promedios mensuales (Jalota, Vashisht, Sharma, y Kaur, 2018).

Debido a que cada institución utiliza métodos y enfoques únicos en sus modelamientos, obteniendo diferentes resultados en cada simulación, en 1995 se creó el Proyecto de Intercomparación de Modelos Acoplados (CMIP), organizado por el Programa Mundial de Investigación Climática. Su principal función es centralizar estos modelos para almacenar y comparar sus resultados, lo que permite una comprensión más profunda del sistema climático a nivel global (Peng y cols., 2023).

El CMIP proporciona una base de referencia fundamental para los informes del Panel Intergubernamental sobre Cambio Climático (IPCC), el principal organismo internacional encargado de evaluar el cambio climático global. El IPCC ofrece diversos escenarios futuros del clima, conocidos como Trayectorias de Concentración Representativas (RCP, por sus siglas en inglés). En su versión más reciente, se introdujeron las Vías Socioeconómicas Compartidas (SSP), que permiten explorar cómo diferentes trayectorias de desarrollo socioeconómico pueden influir en los diferentes futuros escenarios climáticos (Bordera y cols., 2022).

El marco conceptual de las SSP representa un avance significativo en la construcción de Escenarios de Cambio Climático, proporcionando una base estructurada para evaluar los posibles impactos del cambio climático bajo diferentes trayectorias socioeconómicas globales. Estas vías permiten una mejor integración de factores socioeconómicos en las evaluaciones de los impactos hidrológicos y climáticos (Estoque, Ooba, Togawa, y Hijioka, 2020).

Estas SSP se utilizan para generar proyecciones futuras de los cambios climáticos que podrían ocurrir en función de distintos niveles de desarrollo socioeconómico y la interacción de múltiples factores a escala global. Estas trayectorias consideran la evolución de la población, el crecimiento económico, el uso del suelo y las políticas ambientales, entre otros aspectos, que determinan la magnitud y velocidad del cambio climático. Así, la evaluación de los efectos futuros del cambio climático se basa en estos métodos para capturar la vulnerabilidad de los sistemas socioambientales, explorando cómo diferentes escenarios de desarrollo global pueden modificar la intensidad, frecuencia y distribución de los impactos climáticos en distintas regiones del planeta (Rohat, 2018).

Las ecuaciones empleadas en los GCM generan múltiples variables climatológicas y ambientales, como la precipitación, la temperatura, la radiación solar y los componentes del viento. Sin embargo, debido a que estas simulaciones se producen en una malla de cuadrículas de resolución relativamente gruesa, es necesario emplear técnicas de reducción de escala, conocidas como *downscaling* (Mhaweji y Abunnasr, 2022). Estas técnicas permiten obtener proyecciones más detalladas y específicas para áreas geográficas concretas, mejorando así la aplicabilidad de los modelos a nivel regional y local.

Para este caso, las técnicas de *Downscaling* tienen como objetivo mapear estas simulaciones globales de gran escala en una región local a una resolución más fina. El *downscaling* se realiza utilizando dos enfoques, el dinámico y el estadístico; por un lado el dinámico que imita las metodologías utilizadas en los GCM son muy costosas y requieren de altos recursos computacionales. Por el contrario, los enfoques de *downscaling* estadísticos se basan en modelos o algoritmos estadísticos que suponen una relación de estacionariedad que vinculan los resultados de granularidad gruesa de los GCM con las estimaciones de modelos de una menor resolución Espacio-Temporal como los de reanálisis de ERA5-Land y observaciones locales de estaciones meteorológicas distribuidas sobre el área de estudio (Baghanam, Eslahi, Sheikhabaei, y Seifi, 2020).

Colombia se encuentra ubicada sobre el paralelo del Ecuador, lo que, sumado a sus características geográficas, se presentan complejas condiciones hidroclimatológicas, que son generadas por las circulaciones atmosféricas y el aumento de la temperatura de los océanos, afectando fenómenos naturales como El Niño Oscilación Sur (ENOS). El ENOS es un evento climático de origen oceánico y atmosférico que impacta en los patrones climáticos a nivel global, siendo esta una oscilación irregular entre dos fases opuestas: El Niño, asociado al calentamiento del océano Pacífico ecuatorial, y La Niña caracterizada por un enfriamiento de estas aguas, que también afectan la circulación atmosférica asociada (Hoyos, Baquero-Bernal, Jacob, y Rodríguez, 2013). Estas variaciones climáticas y una fuerte orografía propician que la Región Pacífica del país sea una de las más lluviosas del mundo y con diversos pisos climáticos.

La región del Pacífico colombiano se caracteriza por presentar dos temporadas climáticas bien definidas a lo largo del año. La temporada seca, que se extiende entre enero y junio, se asocia con una reducción relativa de las precipitaciones, aunque sigue siendo una de las regiones más lluviosas del mundo. Durante estos meses, los acumulados mensuales de precipitación oscilan entre 300 mm y 600 mm, con un valor anual promedio que ronda los 7,000 mm a 8,000 mm. Las temperaturas durante este periodo muestran una mayor estabilidad, con valores medios que varían entre 26°C y 28°C, aunque pueden superar los 30°C en áreas de menor altitud. En años de El Niño, la disminución de las lluvias puede llegar hasta un 30 % por debajo de los valores normales, intensificando las condiciones secas y favoreciendo eventos de déficit hídrico en ríos y ecosistemas (Trojer, 2017).

Por otro lado, la temporada húmeda, comprendida entre julio y diciembre, se caracteriza por un aumento significativo en las precipitaciones, con acumulados mensuales que pueden superar los 1,000 mm en algunas áreas. En este periodo, el promedio anual de precipitación puede alcanzar 9,000 mm o más, especialmente en zonas cercanas a la costa. En años de La Niña, las lluvias pueden incrementarse hasta en un 40 % por encima de los valores normales. Las temperaturas durante esta temporada se mantienen dentro del mismo rango de 26°C a 28°C, pero con mayor sensación térmica

debido a la alta humedad relativa. Este comportamiento interanual, influenciado principalmente por el fenómeno de ENOS, determina la disponibilidad de agua en la región (Enriquez, Guzmán, y Narváez, 2014).

El proyecto se desarrolla sobre la región Pacífica del departamento del Valle del Cauca, con caso de estudio en la cuenca del río Dagua. La cuenca del río Dagua nace sobre la cordillera de los Andes y se extiende hacia la costa del océano Pacífico. Esta cuenca presenta una topografía y cobertura vegetal muy heterogénea, que desarrollan una gran diversidad climática y numerosos ríos que hacen de la cuenca uno de los principales productores de oferta hídrica del departamento del Valle del Cauca (. CVC, 2020). La cuenca se divide en dos subregiones, la zona alta y la zona baja, donde se presentan grandes cambios de temperatura y precipitación, entre el promedio anual multianual de los 10 a 20 °C de temperatura y una precipitación de 900 a 1600 mm por año en la parte alta. En cuanto a la parte baja, la temperatura anual multianual se promedia entre los 23 y 30 °C y la precipitación desde los 1800 a 10000 mm al año (Aguirre, Ibarra, Trochez, Guevara, y Bermudez, 2017).

La precipitación y la temperatura son reconocidas como las variables más significativas del ciclo global del agua y motor central de los procesos hidrológicos de la superficie terrestre, ellas juegan un papel importante en la investigación hidrológica, meteorológica y ecológica de la cuenca (Eekhout, Hunink, Terink, y Vente, 2018) (Osborn y cols., 2021). Igualmente los datos relacionados con la gestión del agua y los peligros que esta pueda ocasionar son de gran significancia (Arun, Chowdary, Kesarwani, y Neeti, 2023), de modo que, el objetivo de este proyecto es representar los datos climatológicos de precipitación y temperatura con una resolución espacial y temporal más adecuada para el modelamiento dinámico y aplicaciones de estudio y comportamiento climático, con técnicas de *Downscaling* para el estudio de la región Pacífica del departamento y la cuenca del río Dagua.

El presente documento está compuesto por la siguiente estructura: en el primer capítulo se presenta la descripción de la problemática planteada, la construcción del árbol de problemas, la formulación de la pregunta y el enfoque de investigación. Luego se presentan el objetivo general y los objetivos específicos con el diseño del árbol de objetivos con los resultados esperados en este proyecto y los medios con los cuales se busca dar solución a la problemática propuesta. Además se expone la justificación, donde se presentan las razones y conveniencia de resolver esta problemática.

En el capítulo dos se describen los marcos conceptuales, teóricos y referenciales involucrados en la fundamentación del problema y solución planteada. El capítulo tres contiene la estrategia del marco experimental , presentando el desarrollo de la metodología en sus diferentes fases. En el cuarto capítulo se presentan los resultados, el análisis y las discusiones desarrolladas. Para finalizar con el capítulo 5 se presentan las conclusiones y por último la bibliografía consultada.

Capítulo 1

PROBLEMA

1.1. Definición del problema

En el escenario actual del Cambio Climático Global (CCG) y la Variabilidad Climática (VC) se ha producido un incremento sustancial en la frecuencia, severidad y magnitud de los eventos hidrometeorológicos extremos, generando mayor impacto climático sobre los ecosistemas y afectando considerablemente el paisaje (Ghil y Lucarini, 2020). Como lo menciona el IPCC, cada vez es más probable que la temperatura promedio de la Tierra se incremente en 2 °C para el año 2100 (Kamworapan, Thao, Gheewala, Pimonsree, y Prueksakorn, 2021). La Variabilidad Climática y Cambio Climático Global, generan frecuentes anomalías en la precipitación y temperatura que originan eventos hidrometeorológicos extremos que impactan con mayor severidad a los países en desarrollo que causan cambios profundos en su territorio, provocando consecuencias perjudiciales al ser humano, su entorno y principales actividades (X. Yang y cols., 2022).

Este efecto del aumento en la temperatura puede ser debido al incremento de los gases de efecto invernadero, el dióxido de carbono CO₂ que se acumulan en la atmósfera tanto por causas naturales como antrópicas principalmente por el consumo de combustibles fósiles y otras variables antrópicas asociadas como la deforestación y urbanización (Yuan, Zhang, Busteed, y Flanagan, 2022). Una de las principales limitaciones en la comprensión del sistema climático a nivel global radica en las escalas espaciales y temporales de los modelos climáticos globales y la distribución dispersa de las estaciones meteorológicas. Estos modelos suelen basarse en datos históricos obtenidos de sensores remotos con resoluciones espaciales y temporales relativamente bajas, complementados por registros puntuales de estaciones distribuidas en distintas regiones del mundo. Aunque los modelos generan productos a nivel global, su aplicabilidad a nivel regional es limitada, lo que restringe su uso en estudios específicos de ciencias de la Tierra y en aplicaciones locales o regionales. (Carlsonweb, 2022). Por ejemplo, la resolución espacial representada por el

GCM son alrededor de 100 kilómetros de lado de píxel o más (Hooker, Duveiller, y Cescatti, 2018), (Olén y Lehsten, 2022). Además, se considera que las simulaciones de GCM incluyen sesgos con relación en los datos utilizados en el desarrollo del modelo y se sabe que estas simulaciones no logran abordar adecuadamente las cuestiones climáticas interdisciplinarias locales, aportando bajo potencial para construir e implementar capacidades adaptativas a la VC y el CCG, particularmente a escalas de interés a nivel local, donde el clima tiene fluctuaciones significativas, incluso dentro de una celda de cuadrícula de GCM (Potgieter, Schepen, Brider, y Hammer, 2022).

Con respecto al departamento de Valle del Cauca y la cuenca del río Dagua, al estar ubicada en la zona intertropical del globo terrestre y en la región del Pacífico colombiano, presenta dos temporadas climáticas frecuentes denominadas temporada seca y temporada húmeda, que en las últimas décadas han sido afectadas fuertemente por la VC y CCG (Ministerio de Ambiente y Desarrollo Sostenible, 2017). Cabe considerar que la cuenca cuenta con características peculiares de climas ecosistémicos, un paisaje muy heterogéneo y escarpado que imprime una alta complejidad de condiciones ambientales, presentando constantemente inundaciones, deslizamientos y una significativa carga de sedimentos a la bahía de Buenaventura (Arias, Ortega, Villegas, y Martínez, 2021). Por otra parte, el departamento posee pocas observaciones puntuales y alta dispersión de medios de captura de datos climáticos primarios, debido tanto a las limitaciones de las observaciones de los sensores remotos como al alto déficit de estaciones meteorológicas que miden variables meteorológicas que recopilan datos puntuales (in situ). La ubicación de las estaciones que se encuentran en la cuenca son muy limitadas y están dispersas espacialmente y sus registros son de poca frecuencia en el tiempo, especialmente en la región Pacífica (Baquero y cols., 2018).

En términos generales, las bases de datos son el componente principal en las técnicas de Downscaling estadístico, donde a la fecha son utilizados modelos frecuentes como los lineales, análogos, máquina de vectores de soporte, bosques aleatorios y redes neuronales artificiales (Ebrahimi y Azadbakht, 2019). Pese a que sus resultados son satisfactorios, estos métodos no tienen la capacidad de manejar automáticamente los espacios de entrada de alta dimensionalidad y la no linealidad de los datos, sin llevar el modelo a un sobreajuste. Esto ocasiona que los datos de entrada al modelo, deban ser sometidos a tratamientos de selección y reducción de características tediosos y "guiados por humanos", lo que generalmente trae consigo una pérdida de información que puede ser relevante para explicar la variabilidad local de los datos para el valor pronosticado (Ballarin y cols., 2023).

Con el objetivo de consolidar un conjunto de datos climáticos proyectados con mayor resolución Espacio-Temporal, se busca mejorar la representatividad a una escala más detallada de la región Pacífica del departamento del Valle del Cauca, ya que estos datos son fundamentales para servir como insumo en modelos hidrológicos dinámicos, permitiendo análisis más precisos y ajustados a las condiciones específicas de la zona.

1.2. Árbol del problema

Para ilustrar el cómo se llegó a la identificación del problema, a continuación se presenta el árbol del problema construido con la metodología del marco lógico propuesto por Planeación Nacional.

Tabla 1.1: Árbol del problema

Impacto	Poco detalle para representar información climatológica histórica y a futuro útil para modelamiento hidrológico		
Efectos indirectos o a largo plazo	Datos climatológicos generalizados y de bajo potencial para implementar capacidades adaptativas a la VC y CCG		No se generan relaciones que unen los datos de entrada con las predicciones que determinan la climatología local
Efectos o síntomas	Baja resolución espacial y temporal de la precipitación y temperatura	Poca disponibilidad de información ambiental local para la cuenca	Alta incertidumbre de los datos ambientales de la cuenca a una resolución más detallada
P R O B L E M A	Limitada información espacial y temporal de precipitación y temperatura de la cuenca del río Dagua dadas las condiciones particulares presentadas en el paisaje		
Causas directas	Gruesa resolución espacial y temporal de los datos globales ambientales disponibles de los GCM	Pocas estaciones de captura de información climatológica para la cuenca	Alta dispersión de los datos ambientales disponibles a nivel local
Causas indirectas	Poca disponibilidad y conocimiento de datos ambientales frecuentes del clima a nivel local	Falta innovación en la integración de diferentes fuentes de información ambiental para la cuenca	Insuficiente información ambiental confiable distribuida para la cuenca

1.3. Pregunta de investigación

Teniendo en cuenta las características actuales de los datos climáticos derivados de los GCM, ¿Cómo podemos mejorar la representación Espacio-Temporal de las variables de la temperatura y la precipitación en escenarios climáticos donde el paisaje presenta alta complejidad como lo es el Pacífico colombiano?

1.4. OBJETIVOS

Dando respuesta a lo planteado en el árbol del problema, se realiza la siguiente propuesta de objetivos para el presente proyecto:

1.4.1. Árbol de objetivos

Tabla 1.2: Árbol de objetivos

Impactos a mediano plazo	Datos de temperatura y precipitación más detallados y de mayor aplicabilidad	Mayor potencial para construir e implementar evaluaciones de capacidades adaptativas a la VC y CCG	Aprender patrones complejos y no-lineales que están presentes en los datos espaciales	Extender el acceso a los datos resultantes del proyecto, que sean fácilmente encontrados y accedidos en los formatos usuales con herramientas de gestión de datos espaciales
Resultados esperados	Mejorar la resolución espacial y temporal de precipitación y temperatura, resultante de la información ambiental disponible	Mejor granularidad y disponibilidad de información ambiental para la cuenca	Menor incertidumbre en aplicaciones y resultados de la modelación hidrológica	Mayor acceso de información ambiental a nivel local de la cuenca del río Dagua
Objetivo General	Disponer de información espacial y temporal adecuada de la cuenca del río Dagua dadas las condiciones particulares encontradas en el paisaje			
Objetivos específicos	Disponer de información ambiental a una resolución espacial y temporal más fina	Obtener proyecciones climáticas de mayor resolución espacial y temporal mediante técnicas de inteligencia artificial utilizando fuentes de datos existentes	Desarrollar modelos predictivos basados en técnicas de Inteligencia Artificial para generar información ambiental más confiable	Implementar una plataforma web digital que permita visualizar, consultar y descarga de información ambiental proyectada más detallada de la cuenca del río Dagua
Medios	Implementar Modelos que integren los datos disponibles para la reproducción de los escenarios ambientales, mejorando sus características Espacio-Temporales		Aplicar las arquitecturas de Deep Learning que mejor relaciona las variables ambientales con las condiciones particulares de la cuenca	Desarrollar una aplicación web de libre acceso con la integración del modelo de Inteligencia artificial más preciso

1.4.2. Objetivo general

Disponer información climática de temperatura y precipitación proyectada a una resolución espacio-temporal más apropiada para la cuenca del río Dagua, generada a partir de datos de GCM, ERA5-Land y observaciones locales, mediante el uso de técnicas de inteligencia artificial.

1.4.3. Objetivos específicos

- Interpretar las bases de datos espaciales de los GCM, ERA5-Land y estaciones 'in situ' a través de un análisis exploratorio y descriptivo de las variables climáticas, para establecer las interacciones en los dominios espacial y temporal de la zona de estudio.
- Aplicar técnicas de Inteligencia Artificial con arquitectura LSTM a los datos de GCM y ERA5-Land para la reducción de escala de la temperatura y precipitación, evaluando su rendimiento y aplicabilidad en comparación con técnicas convencionales de Regresión Lineal, Random forest y Regresión de vectores de soporte.
- Desarrollar una plataforma digital para disponer datos climatológicos proyectados de los GCM hasta el año 2035 de precipitación y temperatura más confiables y detallados para la región Pacífica del Valle del Cauca y la cuenca del río Dagua dada la mejor precisión de los modelos de Inteligencia Artificial implementados.

1.5. JUSTIFICACIÓN

La temperatura y la precipitación son dos elementos de estado clave en el ciclo del agua y de energía entre la atmósfera y la superficie terrestre (L. Wang y cols., 2022)(Kamworapan y cols., 2021), pero los productos de los Modelos de Circulación Global existentes de temperatura y precipitación suelen tener resoluciones espaciales de cientos de kilómetros, lo que no obedece con las necesidades de las investigaciones y aplicaciones a nivel local (H. Zhang y cols., 2021). La información más precisa y detallada de la distribución Espacio-Temporal de la temperatura y precipitación mejora sustancialmente la comprensión que tenemos de la dinámica del sistema climático actual y los diferentes escenarios representados a futuro.

Una gran parte del departamento del Valle del Cauca se encuentra ubicada en la región Pacífica colombiana y como una de las cuencas más importantes de la región, la cuenca del río Dagua se nutre los ríos Bitaco, San Cipriano, Pavas, Río Grande, Mozambique, Pepitas, San Cristóbal, Escalerete que son una de las fuentes productoras principales de agua del departamento del Valle del Cauca (CVC, 2017). Esta gran materia hídrica generada en la cuenca hace que en esta se presenten constantemente fuertes inundaciones, crecientes y deslizamientos que afectan las actividades socioeconómicas de la región, que muchas veces cobran vidas e interfieren con el transporte hacia el principal puerto marítimo de la región y aportando una gran cantidad de sedimentos al puerto de Buenaventura (Fedesarrollo, 2013). Por lo tanto, es importante evaluar la variabilidad del clima a futuro, para comprender sus impactos sobre sistemas naturales y socioeconómicos (Chen, Chen, Qin, Zhao, y Duan, 2020), y tomar medidas que permitan mitigar

sus impactos sobre el desarrollo de la región, la seguridad alimentaria y la sustentabilidad del medio natural de la región (Aguirre y cols., 2017).

La urgencia de la crisis climática despierta gran interés entre investigadores y gobiernos para enfrentar y reconocer los efectos del CCG y la VC. Es así como los datos Espacio-Temporales de precipitación y temperatura más confiables y continuos con una granularidad más fina, son particularmente necesarios e implican un mayor nivel de detalle en regiones heterogéneas y de topografía compleja. En general, cuanto más fina sea la resolución espacial y temporal mejor serán los resultados y menor la incertidumbre de la modelación en una amplia gama de aplicaciones donde se involucra principalmente los recursos hídricos (Karaman y Akyürek, 2023).

El terreno complejo y escarpado como el que encontramos en la cuenca alta y baja del río Dagua, son primordiales para la investigación relacionada con los estudios de la dinámica ambiental y la variabilidad climática local sobre la costa del Pacífico colombiano (González, 2014). De este modo, se busca brindar una estimación más precisa y continuidad en el espacio y el tiempo de las variables de temperatura y precipitación en las proyecciones del cambio climático del IPCC, enfocada en un área de interés más específica que la considerada por los GCM. Esto permitirá fortalecer los esfuerzos locales para comprender el cambio climático global, gestionar el riesgo de desastres y la conservar los recursos hídricos (H. Zhang y cols., 2021).

Capítulo 2

REVISIÓN DE LITERATURA

Para el siguiente capítulo se presentan los marcos conceptual, teórico y referencial del proyecto de grado que fundamentan la elaboración de la tesis de investigación. En la primera sección se encuentra el marco conceptual, donde se destacan los conceptos básicos del clima, en cómo es su comportamiento alrededor del planeta y los modelos climáticos globales actuales que simulan todo el proceso de las variables ambientales que componen el clima. Por otra parte, se plantean las técnicas y modelos para reducir la resolución espacial y temporal de las variables de temperatura y precipitación.

En el marco teórico se plantean los procedimientos, algoritmos y ecuaciones consideradas en los modelos de *downscaling* estadístico y las métricas de validación que evaluarán la precisión de los resultados obtenidos en cada uno de los diferentes modelos.

Por último en el marco referencial se expone el estado del arte de la investigación, dando a conocer las metodologías y autores más relevantes consultados en este estudio, que serán relacionados con el cumplimiento de los objetivos de la investigación. Es así como se enfatiza en el análisis de la variabilidad climática y el cambio climático en la zona de estudio, la reducción de la resolución Espacio-Temporal de la temperatura y precipitación y la publicación en la web de los productos generados.

2.1. MARCO CONCEPTUAL

El clima describe las condiciones climáticas o el tiempo promedio medido para un lugar en particular durante un período que va desde varios meses hasta los miles de años (FAO, 2022). Por lo general los actuales estudios sobre el clima promedian dos variables elementales de la superficie terrestre, la precipitación y la temperatura en un ciclo de 30 años como lo establece World Meteorological Organization (WMO).

El clima en el mundo se puede describir como el funcionamiento de un sistema, siendo el

resultado de la interacción de cinco componentes clave: la atmósfera, la hidrósfera, la criósfera, la litósfera y la biósfera. (Randall y cols., 2007). Asimismo, el clima en la Tierra está determinado por factores como el ingreso de la radiación solar y la retención de esta energía dentro de la atmósfera, conocido como el efecto invernadero que proporciona una distribución global de energía como lo son el calor y el viento; además de una distribución de masa como la humedad del aire, nubosidad y la precipitación (Hannah, 2015).

Ahora bien, el clima de una determinada región no solo depende de esta distribución de masa y energía, sino también de sus características geofísicas como la longitud, la latitud, la altura, la orografía, su cercanía al mar y sobre determinada circulación atmosférica o bajo la influencia de alguna corriente marina. Además la esfericidad de la tierra hace que más energía y calor se acumulen en el trópico que en las zonas polares; En el océano el calor se acumula en promedio en un grosor de agua de 100 a 200 metros y en el suelo en una capa de 1 a 2 metros, transfiriendo este calor a la atmósfera, produciendo así esta diferencia de distribución espacial de la presión atmosférica que a su vez produce la circulación del aire y las corrientes marinas (IDEAM - UNAL, 2018).

Esta distribución a nivel global de calor y humedad ocasionan patrones Espacio-Temporales de temperatura y precipitación, que inducen características particulares a cada región en diferentes partes del mundo, produciendo diferentes índices climáticos, lo que se conoce como la clasificación de Köppen-Geiger (Hamed y cols., 2023). Esta distribución de la temperatura y la precipitación organizan una repartición de los biomas y funcionan de acuerdo con la estacionalidad del clima y en cada continente está asociada a su tipo de vegetación, flora, fauna y ecosistemas que determinan su productividad y desarrollo, de tal manera que las diferentes actividades humanas por región se especializan en algún tipo de disciplina como la agricultura, ganadería, minería, turismo, entre otras, que son efectos socioeconómicos favorables para cada territorio (Yin, Liang, Wang, y Zhou, 2022).

Al estudiar estos parámetros de distribución climática durante diferentes periodos, hacemos dos distinciones muy importantes, la Variabilidad Climática (VC) y el Cambio Climático Global (CCG). Cuando se observa la precipitación y la temperatura en el transcurso de unos días, meses e incluso pocos años, y se condiciona a causas naturales del ciclo hidrológico se define como VC. Por otro lado, al estudiar los cambios a largo plazo y atribuidos a efectos antrópicos que alteran la composición atmosférica y los patrones climáticos se especifica como CCG. Cuando observamos el clima, estudiamos los patrones promedio de temperatura y precipitación en un lugar durante 30 años o más (Ghil y Lucarini, 2020).

La VC siempre ha estado presente entre nosotros y esta hace referencia a un comportamiento anormal del clima que se presenta con cierta frecuencia, pero esta es temporal y transitoria como es el caso del fenómeno de El Niño Oscilación Sur (ENOS) (Canchala y cols., 2022). ENOS es

un fenómeno de variabilidad climática, mientras que el CCG está asociado con la variación de la temperatura promedio que es permanente y para ser verificada se necesita de varias décadas de registros climatológicos recolectados desde el siglo pasado.

El clima global es un sistema dinámico e interconectado, en el cual cada uno de sus componentes mantiene una relación de interdependencia. Esto implica que una alteración en una región puede incidir en variaciones climáticas en otras partes del mundo. Por ejemplo, mientras en algunas zonas se experimentan sequías extremas y escasez hídrica, en otras regiones pueden manifestarse precipitaciones intensas, tormentas severas e inundaciones (Feser, Rockel, Storch, Winterfeldt, y Zahn, 2011).

Desde esta perspectiva, el fenómeno El Niño-Oscilación del Sur (ENOS) surge a partir de la interacción acoplada entre el océano y la atmósfera en el Pacífico ecuatorial. Se distingue por un anómalo calentamiento de las aguas superficiales del Pacífico, lo que modifica la circulación atmosférica y los patrones climáticos a escala planetaria. En contraste, La Niña representa la fase opuesta del fenómeno, caracterizándose por un enfriamiento anómalo de las mismas aguas (Z. Wang, Wu, Chen, y Qiao, 2013).

El ENOS está estrechamente asociado con fluctuaciones en la intensidad y dirección de los vientos alisios, también conocidos como vientos del este (provenientes del noreste en el hemisferio norte y del sureste en el hemisferio sur). Estas variaciones modifican la estructura térmica del océano, particularmente la termoclina, que actúa como una capa de transición entre las aguas cálidas superficiales y las frías profundas, que en conjunto con los cambios en la presión atmosférica, son factores determinan la magnitud y el alcance del impacto del ENOS en los patrones climáticos a nivel global (Zhou y cols., 2023).

El Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) declara que el fenómeno ENOS se consolida cuando las condiciones de aumento de la temperatura superficial del mar y la disminución de los vientos alisios se mantienen durante 5 meses continuos. Colombia es un país tropical que, debido a su posición geográfica cerca de la línea ecuatorial, no tiene estaciones sino dos períodos climáticos durante el año, uno de más lluvia (época húmeda) y otro de menos lluviosa (época seca). Precisamente por esta posición natural, en el país los efectos de un fenómeno de variabilidad climática presentan unas cualidades muy particulares y diferentes a la mayoría de territorios de este continente (Urrea, Ochoa, y Mesa, 2019).

El territorio de Colombia se localiza entre las latitudes 4 grados sur y 13 grados norte, donde la franja ecuatorial predomina sobre la mayor parte del país, lo que es un factor determinante en la distribución espacial y temporal de la precipitación, temperatura, nubosidad y diferentes variables climatológicas (Hoyos y cols., 2013). Además, Colombia al estar ubicado sobre complejas características hidroclimatológicas, tiene mucha influencia de las circulaciones atmosféricas de los procesos que ocurren en los océanos Atlántico y el Pacífico tropical, el fuerte gradiente topográfico

que se observa en sus regiones e importantes cuencas de los ríos Cauca, Orinoco y Amazonas, que combinado con el aumento de la temperatura de los océanos, afecta fenómenos naturales como ENOS (Feijt, 2015).

Debido a las variaciones climáticas a través de todo país se induce a que la región pacífica colombiana sea una de las zonas más lluviosas del mundo y particularmente por los ciclos extremos de El Niño y La Niña. La compleja topografía del país también comprende innumerables cantidad de montañas y valles que propician una diversidad de pisos térmicos y microclimas (IDEAM, 2005).

La cuenca de Dagua nace alrededor de los 2.200 m.s.n.m a lo alto de los Andes occidentales, sobre el Parque Nacional Natural Farallones de Cali y se extiende hacia el occidente en su demarcación más baja sobre la cota cero sobre el nivel del mar, generando comportamientos únicos y que en conjunto, modifican el equilibrio térmico del clima a nivel regional. Esta se divide en dos partes: cuenca alta y cuenca baja (Productora y consumidora) (Andina y Pacífica). En donde su relieve se puede encontrar altas cumbres y profundas simas que le permiten contar con una gran diversidad climática, abundantes fuentes hídricas y bosques primarios sin intervenir en los cuales se produce un conjunto de subcuencas y microcuencas cuyas aguas se trasvasan al río Dagua para así desembocar en el océano Pacífico sobre la bahía de Buenaventura (Gutiérrez, Carvajal, y Ávila, 2013).

Esta cuenca presenta un paisaje fuertemente escarpado, pendientes pronunciadas, una cobertura del suelo de bosque denso y selvas tropicales por las cuales corren caudalosos ríos. Por tal motivo la cuenca cuenta con abundante oferta hídrica que determina una alta diversidad biológica de flora y fauna, que ayuda al mantenimiento de los procesos ecológicos, siendo una de las zonas más privilegiadas en materia de recursos naturales (Xie, Li, Tian, Wu, y Jin, 2023).

Las características únicas del relieve que se exhiben en la cuenca compone un enorme factor modificador del clima, al mismo tiempo estos diversos tipos climáticos se relacionan sobre todo con la altitud presentada por el paisaje y cuya ubicación determinada entre la zona intertropical (Franja entre los trópicos de capricornio y cáncer) hacen que en esta se presenten y surjan tres pisos térmicos, el piso macrotérmico, el piso subtropical y el templado, donde la temperatura y precipitación varían considerablemente. Cabe destacar que en esta delimitación de ecosistemas encontramos las selvas tropicales o pluviales que se encuentran entre los 0 y 1200 m.s.n.m con una temperatura de 23 a 30 °C y una alta precipitación entre 1800 y 10000 mm al año. Por otra parte, los bosques muy secos tropicales o subxerofíticos se presentan específicamente en la zona correspondiente al cañón del río Dagua a una altura entre los 400 y 1400 m.s.n.m con pocas precipitaciones de 900 a 1600 mm por año y un rango de temperatura de 10 a 20 °C (Aguirre y cols., 2017).

Como variables esenciales del ciclo global del agua y motor central de los procesos hidrológicos de la superficie terrestre, la precipitación y la temperatura juega un papel importante en la

investigación hidrológica, meteorológica y ecológica de la cuenca. Más aún los datos apropiados de precipitación y temperatura son cruciales para la mayoría de los estudios ambientales, como el pronóstico de inundaciones, el monitoreo de sequías, la gestión de recursos de cuencas hidrográficas y los datos relacionados con la gestión del agua y los peligros que esta pueda ocasionar, de modo que, es de gran interés comprender las características y fluctuaciones de la precipitación y temperatura especialmente la distribución espacial y el cambio temporal (Eekhout y cols., 2018) (Osborn y cols., 2021).

Por otro lado la simulación del clima es desarrollada con herramientas muy potentes como los Modelos de Circulación General (GCM), estos son ampliamente aplicados, ya que generan las proyecciones para comprender y estudiar el sistema climático a nivel mundial, en el pasado, presente y futuro (Guo y cols., 2021). Los GCM son modelos numéricos que representan el sistema climático a nivel mundial con las interacciones entre la atmósfera, el océano, la criósfera, la superficie terrestre, basados en procesos físicos, químicos y biológicos (Goosse, Barriat, Loutre, y Zunz, 2010). Estos modelos de simulación del CCG proyectan los efectos de la acumulación de Gases de Efecto Invernadero (GEI), con una resolución típica de 100 a 200 Km (Toigo, Lee, Newman, y Richardson, 2012).

Como se puede apreciar en la figura 2.1 esta representa la temperatura promedio diaria a nivel global del GCM GFDL-ESM4, y la figura 2.2 representa la precipitación promedio diaria, a una resolución espacial de 100 Km.

Esta versión desarrollada por la NOAA Geophysical Fluid Dynamics Laboratory (GFDL), un modelo de sistema terrestre acoplado que integra la química atmosférica, los ciclos biogeoquímicos y la dinámica climática. Este modelo, que forma parte de las contribuciones al Sexto Informe de Evaluación del IPCC (IPCC AR6), incorpora avanzadas representaciones de los ecosistemas terrestres y marinos, permitiendo una simulación más realista del balance de carbono y las interacciones entre la atmósfera, los océanos y la biosfera. Su enfoque multidisciplinario lo convierte en una herramienta fundamental para el estudio del cambio climático y la evaluación de escenarios futuros a nivel global.

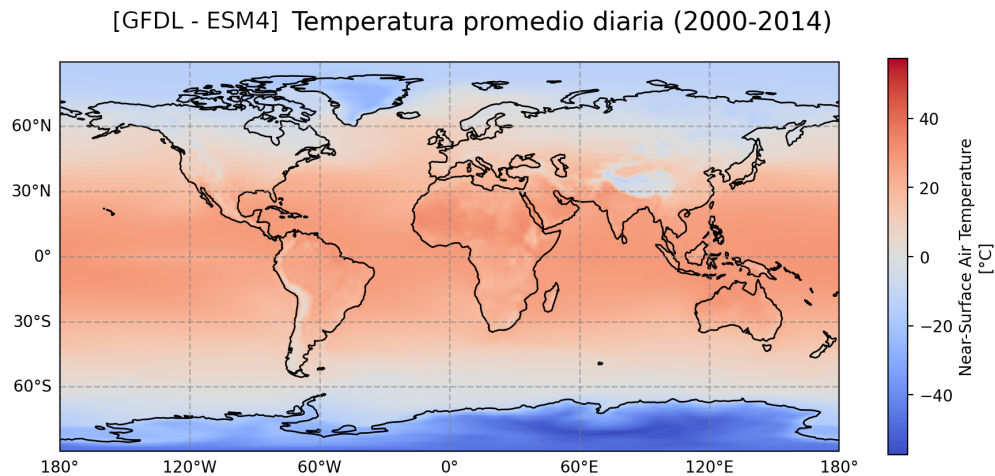


Figura 2.1: Temperatura promedio diaria cerca a la superficie a nivel global del Modelo GFDL-ESM4 de la NOAA. Datos históricos con una resolución espacial por píxel de 1°.

Fuente: NOAA (2024)

En 1995, se estableció el Proyecto de intercomparación de modelos acoplados (CMIP), organizado por el Programa Mundial de Investigación Climática (WCRP), coordinando a la comunidad global de modelado climático para recopilar y archivar los resultados de GCM, en un formato estandarizado y un alcance definido, a disposición de los usuarios con dominio público, convirtiéndose en un elemento central de las investigaciones sobre el cambio climático a nivel global y regional (Tian y Dong, 2020). Debido a que cada institución alrededor del mundo genera diferentes GCM, al propósito de este proyecto es centralizar y llevar a cabo un conjunto común de simulaciones climáticas, ayudando a evaluar la precisión de cada modelo, conduciendo a mejoras en las simulaciones y proporcionando una mejor comprensión y representación del sistema climático global.

Los modelos climáticos del CMIP en su fase seis (CMIP6) lanzan nuevas versiones de GCM que proporcionan una importante base de referencia para el sexto informe de evaluación del Panel Intergubernamental sobre el Cambio Climático (IPCC) (Rivera y Arnould, 2020). El IPCC en su versión anterior (IPCC5) presenta diferentes escenarios denominados Trayectorias de Concentración Representativas (RCP, Representative Concentration Pathways), los cuales son un conjunto de vías de concentración y emisión de gases de efecto invernadero que se definen por el nivel de forzamiento radioactivo. Para su versión actual el IPCC6 integra las Vías Socioeconómicas Compartidas (SSP, Shared Socioeconomic Pathways), que proyecta posibles futuros climáticos en función de las anteriores RCP, las condiciones socioeconómicas y las políticas de mitigación, diseñadas para

apoyar la investigación sobre los impactos y las posibles respuestas políticas al cambio climático (Kim, Choi, Choi, y Park, 2013) (L. Liu y cols., 2022).

Los productos originales de los GMC como se mencionó anteriormente son de escala bastante gruesa, por lo general deben reducirse a una escala más fina, para ser utilizados en simulación a un nivel de escala más local (D. L. Liu y Zuo, 2012). Los métodos de *downscaling* busca ajustar las simulaciones de los GCM a una resolución más alta en el ámbito local, mejorando su aplicabilidad en diferentes estudios. El *downscaling* se realiza utilizando dos enfoques, por un lado el dinámico que imita las metodologías utilizadas en los GCM que son muy costosas y requieren de altos recursos computacionales. Los enfoques de *downscaling* estadísticos se basan en modelos o algoritmos estadísticos que suponen una relación de estacionariedad que vinculan los resultados de granularidad gruesa de los GCM con las observaciones locales de estaciones meteorológicas distribuidas sobre el área de estudio (Baghanam y cols., 2020).

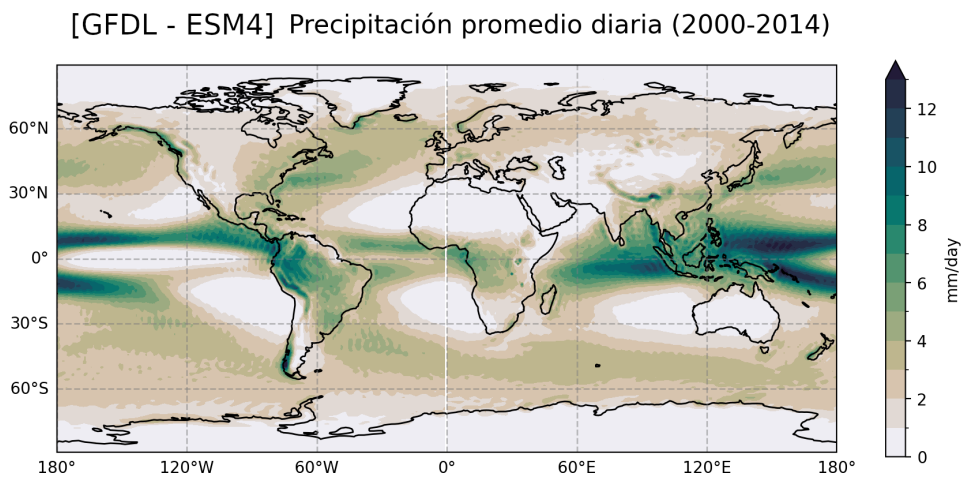


Figura 2.2: Precipitación promedio diaria a nivel mundial del modelo GFDL- ESM4 con una resolución espacial por píxel de 1°.

Fuente: (NOAA, 2024)

Actualmente existen diferentes técnicas de *downscaling* estadístico que emplean modelos lineales, análogos, máquinas de vectores de soporte (MSV), bosques aleatorios y redes neuronales artificiales (RNA) (Ebrahimi y Azadbakht, 2019). Las RNA es una metodología de inteligencia artificial, en la que los algoritmos se inspiran en la estructura del cerebro biológico, donde estos programas aprenden y se forman a sí mismos, en lugar de ser programados de forma explícita con el objetivo de resolver problemas de la misma manera que el cerebro humano. Las RNA toman los diferentes datos recolectados y se pasan por las diferentes capas de neuronas para que estas se

entrenen y aprendan a reconocer los patrones que en los datos se presenten, para luego pronosticar los resultados en un nuevo conjunto de datos y validar el modelo diseñado. Dentro de este contexto la comunidad científica ha enfocado sus estudios en las últimas topologías de redes neuronales profundas o Deep Learning con los modelos de Redes Neuronales Convolucionales (RNC) y *Long Short Term Memory* (LSTM), siendo estos candidatos potencialmente poderosos para diferentes disciplinas y aplicaciones orientadas al clima, incluido la reducción a escala (Aguirre y cols., 2017).

El modelo Random Forest (RF) es una técnica de aprendizaje supervisado ampliamente utilizada en el ámbito del *downscaling* estadístico debido a su capacidad para manejar grandes cantidades de datos y múltiples variables explicativas. RF se basa en la construcción de múltiples árboles de decisión, donde cada árbol es entrenado con un subconjunto diferente de datos y variables. Esta combinación de árboles permite que el modelo encuentre patrones robustos, mejorando significativamente el rendimiento y la precisión en comparación con métodos individuales (W. Xu, Zhang, Long, y Qin, 2021). Por su capacidad para reducir el sobreajuste y manejar tanto datos lineales como no lineales, este modelo es una herramienta clave para mejorar la resolución espacial y temporal de datos climáticos, permitiendo una representación más detallada de fenómenos locales.

La Máquina de soporte vectorial (MSV) es un algoritmo de Machine Learning utilizado para problemas de regresión, utilizando una función kernel que resuelve problemas complejos de una forma lineal, transformando los datos de entrenamiento desde un espacio de menor dimensión, a un espacio de mayor dimensión, donde es más fácil encontrar un hiperplano separador, buscando evitar problemas de sobreajuste, desarrollando un algoritmo de función de optimización, donde la convergencia ocurre en los mínimos globales y no locales utilizando funciones de activación lineales, polinómicas, función de base radial y sigmoidal (Ebrahimi y Azadbakht, 2019).

2.2. MARCO TEÓRICO

Las técnicas estadísticas de reducción de escala utilizan funciones lineales y no lineales para transformar las características de los GCM, que operan a una escala amplia, en proyecciones más detalladas a nivel local. Dentro de esta categoría, los métodos de regresión aplican herramientas estadísticas y de aprendizaje automático para mapear las características de una cuadrícula amplia a una cuadrícula más pequeña, permitiendo estimar variables climatológicas como la precipitación y la temperatura de manera más localizada.

Estos métodos estadísticos se fundamentan en funciones que representan la relación entre las variables climáticas atmosféricas de gran escala y los valores observados a nivel local. Entre los métodos de regresión no lineales más utilizados para la reducción de escala se encuentran las Redes Neuronales Artificiales (RNA), las Máquinas de Soporte Vectorial (SVM) y los modelos

de Bosques Aleatorios (Random Forest, RF). Por otro lado, los modelos de regresión lineal, que establecen una relación directa entre un indicador simulado a gran escala y un valor observado a escala local, son metodologías más sencillas. Aunque menos complejos, estos modelos lineales se emplean como una base de comparación para evaluar el desempeño de los modelos más robustos.

2.2.1. Análisis de componentes principales

El análisis de componentes principales (ACP) es una técnica implementada para reducir la dimensión de un conjunto de múltiples variables que pueden estar relacionadas, siendo esta una estrategia útil para la limpieza y preparación de los datos a modelar. El ACP nos permite extraer la información más relevante de un grupo de múltiples variables, eliminando la multicolinealidad que se pueda presentar entre las mismas variables y mejorando así el rendimiento para el procesamiento de datos a la hora de la modelación (Preisendorfer y Mobley, 1988).

El ACP se basa en fundamentos de álgebra lineal y estadística multivariable que permiten conservar la mayor cantidad de varianza posible que se presenten en los datos. El objetivo de ACP es transformar un conjunto de datos X de dimensión $n \times m$ a otro conjunto de datos Y de menor dimensión $n \times l$ con la menor pérdida de información útil posible utilizando para ello el método basado en la matriz de covarianza, donde el número de componentes principales tiene que ser inferior a la menor de las dimensiones de X $l \leq \min\{n, m\}$.

Para el análisis los datos tienen que estar centrados en una media de 0 (restar la media de cada columna) y/o autoescalados (centrados en una media de 0 y dividiendo cada columna por su desviación estándar) 2.1:

$$Z_j = \frac{X_j - \mu_j}{\sigma_j} \quad (2.1)$$

Donde:

Z_j : Variable estandarizada

X_j : Variable original

μ_j : Media de X_j

σ_j : Desviación estándar de X_j

El ACP se basa en la descomposición de vectores propios de la matriz de covarianza la cual se calcula con la ecuación 2.2 :

$$C = \frac{1}{n-1} Z^T Z \quad (2.2)$$

Donde:

C : Matriz de covarianza

n : Número de observaciones

Z : Matriz de datos estandarizados

El siguiente paso es el cálculo de los autovalores y autovectores, resolviendo la ecuación característica 2.3. Luego los autovalores se ordenan de mayor a menor: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Estos autovectores correspondientes forman las columnas de la matriz de componentes principales.

$$(C - \lambda I)v = 0 \quad (2.3)$$

Donde:

λ : Autovalores

v : Autovectores

I : Matriz identidad

Por último se seleccionan las componentes principales, calculando la proporción de varianza explicada por cada componente:

$$\text{Proporción de varianza } a_i = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j} \quad (2.4)$$

Seleccionando los primeros k componentes que explican una proporción acumulada de varianza deseada que en términos ideales se encuentra entre un 80 % y 95 % (Yazdian, Salmani-Dehaghi, y Alijanian, 2023).

2.2.2. Interpolación bicúbica

La interpolación bicúbica es una técnica basada en el uso de polinomios cúbicos para realizar el cálculo de valores intermedios en una función bidimensional, realizando la interpolación tanto en la dirección vertical, como horizontal de una cuadrícula de datos raster (Y. Zhang, Feng, Zhang, Wang, y Wang, 2021).

Esta interpolación considera los 16 píxeles más cercanos en una cuadrícula de 4×4 para estimar el valor en un punto intermedio, donde su fórmula general se define como:

$$f(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j \quad (2.5)$$

Donde:

a_{ij} son los 16 coeficientes que se deben determinar.

2.2.3. Regresión lineal

Los modelos estadísticos basados en regresión son la técnica más comúnmente utilizada en la reducción de escala estadística. Estos pueden emplear diferentes variables climáticas, como la humedad relativa, presión atmosférica y el viento, como variables predictoras para estimar de otras variables como la temperatura y la precipitación (Sun, Lan, y Jiang, 2023).

El método estadístico de regresión lineal es utilizado para estudiar la relación de una variable dependiente Y , entre una o diferentes variables independientes X . El cálculo de la ecuación corresponde a la línea o hiperplano en mayor dimensiones que mejor describe la relación entre la variable respuesta y la variable explicativa. El modelo se expresa como ecuación 2.6:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \xi \quad (2.6)$$

Donde:

Y : variable dependiente o respuesta

X : variables explicativas o independientes

$\beta_0, \beta_1, \beta_n$: parámetros del modelo

ξ : término de error asociado

En el proceso de modelación y evaluación de un modelo de RL, es fundamental verificar que los residuos o errores cumplan con ciertos supuestos estadísticos, lo cual ayuda a interpretar el comportamiento e interpretabilidad del modelo (Li y Yan, 2017), (Pahlavan, Zahraie, Nasser, y Mahdipour Varnousfaderani, 2018). En esta investigación, se evalúan los siguientes supuestos: Normalidad, Homocedasticidad e Independencia. Estos supuestos proporcionan información sobre la distribución y estructura de los errores, lo que permite diagnosticar posibles deficiencias en el modelo y establecer una base a el uso de modelos alternativos.

2.2.4. Random forest

La metodología de Random Forest se compone por el conjunto de múltiples árboles de decisión de forma aleatoria, en el que cada uno de los árboles tiene la capacidad de predecir una variable respuesta a partir de un conjunto de diferentes variables, como se representa en el diagrama de la figura 2.3. Este método de aprendizaje automático es potente y versátil que funciona bien para diferentes tipos de problema de predicción incluyendo las variables climáticas como la temperatura y precipitación. Esta predicción tiende a mejorar utilizando una gran variedad de árboles de decisión (Y. Liu, Jing, Wang, y Xia, 2020).

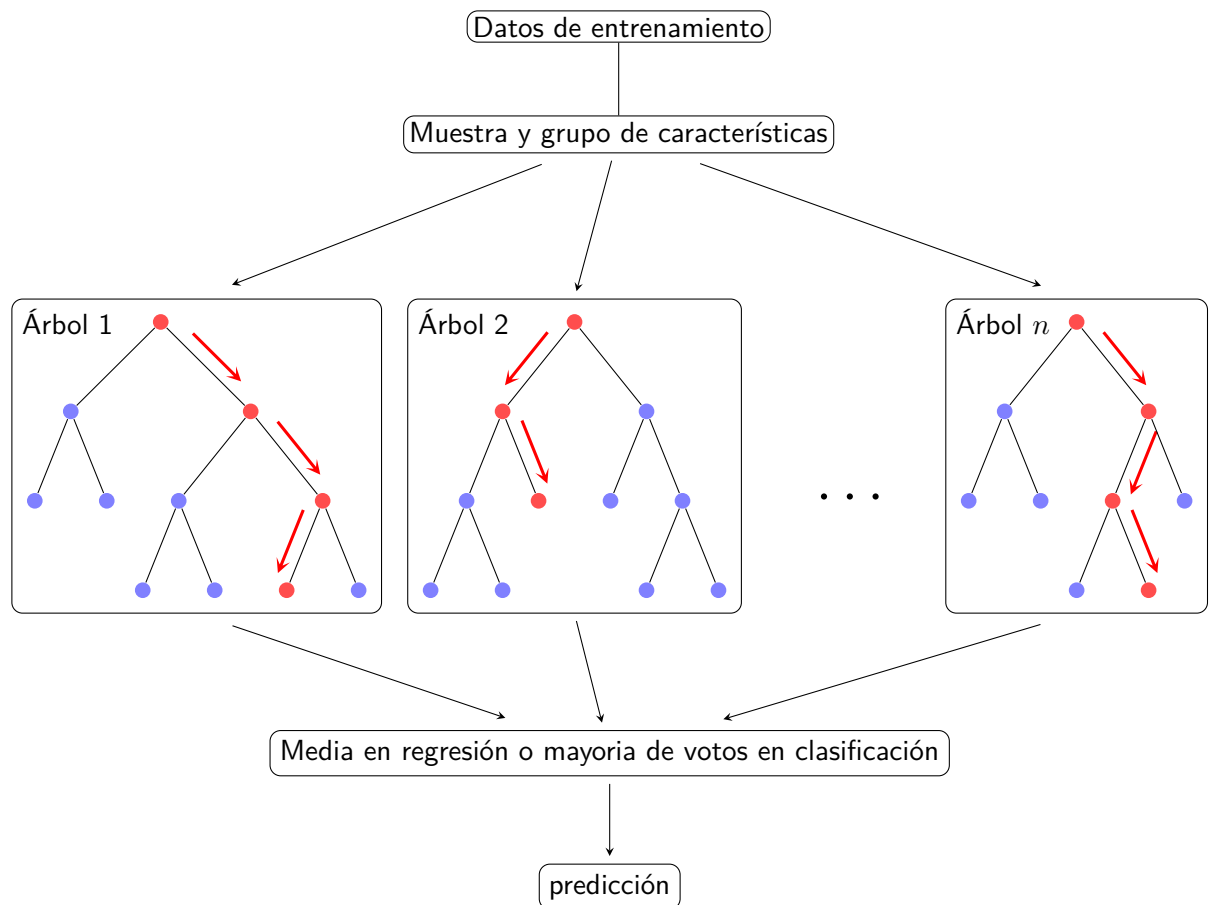


Figura 2.3: Diagrama simple de un Random Forest

Fuente: (Riebesell, 2024)

Los árboles de decisión como base de construcción de este método, es una estructura que toma decisiones basadas en preguntas sobre las características de los datos. En el contexto de la predicción, un árbol realiza diferentes preguntas entrenándose con una muestra aleatoria de los datos y un subconjunto aleatorio de características, cuya aleatoriedad ayuda a reducir el sobre ajuste y mejorar la generalización del modelo.

2.2.5. Máquina de soporte de vectores

Las Máquinas de Soporte Vectorial (MSV) son un conjunto de algoritmos de aprendizaje supervisado muy efectivos para resolver problemas de regresión, figura 2.4. A partir de un conjunto de muestras, el modelo se entrena para predecir valores futuros. Una de las principales ventajas de esta técnica es su flexibilidad, que permite definir el margen de error aceptable en el modelo, ajustando una línea o hiperplano óptimo que se adapta a los datos, incluso en espacios de alta

dimensionalidad.

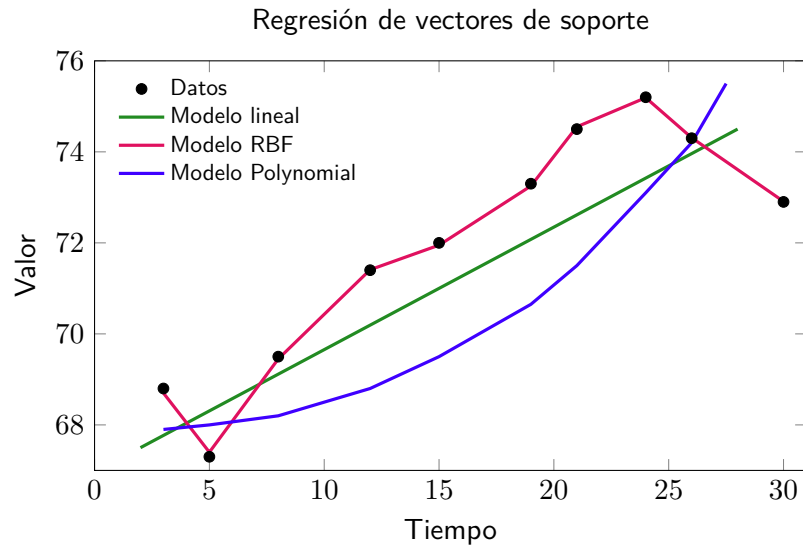


Figura 2.4: Gráfico del comportamiento de los Kernel por modelo de SVR

Fuente: (Dash, Nguyen, Cengiz, y Sharma, 2023).

El modelo MSV estima la regresión basándose en una serie de funciones kernel, que son capaces de convertir los datos de entrada originales, de menor dimensión, a un espacio de características de mayor dimensión de forma implícita (Fan y cols., 2018). La estimación del modelo de regresión podría expresarse con la Ecuación 2.7:

$$\hat{y} = \langle w, x \rangle + b \quad (2.7)$$

Donde:

w: vector de ponderaciones de características

$\langle \rangle$: paréntesis angulares indican un producto escalar

b: término de sesgo.

La MSV integra una función de pérdida para minimizar el error de predicción y tiene la intención de crear un límite para incluir tantas muestras de confiabilidad como sea posible (W. Wang y cols., 2019). Siguiendo la teoría de la regularización, los parámetros w y b se estiman minimizando la función de costo como se muestra en la ecuación 2.8.

$$Min : \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (2.8)$$

Donde:

ξ_i y ξ_i^* : variables de holgura positivas, que capturan la magnitud de los residuos más allá de la tolerancia prescrita

ε : garantiza una solución para todo ε .

C: término de regularización que determina el grado de la sanción lineal aplicada al exceso residual.

En consecuencia, la ecuación 2.8 está sujeta a las siguientes restricciones:

$$\begin{aligned} y_i \langle w, x_i \rangle - b &\leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i &\leq \varepsilon \xi_i^* \\ \xi_i^* &\geq 0, \xi_i \geq 0 \end{aligned}$$

Donde:

x_i : punto de datos en el espacio de entrada X

y_i : salida correspondiente.

Las funciones kernel comúnmente aplicadas incluyen funciones lineales, polinómicas, de base radial y sigmoideas. La función de base radial gaussiana (FBR) es una de las funciones kernel más utilizadas que podría generalizar funciones no lineales y se ha desempeñado bien en grandes conjuntos de datos. El kernel RBF está formulado de la siguiente manera:

$$\varphi(x, x') = \exp(-\gamma |x, x'|^2), \quad \gamma > 0 \quad (2.9)$$

Donde:

x, x' : son vectores en el espacio de entrada.

γ (*gamma*): es un parámetro de ajuste que controla la influencia de cada muestra en la transformación.

2.2.6. Redes neuronales artificiales

Las redes neuronales artificiales (RNA) son modelos matemáticos inspirados en el sistema nervioso del cerebro humano. Las redes neuronales en el cerebro humano consisten en células nerviosas. Cada neurona está conectada con otras células nerviosas. Al igual que las neuronas

humanas, las RNA consisten en neuronas que están conectadas a las neuronas de la capa anterior o la capa siguiente. Se asigna un peso a cada conexión. La salida de cada neurona es una función de estos pesos (Sathe, Upadhye, y Agnieszka, 2023).

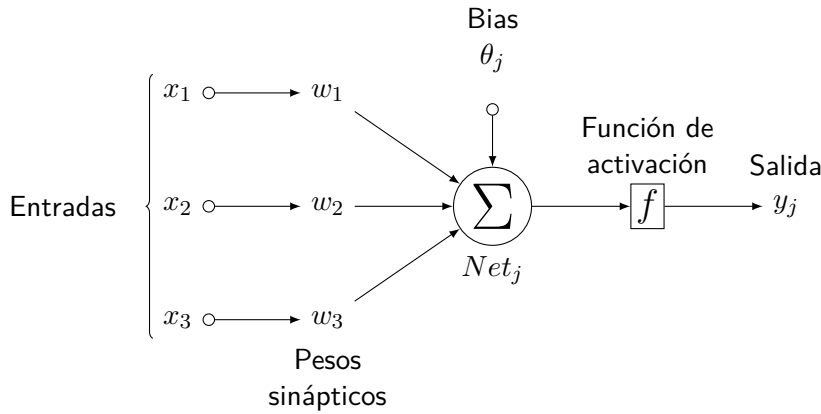


Figura 2.5: Representación del modelo perceptrón

Fuente: (Caicedo y López, 2017)

Cada capa tiene un número de nodos. La Figura 2.6 muestra una estructura de red neuronal. Cada nodo contiene una función de activación. La función de activación especifica la salida de una neurona. Se pueden utilizar diferentes funciones de activación en una red, algunas funciones de activación comunes son sigmoide, tangente y lineal. Las ecuaciones 2.10, 2.11, 2.12, 2.13 describen estas funciones (Caicedo y López, 2017).

$$\text{Lineal} : f(x) = x \quad (2.10)$$

$$\text{Sigmoide} : f(x) = \frac{1}{1 + e^{-x}}; \quad 0 < f(x) < 1 \quad (2.11)$$

$$\text{Sigmoide logística} : f(x) = \frac{e^{-x} - e^{-x}}{e^{-x} + e^{-x}}; \quad -1 < f(x) < 1 \quad (2.12)$$

$$\text{ReLU} : \text{ReLU}(x) = \max(0, x) \quad (2.13)$$

En este modelo se define el vector de entrada como $X = [x_1, x_2, \dots, x_n]$, donde estos datos son modificados por un vector w de pesos sinápticos, que se asimilan como ganancias que pueden atenuar o amplificar los valores que se desean propagar hacia la siguiente neurona. El parámetro θ_j se conoce como el sesgo o el umbral. Los diferentes valores que recibe la neurona, modificados por los pesos sinápticos, son sumados para producir lo que se denomina la entrada neta. Esta entrada neta es la que va a determinar si la neurona se activa o no. La activación o no de la neurona

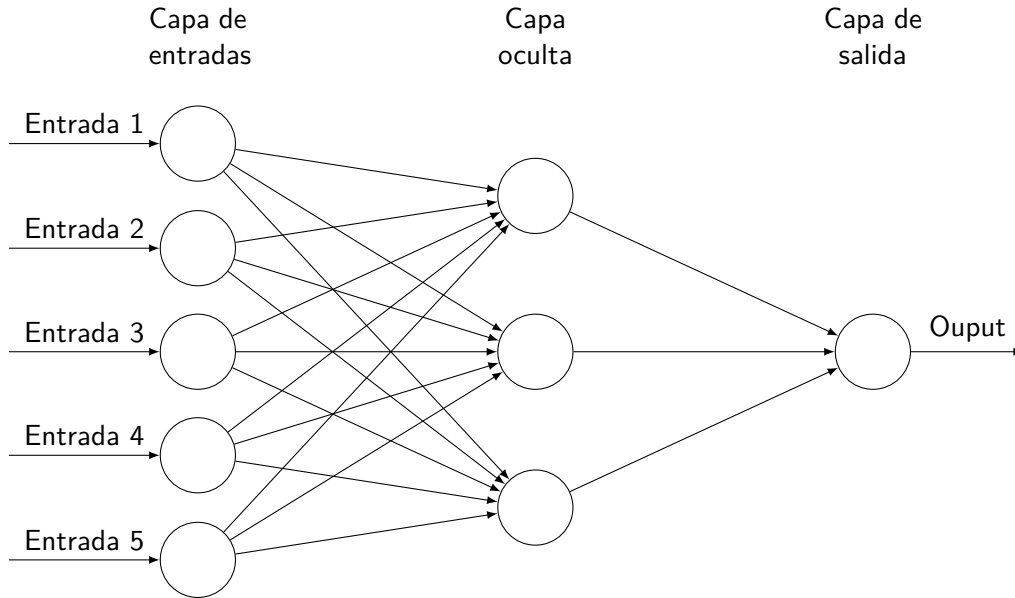


Figura 2.6: Modelo de una red neuronal artificial

Fuente: (Caicedo y López, 2017)

depende de lo que se llama Función de Activación. La entrada neta se evalúa en esta función y obtenemos la salida de la red. De acuerdo a lo mencionado, la salida y_j de la neurona se genera al evaluar la neta en la función de activación.

En una neurona artificial el procesamiento matemático para el cálculo de la entrada neta se puede representar con la ecuación 2.14, y en forma vectorial se representa con la ecuación 2.15.

$$Net_j = \sum_{i=1}^N x_i w_{ji} + \theta_j \quad (2.14)$$

$$Net_j = w_1 x_{j1} + w_2 x_{j2} + w_i x_{ji} + \dots + w_N x_{jN} \quad (2.15)$$

La salida de la neurona artificial está determinada por una función de activación (Fact), tal como se aprecia en la ecuación 2.16.

$$y_j = Fact_j(Net_j) \quad (2.16)$$

La función de activación generalmente es de tipo lineal, sigmoideal, sigmoide bipolar o ReLU, como se presenta en las siguientes Figuras:

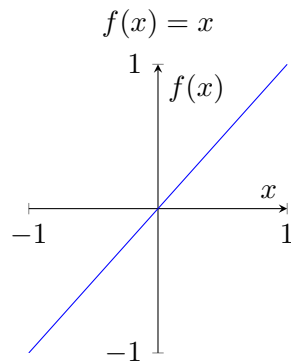


Figura 2.7: Función lineal

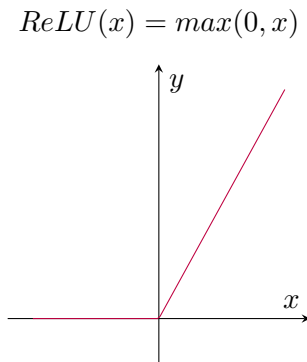


Figura 2.8: Función ReLU

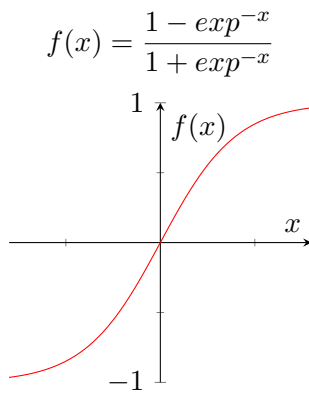


Figura 2.9: Función Sigmoide bipolar

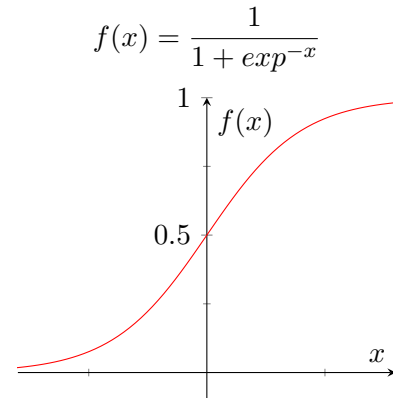


Figura 2.10: Función Sigmoide

Fuente: (Caicedo y López, 2017)

El objetivo principal de la red neuronal es analizar los datos y ajustar los pesos en un proceso llamado "entrenamiento de red". Es así como las redes neuronales de tipo convolucionales (RNC) aprovecha la estructura espacial inherente a los datos climáticos, otorgando importancia a cada píxel dentro de las imágenes que representan variables meteorológicas. A diferencia de las redes neuronales tradicionales, las RNC capturan relaciones espaciales al considerar la dependencia entre píxeles vecinos en las dimensiones espaciales (ancho y alto). Esto permite identificar estructuras, patrones y características relevantes que pueden ser utilizadas para mejorar la comprensión del comportamiento del clima.

El núcleo de una RNC es la operación de convolución, la cual aplica un filtro o kernel K a la imagen de entrada X , generando una nueva representación de la misma. Matemáticamente, la convolución se define como:

$$Y_{(i,j)} = \sum_m \sum_n X(i+m, j+n) \cdot K_{m,n} \quad (2.17)$$

Donde:

$Y_{(i,j)}$: es el valor del píxel en la salida después de aplicar la convolución.

$X(i+m, j+n)$: representa los valores de la imagen de entrada en una ventana definida por el filtro.

$K(m,n)$: kernel o matriz de pesos que extrae características específicas.

m, n : recorren las dimensiones del filtro.

Esta estructura de RNC son utilizadas en la predicción de fenómenos meteorológicos, facilitando la identificación de patrones espaciales y la mejora en la precisión de modelos de pronóstico climático.

Así mismo el algoritmo LSTM (del inglés, *Long Short-Term Memory*) es un tipo de red neuronal recurrente muy utilizada en la actualidad, Figura 2.11. Una LSTM es capaz de ‘recordar’ un dato relevante en la secuencia y de preservarlo por varios instantes de tiempo. Por lo tanto, esta arquitectura puede tener una memoria tanto de corto plazo como también de largo plazo (Hochreiter y Schmidhuber, 1997).

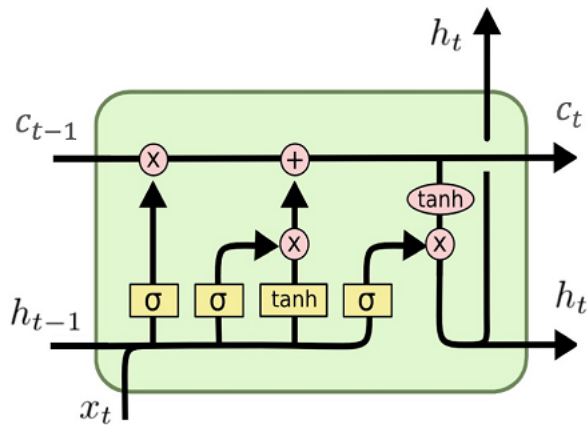


Figura 2.11: Arquitectura de una red neuronal LSTM

Fuente: (Hamarashid, 2021)

Donde:

x_t : vector de entrada en el tiempo t .

h_{t-1} : estado oculto de la celda en el tiempo anterior $t - 1$.

c_{t-1} : estado de la celda en el tiempo $t - 1$, que transporta información a largo plazo.

Puerta de olvido σ : determina cuánta información del estado de la celda anterior c_{t-1} se debe olvidar.

Puerta de entrada σ y $\tanh$: decide qué nueva información se agrega al estado de la celda.

Actualización del estado de la celda C_t : donde se combina la información de la puerta de olvido y la puerta de entrada.

Puerta de salida σ y $\tanh$: decide qué parte del estado de la celda se usa para la salida h_t .

La mejora detrás de la arquitectura LSTM es crear un módulo adicional en una red neuronal que aprenda cuándo recordar y cuándo olvidar información pertinente. Es así como la red aprende eficientemente, cuál información podría ser necesaria o no, en una siguiente secuencia o iteración. La arquitectura básica del modelo clasificador de redes neuronales profundas LSTM, se construye usando la siguiente función: (Misra, Sarkar, y Mitra, 2018).

$$S_t = f(P_t, P_{t-1}, S_{t-1}) \quad (2.18)$$

Donde:

S_t : representa la salida o el estado interno de la red en el tiempo t .

f : función de activación realizadas dentro de la red.

P_t : variable de entrada en el tiempo t .

P_{t-1} : valor de la variable de entrada en el instante anterior.

S_{t-1} : contiene la información histórica acumulada hasta el tiempo $t - 1$.

2.2.7. Medidas de desempeño y métricas de validación

Cada modelo da origen a una predicción del valor desconocido al cual se pretende estimar produciendo una incertidumbre en cada paso del modelamiento, por lo tanto es importante medir la precisión y variabilidad de cada predicción.

Para la verificación de las metodologías planteadas, se presentan las medidas de desempeño aplicadas a los modelos y métricas de validación para la evaluación de los resultados de las predicciones, con las observaciones reales, realizando la comparación entre los diferentes métodos utilizados.

Las evaluaciones más utilizadas por los autores consultados en esta investigación son: Error medio cuadrático (EMC), Raíz del error medio cuadrático (REMC), Error medio absoluto (EMA), Error porcentual absoluto medio (EPAM) Coeficiente de correlación de Spearman (S), Coeficiente

de correlación de Pearson (r), Coeficiente de determinación (R^2) y el Diagrama de Taylor (Tran Anh, Pham, Dang, y Hoang, 2019), (Chou, Park, y Chou, 2021), (Sun y cols., 2023), (Kheir, Elnashar, Mosad, y Govind, 2023).

2.2.7.1. Error cuadrático medio (ECM)

El error medio cuadrático mide el promedio de las diferencias entre los valores estimados y el valor que se estima o el promedio de los errores al cuadrado.

$$EMC = \frac{\sum_{i=1}^n (y_i - x_i)^2}{n} \quad (2.19)$$

Donde:

y_i : vector de n predicciones

x_i : vector de los verdaderos valores

Esta métrica penaliza más los errores grandes debido a la elevación al cuadrado, donde un ECM bajo indica que las están generalmente cerca de los valores reales y un ECM alto indica una gran dispersión de las predicciones respecto a los valores reales.

2.2.7.2. Raíz del error cuadrático medio (RECM)

La raíz del error medio cuadrático promedio muestra la diferencia entre los valores previstos y los valores observados en los modelos planteados.

$$REMC = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (2.20)$$

Donde:

y_i : es un vector de n predicciones.

x_i : es el vector de los verdaderos valores.

Esta es la raíz de ECM, por lo que mantiene las mismas unidades de la variable objetivo.

2.2.7.3. Error absoluto medio (EAM)

El error medio absoluto es una medida estadística que sirve para cuantificar la precisión de una técnica de predicción, comparando los valores predichos con los valores observados.

$$EMA = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (2.21)$$

Donde:

y_i : es un vector de n predicciones

x_i : es el vector de los verdaderos valores

Mide el error absoluto promedio sin elevarlo al cuadrado, por lo que no sobrepondera errores grandes. A diferencia del MSE y RMSE, no castiga errores grandes en exceso, por lo que es útil cuando se desea un error más interpretable en términos de unidades.

2.2.7.4. Error porcentual absoluto medio (EPAM)

El error porcentual absoluto medio mide el error en porcentaje del promedio entre las diferencias de las predicciones del modelo planteado y el valor real observado.

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{y_i - x_i}{x_i} \right|}{n} \times 100 \quad (2.22)$$

Donde:

y_i : es un vector de n predicciones.

x_i : es el vector de los verdaderos valores.

n : número total de observaciones.

Los posibles resultados del EPAM se consideran en:

- EPAM bajo: menor a 10 % (Excelente predicción).
- EPAM moderado: entre el 10 % y 20 % (Predicción aceptable).
- EPAM Alto: mayor a 20 % (Baja precisión).

2.2.7.5. Coeficiente de correlación de Pearson (r)

El coeficiente de correlación de Pearson es una medida de dependencia lineal entre dos variables, midiendo el grado de relación entre las dos variables.

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (2.23)$$

Donde:

y_i : es un vector de n predicciones.

x_i : es el vector de los verdaderos valores.

Los valores de correlación r pueden ubicarse entre -1 y 1. Una baja correlación no necesariamente indica que el modelo sea erróneo, ya que puede haber relaciones no lineales en los datos, donde:

- $r \approx 1$: correlación fuerte.
- $r \approx 0$: no hay correlación.
- $r \approx -1$: correlación inversa fuerte.

2.2.7.6. Coeficiente de determinación (R^2)

El coeficiente de determinación es la porción de la varianza total de la variable explicada por las técnicas aplicadas, reflejando la bondad del ajuste del modelo a la variable que se pretende explicar.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Donde:

- y_i son los valores observados de la variable dependiente.
- $\hat{y}_i$ son los valores predichos por el modelo.
- $\bar{y}$ es la media de los valores observados.

Si los datos tienen alta variabilidad o no tienen una relación lineal, un R^2 puede ser normal, aunque el modelo tenga un ECM aceptable, donde se interpreta los valores de R^2 como:

- $R^2 \approx 1$: el modelo explica casi toda la variabilidad.
- $R^2 \approx 0$: el modelo no explica la variabilidad mejor que un promedio simple.
- $R^2 < 0$: el modelo es peor que una predicción basada en el promedio.

2.2.7.7. Coeficiente de correlación de Spearman (S)

El coeficiente de correlación de Spearman es la versión no paramétrica de la correlación de Pearson (r), es una medida de relación no lineal entre dos variables, midiendo el grado de relación entre las dos variables de distribución no gaussiana.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2.24)$$

Donde:

d_i : $R(X_i - Y_i)$, es la diferencia entre los dos rangos de cada observación.

Este coeficiente mide la relación monótona entre las variables, donde sus posibles resultados son:

- $\rho = 1$: relación monótona fuerte.
- $\rho = 0$: no hay relación monótona.
- $\rho = -1$: relación monótona fuerte decreciente.

2.2.7.8. Diagrama de Taylor

El diagrama de Taylor es una representación matemática diseñado para evaluar el rendimiento de modelos de predicción climática, facilitando visualmente la comparación de las diferentes medidas estadísticas que describen la relación entre los datos modelados y los datos observados, determinando cuál de los modelos aproximados de un fenómeno es más realista.

Este diagrama cuantifica el grado de correspondencia entre el comportamiento del modelo y el valor real en términos de las tres métricas estadísticas mencionadas: Coeficiente de correlación Pearson, Raíz del error cuadrático medio y la desviación estándar.

El diagrama de Taylor traza estos tres estadísticos en un solo gráfico, utilizando coordenadas polares Figura: 2.12:

- Radio: Representa la desviación estándar.
- Ángulo: Representa la correlación.
- Curvas concéntricas: Representan líneas de igual RMSE.

2.3. MARCO REFERENCIAL

Actualmente se han desarrollado diversas técnicas de IA en *Downscaling* Estadístico (DE) para la investigación de diferentes aplicaciones de ciencias de la Tierra, como lo es el modelamiento del clima y su comportamiento a nivel local (X. Zhang y cols., 2022); siendo la temperatura y la precipitación, las dos variables elementales, que buscan orientar a los investigadores y científicos a estudios más sofisticados sobre los diferentes ecosistemas, con el fin de reducir errores e incertidumbres en sus resultados y conclusiones (Saraf y Regulwar, 2016).

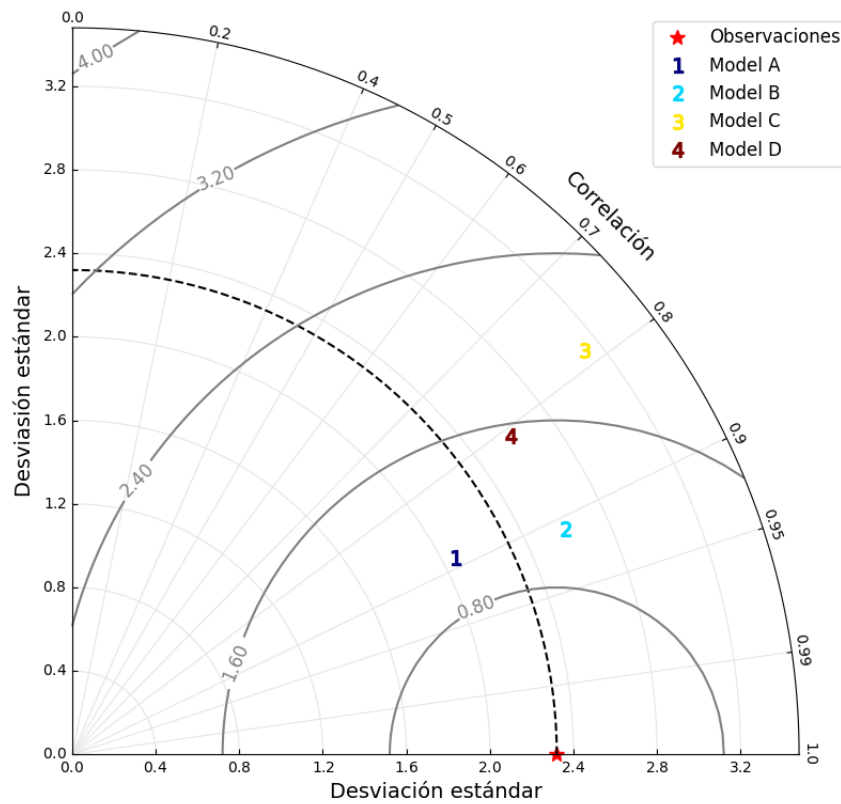


Figura 2.12: Diagrama de Taylor

Fuente: (Ortega, Arias, Villegas, Marquet, y Nobre, 2021), (L. Wang y cols., 2022)

Comprender el comportamiento de la temperatura y la precipitación de un periodo histórico determinado, permite reconocer los patrones y tendencias del clima que provocan la variabilidad climática del área de estudio (Agrawal, Verma, y Gupta, 2016)(Talhar, Ramgirkar, y Jena, 2022). Es así como por medio de un análisis exploratorio y estadístico descriptivo de las variables climáticas, la temperatura máxima del aire, la temperatura mínima del aire y la precipitación total ayuda a establecer las interacciones entre los dominios Espacio-Temporales de la cuenca y las proyecciones futuras establecidas por el comportamiento de los GCM (Ashraf y cols., 2021) (Gonçalves, Pereira, Costa, y Leão, 2023).

Los modelos de aprendizaje automático más utilizados y con precisiones satisfactorias para el DE son la Regresión lineal, Random Forest, Máquina de vectores de soporte y Redes neuronales artificiales (Lima, Kwon, y Kim, 2021), (Pour, Shahid, Chung, y Wang, 2018), (Karaman y Akyürek, 2023), (Muñoz, 2018), siendo estos la base elemental para mejorar la precisión de los productos de la resolución Espacio-Temporal de los GCM, que son importantes para la

comprensión de los procesos de la superficie terrestre y la variabilidad climática a nivel local (Hrisko, Ramamurthy, Yu, Yu, y Melecio-Vázquez, 2020).

Además, numerosos estudios mencionan los dos enfoques principales para la reducción de escala: dinámica y estadística. La reducción de escala estadística a nivel local, ha mostrado resultados comprometedores en términos de precisión, especialmente cuando se enfrentan limitaciones en cuanto a recursos computacionales, adquisición de datos o tiempo. Este enfoque ha sido ampliamente adaptado por varios autores en diversos climas, como se presenta en la tabla 2.1, debido a su capacidad para generar proyecciones climáticas de alta resolución a partir de los GCM.

Por otra parte (Pour y cols., 2018) propone un enfoque de reducción de escala estadística de la precipitación utilizando un modelo de Máquina de Vectores de Soporte (MVS), para la proyección de cambios espaciales y temporales en la precipitación en el país sudasiático de Bangladesh que cuenta con abundante vegetación y numerosos canales hidrológicos. El modelo de Downscaling estadístico utilizado por el autor se enfoca en Estadísticas de Salida del Modelo (ESM), donde los predictores simulados de GCM en lugar de los predictores observados se utilizan para establecer una relación estadística con el predictando observado. Luego aplica la relación estadística para la proyección usando los predictores simulados GCM CMIP5 para un escenario futuro (Turco, Llasat, Herrera, y Gutiérrez, 2017).

(Jing, Lin, Li, Li, y Shen, 2022) plantea una Red Neuronal Convolutacional (RNC) para reducir datos de precipitación mensual de China, que consta de un módulo de atención cruzada global, un módulo de atención cruzada multifactorial y un módulo convolutacional residual, que considera integralmente las posibles relaciones entre la precipitación y las características de la superficie. Esta CNN extrae características de múltiples niveles en datos de alta dimensión y captura las relaciones potenciales entre múltiples variables ambientales de las mediciones puntuales y los datos de GCM, considerando un avance significativo sobre las técnicas convencionales de aprendizaje automático en la reducción de escala (Yuan y cols., 2022).

Teniendo en cuenta el trabajo de (Muñoz, 2018), donde desarrolla un portal en la web sobre el departamento del Valle del Cauca. Este presenta la integración multiescala de precipitación y temperatura, comparando entre cuatro técnicas de inteligencia artificial: Redes Neuronales Artificiales (RNA), Máquina de Soporte Vectorial (MSV), Random Forest y un Modelo Bayesiano. Cada uno de estos modelos obtuvo niveles de correlación aceptables para la precipitación y la temperatura, pero valores bajos de exactitud debido a la falta o insuficiencia de información meteorológica local de las estaciones principalmente en la zona del Pacífico y montañosa del departamento. El modelo de IA que mejores resultados presentó en este trabajo fue el de Random Forest permitiendo estimar de forma más acertada la precipitación y la temperatura, alcanzando un nivel de correlación de 0.81 y 0.93 respectivamente, para ser usado como presentación de la

precipitación y temperatura y presentación de las proyecciones de los escenarios climáticos hasta el 2100 para su portal en la web.

Los métodos estadísticos tradicionales presentan diferentes dificultades para reducir los datos de precipitación y varían significativamente en términos de precisión y confiabilidad, ya que las variables climáticas locales, como la precipitación, dependen de procesos Espacio-Temporales complejos y no lineales (Chou y cols., 2021). Es así como (Chou y cols., 2021) desarrolla una red neuronal de memoria a corto plazo convolucional basada en Super-Resolución generando proyecciones climáticas de alta resolución para la precipitación para datos a 0.25° , probando la solidez y la previsibilidad de este modelo en datos de precipitación mensual en China. El autor concluye que el modelo Super-Resolution ConvLSTM tiene mayor precisión en comparación con otros modelos de Redes Neuronales Convolucionales, aumentando significativamente la precisión de la predicción de precipitaciones tanto en promedios como en eventos extremos de precipitación que es de gran problemática mencionado en toda la literatura consultada.

En la Tabla 2.1 se destacan varios autores que han investigado y aplicado diferentes técnicas de aprendizaje automático para mejorar la resolución Espacio-Temporal de los GCM. Entre estas técnicas se incluyen redes neuronales artificiales, regresión multivariada, y métodos avanzados como el Deep Learning, que han sido reconocidos por su capacidad para modelar de manera efectiva las complejas relaciones no lineales en los datos climáticos. Estas investigaciones no solo subrayan la relevancia de la reducción de escala estadística, sino que también demuestran cómo estas técnicas han sido cruciales para abordar los desafíos del cambio climático con enfoques innovadores y eficientes, optimizando la precisión de los pronósticos a nivel local o regional.

El *downscaling* estadístico actualmente ha evolucionado significativamente con la integración de técnicas de inteligencia artificial, permitiendo mejorar la resolución espacial y temporal de los modelos climáticos globales (GCM) y adaptarlos a escalas locales. Los estudios mencionados anteriormente han demostrado la eficacia de modelos como RL, RF, MVS y RNA, destacando en particular los enfoques basados en aprendizaje profundo, como las RNC y las arquitecturas híbridas con memoria a corto plazo (ConvLSTM). Sin embargo, persisten desafíos en la modelización de la precipitación debido a la complejidad de sus patrones no lineales y a la limitada disponibilidad de datos meteorológicos locales en algunas regiones.

Considerando estos avances y limitaciones, el presente estudio adopta una estrategia basada en redes ConvLSTM2D para mejorar la representación de la temperatura y la precipitación en la costa del Pacífico colombiano, buscando optimizar la precisión del *downscaling* estadístico en la región. La implementación de esta metodología permitirá capturar mejor la variabilidad climática espacio-temporal, contribuyendo al desarrollo de proyecciones climáticas más precisas y útiles para la planificación y gestión del cambio climático a nivel regional.

Tabla 2.1: Cuadro de referencias utilizadas en la investigación, con los modelos, variables y resoluciones abordadas por cada autor.

REFERENCIAS DE INVESTIGACIÓN					
Modelo	Autor	Variable	Resolución	Validación	Año
CNN	Kheir et al.	Temperatura	10 Km Mensual	RECM CC D. Taylor	2023
RLM SWG LSTM	Fouotsa, Potgieter, Watson, Nel	Temperatura Precipitación	Puntual Mensual	RECM PEAM NSE	2023
RF	KaramanAkyürek	Temperatura	0.05° Diaria	RECM EAM CC	2023
RLM MSV RF K-NN CNN	Lei Su	Precipitación	Puntual Mensual	Validación Cruzada D. Taylor	2023
CNN DL4DS	Gomez Carlos	Multiples Variables	Múltiples escalas	RECM EAM SSIM Pearson	2023
GAN ConvGRU	Harder P.	Precipitación	Multiples escalas	RECM EAM SSIM MS-SSIM	2023
LSTM SVM	Li Xinyi et. al.	Temperatura	Diaria	RECM CC	2021
Interpolación L. SRCNN DeepSD	Kumar et. al.	Precipitación	Diaria	ECM Pattern correlation	2021
LSTM	Chou	Precipitación	Multiples escalas	RECM Correlacion	2021
CNN	Baño M.	Temperatura Precipitación	Múltiples escalas	RECM EAM CC	2020
LSTM FNN	Ahn Duong, et. al	Precipitación	Diaria	RECM EMS Correlacion	2019
RF MSV RNA	Muñoz	Temperatura Precipitacion	Diaria 0.05°	RECM EPMA CC D. Taylor	2018

Capítulo 3

METODOLOGÍA

En este capítulo se describen los datos y el marco experimental utilizados en el análisis desarrollado para esta tesis de investigación. La investigación ha sido diseñada en el marco del proyecto de convocatoria interna C21 de Ciencia, Tecnología e Innovación (CTel), orientado a la modelización climática de la región pacífica del departamento del Valle del Cauca, la cual se introduce en la sección 3.1.

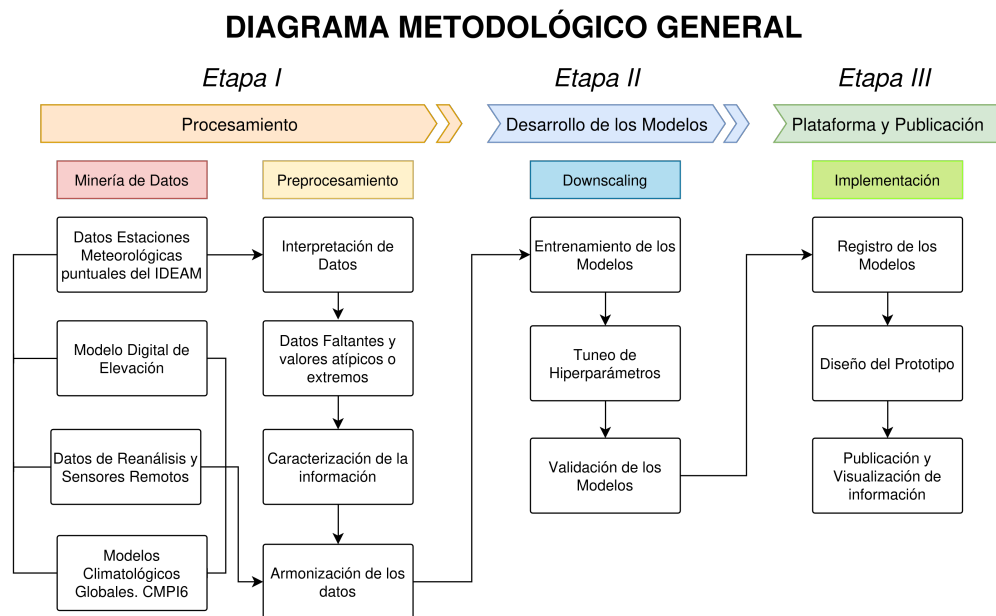


Figura 3.1: Diagrama de flujo de la metodología desarrollada.

Fuente: elaboración propia.

La sección 3.2 presenta el marco experimental implementado, explicando paso a paso cómo

se aborda la problemática planteada. Se detallan los datos utilizados, que incluyen observaciones recopiladas por estaciones meteorológicas, estimaciones de datos de reanálisis de ERA5-Land y resultados de Modelos de Circulación General. Además, se describe el proceso de post-procesamiento de datos, así como la selección de variables predictoras y de respuesta empleadas para el entrenamiento de los modelos. Esta selección se basa en un análisis exploratorio y descriptivo de los datos. Posteriormente, se presentan los distintos modelos utilizados, tanto los tradicionales como RL, RF, RVS y los más avanzados, desarrollados mediante técnicas de inteligencia artificial, como redes neuronales artificiales de aprendizaje profundo *Deep Learning* LSTM, para evaluar de manera comparativa el Downscaling Estadístico. Finalmente, se exponen las métricas empleadas para la evaluación y validación de los modelos de reducción de escala, comparándolos con las estimaciones de reanálisis y las mediciones observadas a nivel local con las estaciones meteorológicas.

La última sección, 3.3, describe el formato de los archivos de almacenamiento de las variables y la estructura del diseño modular de la plataforma web para la presentación y publicación de los datos proyectados de temperatura y precipitación bajo diferentes futuros climáticos.

3.1. Localización

El área de estudio seleccionada comprende la costa del departamento del Valle del Cauca, en estudio a detalle para la cuenca del río Dagua. Esta cuenca cuenta con dos subregiones importantes: la Zona Baja y la Zona Alta sobre la cordillera occidental. En esta última nacen los principales drenajes que desembocan en la costa del océano Pacífico, Figura 3.2.

La cuenca se caracteriza por la presencia de bosque húmedo tropical y bosque seco en la parte baja, mientras que en la zona alta predominan el bosque húmedo y de niebla, extendiéndose por los municipios de Dagua, La Cumbre, Vijes, Restrepo y Buenaventura. El área de estudio también incluye otros municipios del sur del departamento.

3.2. Datos utilizados

Para la realización del proyecto primero presentan los datos de los Modelos de Circulación Global (GCM). Luego se presentan los datos de temperatura y precipitación de ERA5-Land, que funcionan como predictandos o variables de respuesta en los modelos de regresión. Finalmente, se exponen los datos observados por las estaciones meteorológicas distribuidas en la zona de estudio. Estos valores se consideran como referencia 'real' de las variables y se utilizarán para la validación de los modelos.

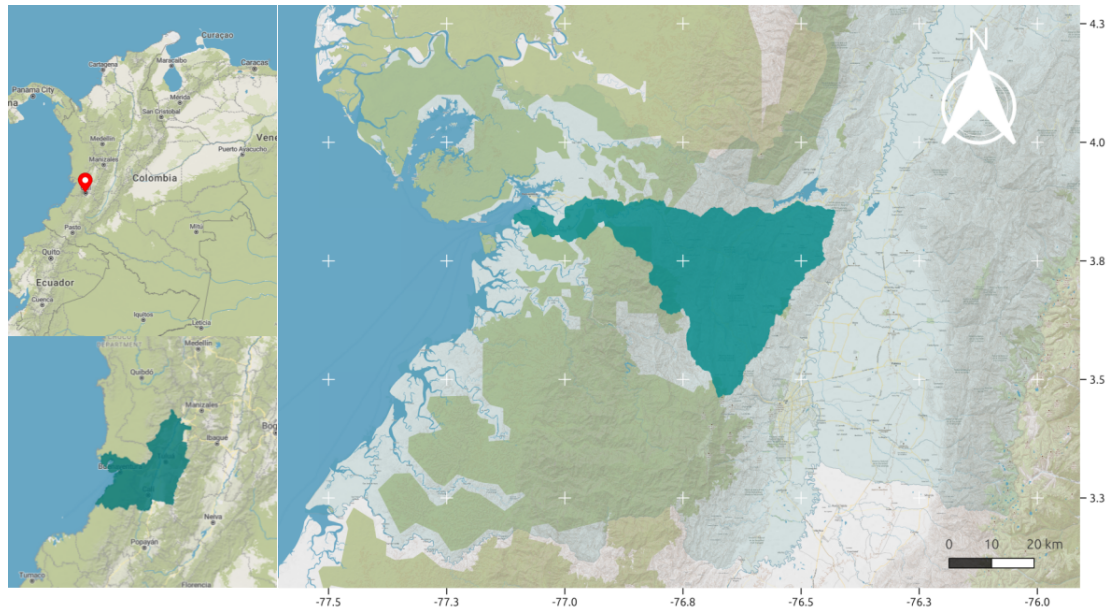


Figura 3.2: Región Pacífica del departamento del Valle del Cauca, ubicación de la cuenca del río Dagua.

3.2.1. GFDL Earth System Model version 4.1 (GFDL-ESM4)

El GFDL-ESM4 es un sofisticado modelo del sistema terrestre, lanzado en 2018 por la agencia NOAA de los Estados Unidos (NOAA, 2024). Este modelo forma parte de la cuarta generación de modelos GFDL y ha contribuido significativamente al CMPI6. Su complejidad radica en la integración de diversos componentes que simulan la atmósfera, los océanos y la superficie terrestre para recrear el clima a nivel superficial. Este modelo tiene gran capacidad para generar datos que abarcan desde la época preindustrial, pasando por el período histórico (1850-2014), hasta proyecciones futuras desde 2015 hasta el 2100 en diferentes variables climáticas. Este modelo trabaja a una resolución espacial de 1 grado (aproximadamente 100 km) como se puede apreciar en la Figura 3.3 y ofrece datos con una resolución temporal de tres horas.

Se seleccionó el modelo global GFDL debido a que es uno de los pocos que ofrece datos con una resolución temporal de tres horas, lo que lo convierte en una opción ideal para estudios que requieren alta frecuencia temporal, como la propuesta en esta investigación. Además, varios autores, como (Ortega y cols., 2021), (Reboita, Ferreira, Ribeiro, y Ali, 2024), han evaluado su precisión en estudios específicos sobre Sur América, mostrando resultados favorables en comparación con otros modelos globales.

Esta investigación adopta todas las variables climáticas del GFDL-ESM4 a través de la plataforma Earth System Grid Federation (ESGF) (Williams y cols., 2016). Esta plataforma es

una fuente importante para los investigadores del cambio climático, ya que alberga una extensa colección de simulaciones, observaciones y datos de diferentes instituciones alrededor del mundo.

El modelo cuenta con variables climáticas listadas en la tabla 3.1 y disponibles en los escenarios climáticos proyectados:

- SSP 1 - 1.9: representa un mundo sostenible con bajos niveles de emisiones, donde se logra un forzamiento radiativo de 1.9 W/m^2 para 2100. Este escenario está alineado con el objetivo más ambicioso del Acuerdo de París de limitar el calentamiento global a 1.5°C sobre los niveles preindustriales.
- SSP 1 - 2.6: también describe un futuro sostenible con esfuerzos significativos hacia la mitigación, pero con un forzamiento radiativo de 2.6 W/m^2 para 2100. Corresponde a un calentamiento global de aproximadamente 2°C , y aunque menos ambicioso que SSP1-1.9, sigue implicando importantes reducciones de emisiones.
- SSP 2 - 4.5: refleja un camino intermedio, donde las tendencias actuales de desarrollo, políticas y emisiones continúan con modestos esfuerzos de mitigación. Este escenario proyecta un forzamiento radiativo de 4.5 W/m^2 para 2100 y un aumento de temperatura entre 2.5 y 3°C , considerando una combinación de progreso y desafíos en sostenibilidad.

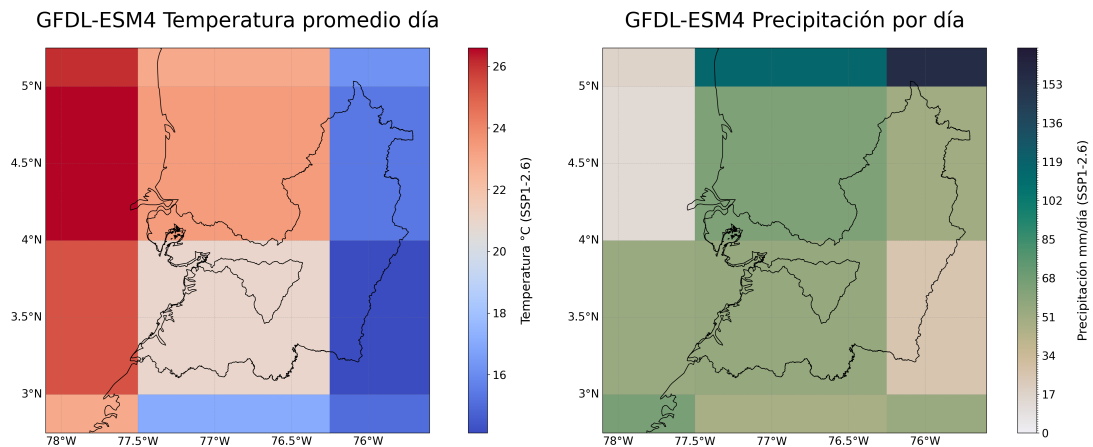


Figura 3.3: Temperatura y precipitación del GCM GFDL-ESM4 para la región Pacífica del departamento del Valle del Cauca. Resolución espacial de 1° y resolución temporal de 3 horas.

Fuente: (Dunne y cols., 2020).

Tabla 3.1: Lista de variables climáticas del GCM GFDL-ESM4 utilizadas en esta investigación. Todas las variables son estimaciones cercanas a la superficie con una resolución espacial de 1° y resolución temporal de 3 horas.

Variable Climatológica	código	Nombre de la variable	Unidades
Temperatura	tas	Near-Surface Air Temperature	K
Precipitación	pr	Precipitation	Kg m-2 s-1
Componente este del viento	uas	Eastward Near-Surface Wind Speed	m s-1
Componente norte del viento	vas	Northward Near-Surface Wind Speed	m s-1
Humedad específica	huss	Near-Surface Specific Humidity	Kg/Kg
Humedad del suelo	mrsos	Moisture in Upper Portion of Soil Column	Kg m-2
Presión	sp	Surface Air Pressure	Pa
Radiación onda corta descendente	rlds	Surface Downwelling Longwave Radiation	W m-2
Radiación onda corta ascendente	rlus	Surface Upwelling Longwave Radiation	W m-2
Radiación onda larga descendente	rsds	Surface Downwelling Shortwave Radiation	W m-2
Radiación onda corta ascendente	rsus	Surface Upwelling Shortwave Radiation	W m-2

Fuente: (Dunne y cols., 2020)

3.2.2. Fifth Generation European ReAnalysis (ERA5-Land)

El proyecto ERA5-Land modela un grupo importante de variables climáticas globales; enmarcado en el Servicio de Cambio Climático de Copernicus (C3S) de la Comisión Europea. El Centro Europeo de Pronóstico Meteorológico de Mediano Plazo (ECMWF), ofrece un conjunto de variables climáticas y ambientales que cubren todo el planeta, a nivel de la superficie terrestre. Este modelo ofrece una alta precisión, validada mediante observaciones in situ y datos satelitales. Su cobertura temporal abarca desde el año 1950 hasta el presente, con actualizaciones constantes que proporcionan datos climáticos por hora, a una resolución espacial de 0.1° como se muestra en la Figura 3.4. Su periodo de propagación es continuo en espacio y tiempo, convirtiéndose en un conjunto de datos valioso para el apoyo de esta investigación (Baker y cols., 2021), (Lyra Fialho Breda, Dias de Paiva, Chou, y Collischonn, 2022). Los datos de ERA5-Land fueron descargados desde su página oficial Europea (Copernicus Climate Change Service, 2024).

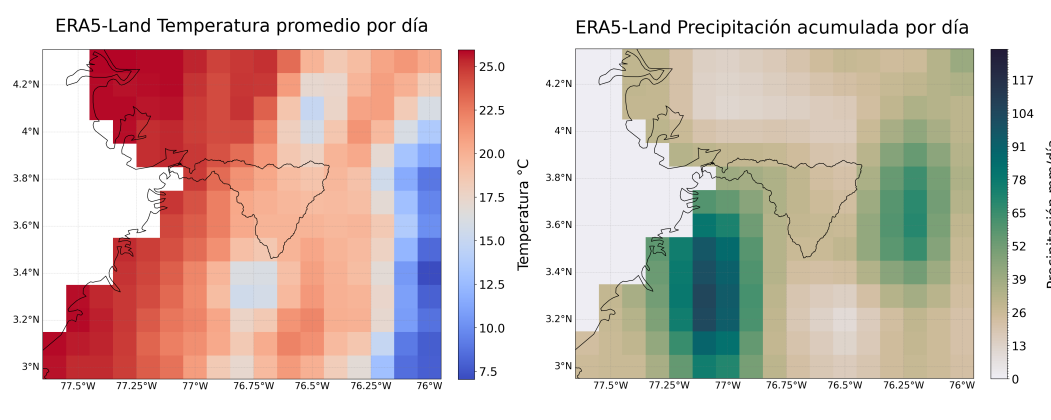


Figura 3.4: Temperatura promedio y precipitación acumulada de ERA5-Land, para el mes de mayo de 2015 en el Pacífico del departamento del Valle del Cauca. Datos con una resolución espacial de 0.1° y temporal de 1 hora.

Fuente: (ECMWF, 2019).

Tabla 3.2: Descripción de las variables temperatura y precipitación de ERA5-Land utilizadas en esta investigación.

Variable Climatológica	código	Nombre de la variable GCM	Unidades
Temperatura	t2m	2 m temperature	K
Precipitación	tp	Total precipitation	m

Fuente: (ECMWF, 2019)

3.2.3. Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM)

Los datos de validación utilizados en este estudio provienen de las estaciones automáticas del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), obtenidos a través de su plataforma especializada Aquarius (IDEAM-AQUARIUS, 2024). Esta plataforma permite el acceso a información hidrometeorológica en tiempo real, centralizando datos de estaciones meteorológicas distribuidas a lo largo del país. En el área de estudio, se identificaron siete estaciones automáticas, ver tabla: 3.3, cada una de las cuales registra variables clave como temperatura y precipitación con una frecuencia horaria. Estas observaciones detalladas y de alta resolución temporal son esenciales para realizar un análisis climático local, pero con una resolución espacial puntual, permitiendo no solo el monitoreo de condiciones actuales, sino también la evaluación y validación de modelos climáticos globales (GCM) y regionales (ERA5-Land). Los datos observados son considerados

como "Valores reales" los cuales también son utilizados como referencia para ajustar y contrastar los resultados obtenidos en los diferentes modelos desarrollados en esta investigación, aportando confiabilidad y rigor científico al análisis.

Tabla 3.3: Lista de estaciones climatológicas del IDEAM disponibles. Toma de registros de datos con frecuencia horaria.

Estación (Código)	Longitud	Latitud	Elevación
AEROPUERTO A. BONILLA - AUT [26075150]	-76° 22' 56"	3° 31' 58"	961 m
FARALLONES - AUT [26055100]	-76° 39' 5"	3° 24' 57"	2275 m
LA CUMBRE - AUT [5311500121]	-76° 33' 53"	3° 38' 42"	1613 m
LA DIANA - AUT [26075120]	-76° 11' 8"	3° 18' 50"	1615 m
SILOE - AUT [23085250]	-76° 33' 38"	3° 25' 31"	1238 m
UNIVERSIDAD DEL PACÍFICO - AUT [5311500056]	-76° 59' 13"	3° 50' 53"	16 m
UNIVERSIDAD DEL VALLE - AUT [26055120]	-76° 32' 2"	3° 22' 40"	996 m

Fuente: (IDEAM-AQUARIUS, 2024)

3.2.4. Datos de elevación

Los datos de elevación fueron obtenidos a partir del Modelo Digital de Elevación (MDE) de alta resolución espacial (10 m), descargados a través de la plataforma de la Alaska Satellite Facility (ASF), utilizando el sensor ALOS PALSAR (JAXA, [2024]). Este sensor, que opera en la banda L de radar, es especialmente valorado por su capacidad para capturar información topográfica precisa, incluso en zonas con densa vegetación o bajo condiciones meteorológicas adversas como se presenta en la costa del Pacífico.

Distribución de las estaciones locales

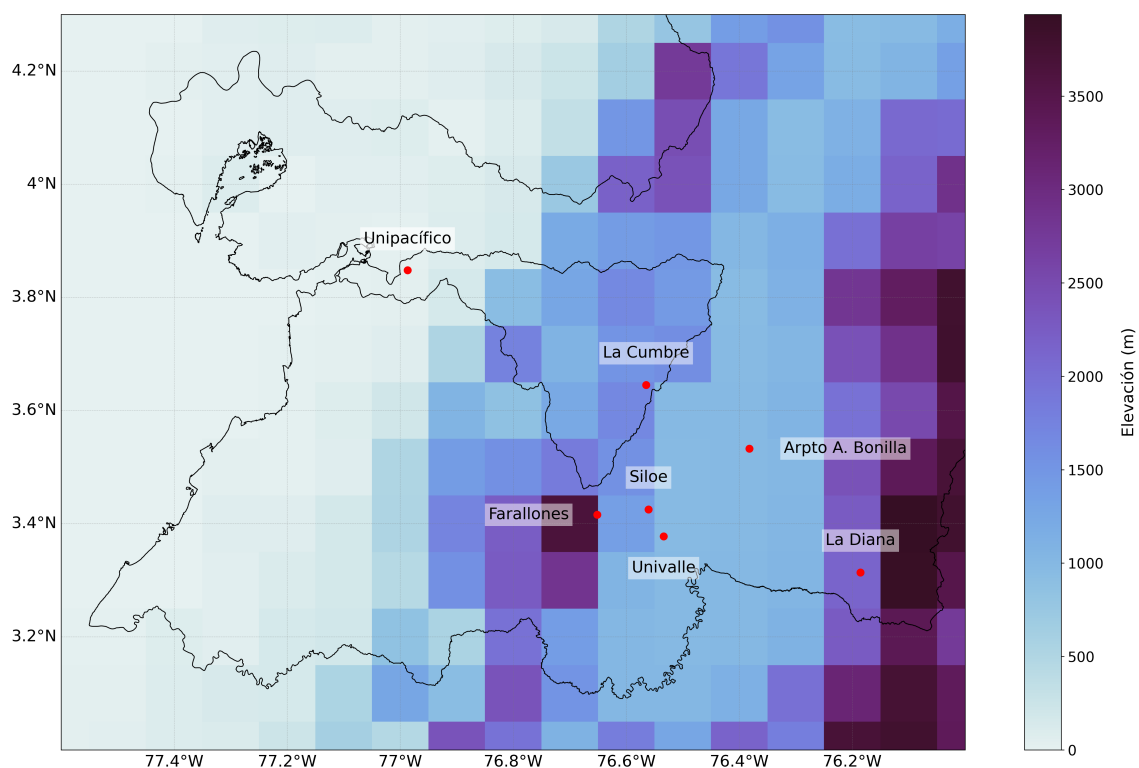


Figura 3.5: Mapa de distribución espacial de las estaciones meteorológicas automáticas en la zona de estudio, superpuesto sobre el MDE de Alos-Palsar. La escala de colores representa la altitud del terreno, desde 0 metros, hasta más de 3500 metros. Las estaciones, distribuidas para capturar la variabilidad climática altitudinal de la región. Los datos climatológicos registrados por estas estaciones son con frecuencia horaria.

Fuente: (JAXA, [2024])

3.3. Marco experimental

La metodología planteada se resume en el diagrama de la figura 3.1, y se organiza en tres niveles o etapas, cada una orientada al cumplimiento de los tres objetivos específicos establecidos.

En la primera etapa, se describen los datos utilizados y el preprocesamiento aplicado a estos. Esto incluye un análisis exploratorio y descriptivo de todas las variables que se emplearán en el modelamiento climático.

La segunda etapa detalla el proceso de desarrollo de los modelos. Se comienza con la selección del horizonte temporal, la partición de los datos, la armonización de estos, y la construcción de los diferentes modelos que serán utilizados.

Finalmente, se diseña una plataforma web destinada a la publicación de los resultados obtenidos, incluyendo proyecciones de temperatura y precipitación hasta el año 2035, en los escenarios del IPCC6 1.19, 1.26 y 2.45.

3.3.1. Horizonte de tiempo

El horizonte temporal del análisis, que abarca de 2015 a 2035, se define considerando tres aspectos clave que aseguran tanto la robustez científica como la viabilidad operativa del estudio.

Primero, se selecciona un período histórico de 9 años (2015-2023) como base para la calibración del modelo, dado que en estudios climáticos es ampliamente recomendado utilizar series temporales de aproximadamente una década. Este intervalo permite captar la variabilidad climática a corto y mediano plazo, proporcionando una referencia para evaluar las condiciones actuales.

Por otro lado, el GCM utilizado en el estudio genera sus proyecciones climáticas a partir de observaciones recopiladas hasta el año 2015. Este hecho establece una limitación temporal en los datos disponibles, dado que las simulaciones posteriores a ese año se basan exclusivamente en modelos de proyección.

Finalmente, el enfoque adoptado, que utiliza datos con resolución horaria, implica un volumen masivo de registros y un procesamiento intensivo, lo que demanda un considerable esfuerzo humano y recursos computacionales. Limitar el horizonte temporal desde 2015 hasta 2035 equilibra la necesidad de generar proyecciones climáticas detalladas y prácticas con las restricciones operativas de los recursos disponibles, garantizando así la calidad y la viabilidad del análisis.

3.3.2. Análisis exploratorio y descriptivo de datos climatológicos

El análisis exploratorio y descriptivo aplicado a los datos climatológicos se enfoca en la extracción de información representativa de cada conjunto de variables, como la temperatura y la precipitación, transformando estos grandes volúmenes de datos en representaciones visuales más comprensibles. Este enfoque permite identificar de manera detallada los patrones y el comportamiento de las variables a lo largo del tiempo (Ashraf y cols., 2021), (Gonçalves y cols., 2023). Debido a la gran cantidad de datos y su dispersión espacial, este proceso debe ser automatizado o semi-automatizado, asegurando que los patrones identificados sean significativos y de alta calidad, orientando así el análisis hacia el descubrimiento de nueva información relevante y precisa.

El análisis exploratorio y descriptivo de los datos se realizó comparando las variables de temperatura y precipitación registradas en cada estación meteorológica con los datos correspondientes al píxel coincidente en los modelos GFDL-ESM4 y ERA5-Land. Este enfoque

permite evaluar la concordancia e identificar las diferencias entre las observaciones in situ y las estimaciones generadas por los modelos, considerando diferentes condiciones topográficas.

La Figura 3.6 ilustra el procedimiento seguido en este análisis. Para cada variable, se desarrollaron diversas estadísticas y gráficos descriptivos que permitieron una comprensión más detallada de los patrones y tendencias presentes en los datos. Con el fin de mantener la claridad del documento, se presentan tablas con el análisis de todas las estaciones. Sin embargo, los gráficos se limitan a tres estaciones representativas de la cuenca.

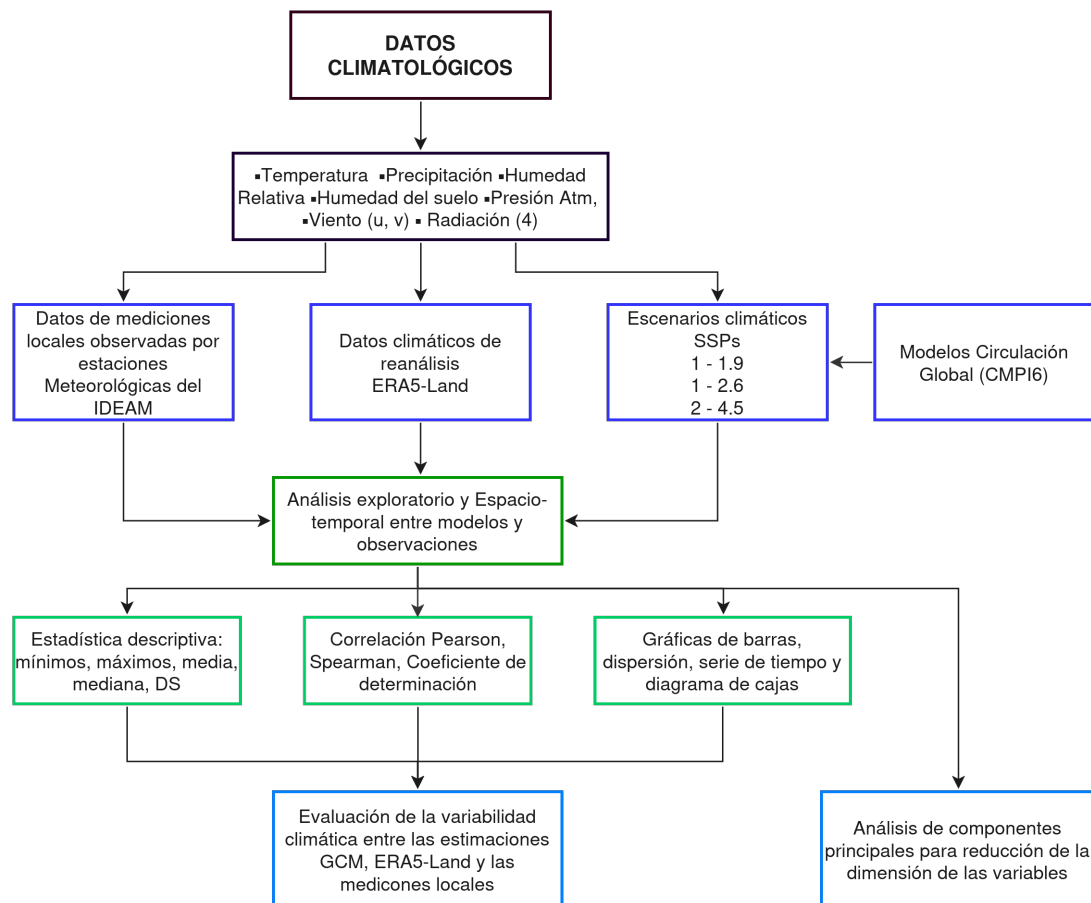


Figura 3.6: Diagrama del procedimiento desarrollado en el análisis exploratorio y descriptivo de los datos y diferentes variables utilizadas.

Fuente: elaboración propia.

3.3.2.1. Distribución espacio-temporal homogénea

Para el análisis de la temperatura y la precipitación, se ajustaron primero las unidades de medición y el tiempo para que las estimaciones de los modelos GFDL-ESM4 y ERA5-Land coincidieran con los datos observados de las estaciones locales. En ambos modelos, la temperatura se transformó de grados Kelvin a grados Celsius, mientras que la precipitación en ambos modelos se convirtió a milímetros por hora (*mm/hora*). Además, el tiempo se ajustó a la hora local, ya que los modelos utilizan el huso horario UTC-00 (Universal Time Coordinated).

Se armonizó la escala espacial de todos los datos, estableciendo una resolución deseada de 0.1°. Para los datos de los diferentes escenarios del GCM, que tienen una resolución más baja, se aplicó una interpolación bicúbica para alcanzar la escala deseada de 0.1°. Asimismo, para el Modelo Digital de Elevación (DEM), que cuenta con una resolución muy detallada, se realizó un remuestreo con el mismo método bicúbico para unificar resoluciones espaciales, como se muestra en la Figura 3.5.

3.3.2.2. Correlación y descripción de los datos

La evaluación de la variable temperatura y precipitación se fundamentó en un análisis de la correspondencia entre tres fuentes de datos: el GCM GFDL-ESM4, el reanálisis ERA 5-Land y las observaciones in situ de las estaciones meteorológicas. El estudio incorporó un análisis de correlación espaciotemporal complementado con estadísticas descriptivas para cuantificar el grado de concordancia entre las fuentes.

La caracterización del comportamiento de la temperatura se realizó mediante el análisis de las series temporales, lo que permitió identificar patrones de variabilidad y tendencias. Esta evaluación se complementa con diagramas de dispersión que evidencian la coherencia entre las diferentes fuentes de datos (Agrawal y cols., 2016).

Al realizar una escala temporal mensual multianual en el análisis de precipitación, se representa una aproximación metodológica del clima. Esta escala temporal permite una mejor caracterización de los patrones estacionales y ciclos anuales, elementos esenciales para comprender el comportamiento climático regional, mientras que efectivamente filtra la alta variabilidad inherente a las observaciones horarias. La agregación mensual no solo facilita la detección de anomalías climáticas significativas, sino que también fortalece la confiabilidad de los análisis estadísticos, incluyendo las correlaciones entre diferentes fuentes de datos (De, Adacheski, y Virgens Filho, 2011).

Esta metodología está respaldada por literatura científica consultada, donde la escala mensual se ha establecido como un estándar en la validación de modelos climáticos y productos de reanálisis, permitiendo comparaciones consistentes entre diferentes estudios y estableciéndose como un marco

temporal óptimo para la evaluación del desempeño de modelos climáticos (Rajanikanth y Kanth, 2017), (El Qorchi y cols., 2023), (Ghaderpour y cols., 2023). Por último se realizó el análisis del ciclo horario durante el período de estudio, revelando los patrones característicos de la evolución térmica y de precipitación horaria en la región.

La identificación del escenario más apropiado del modelo GFDL-ESM4 se fundamentó en una evaluación sistemática de los escenarios SSP 1.19, SSP 2.45 y SSP 3.7 contrastando con las estimaciones ERA5-Land y las observaciones de estaciones meteorológicas. La metodología de selección incorporó un análisis de correlación múltiple, empleando los coeficientes de Pearson para evaluar relaciones lineales, Spearman para capturar asociaciones monótonas no necesariamente lineales, y el coeficiente de determinación (R^2) para cuantificar la varianza explicada (Q. Yang, Zheng, y Liu, 2016).

Esta aproximación multiparamétrica se complementó con un análisis estadístico que incluyó la caracterización de la distribución de los datos mediante estadísticos de posición (mínimos, máximos, media, mediana), dispersión (desviación estándar) y forma (cuartiles Q1 y Q3). Este marco metodológico permitió una evaluación objetiva de la capacidad de cada escenario para reproducir las condiciones observadas, proporcionando una base sólida para la selección del escenario más apropiado para el posterior proceso de *downscaling*.

3.3.2.3. Limpieza de datos, datos faltantes y atípicos

Mediante el análisis exploratorio que incluyó correlaciones, estadísticas descriptivas y visualización de gráficas, se determinó que los datos de los modelos GFDL-ESM4 y ERA5-Land no presentan huecos o datos faltantes. Dado que estos modelos generan datos basados en procesos físicos y parametrizaciones específicas, se considera a los valores que superan las 3 desviaciones estándar en los diagramas de caja como datos atípicos influyentes ni errores, sino como posibles representaciones de eventos climáticos extremos. Además, estos valores constituyen menos del 1 % del conjunto total de los datos, por lo que eliminarlos podría comprometer la representación de eventos extremos en la zona de estudio para el entrenamiento de los modelos. En consecuencia, se opta por conservarlos y modelar con ellos, generando así un análisis más consistente y alineado, respetando la naturaleza de los datos climáticos simulados.

Se realiza el ajuste en los datos de ERA5-Land eliminando los días bisiestos, ya que los GCM no los considera en sus modelaciones. Los valores nulos de la cuadrícula se eliminan en los entrenamientos de RL, RF y RVS; para el uso de las redes neuronal ConvLSTM2D los valores nulos de temperatura se transforman a 0.0 y los de precipitación a -99 para que no sean tenidos en cuenta por el entrenamiento.

En cuanto a los datos de las estaciones climatológicas, aunque no presentan valores atípicos, contienen algunos datos faltantes y no abarcan la totalidad del período de estudio debido a

fallos en las estaciones o interrupciones en la toma de datos. Estos datos faltantes se eliminan tras un proceso de revisión realizado por el IDEAM. Es importante señalar que no se realizaron imputaciones para los datos faltantes, ya que se consideran como el "valor real" de las variables y son fundamentales para la validación de las predicciones de los modelos, permitiendo seleccionar el que mejor se ajusta a la realidad observada.

3.3.3. Componentes principales

Se implementó un análisis de componentes principales (ACP) para las nueve variables climáticas proporcionadas por el modelo GCM (uas, vas, huss, mrsos, sp, rlds, rlus, rsds, rsus), detalladas en la tabla 3.1. El ACP es una técnica estadística ampliamente utilizada para reducir la dimensionalidad de conjuntos de datos al transformar las variables originales en un nuevo conjunto de variables linealmente combinadas, denominadas componentes principales. Estas componentes capturan la mayor parte de la varianza presente en los datos, preservando la información más relevante. En este análisis, se seleccionaron aquellas componentes principales que, de forma acumulativa, explican al menos el 95 % de la varianza total. Este umbral garantiza una representación simplificada del conjunto de datos, manteniendo su integridad informativa y minimizando la redundancia entre las variables originales (Tadić, Bonacci, y Brleković, 2019), (Schoof y Pryor, 2001).

Esta reducción no solo disminuye la dimensionalidad del espacio de características, sino que también ayuda a mitigar la multicolinealidad entre las variables originales, lo que resulta en modelos más eficientes y estables. Las componentes principales seleccionadas se utilizan en el entrenamiento de cada uno de los modelos para evaluar su rendimiento y contribución a las predicciones, optimizando el proceso de modelado al enfocarse en las variaciones más relevantes de los datos (Erlandsen, Parding, Benestad, Mezghani, y Pontoppidan, 2020), (Cao, Cao, Li, y Zhu, 2023).

3.3.4. Escalamiento de datos

En el proceso de preparación de los datos, para mejorar y optimizar el entrenamiento de los modelos, se aplican técnicas de escalamiento como la estandarización y normalización de la temperatura y la precipitación (Zhao y Wu, 2023) (Gu y cols., 2023), (Hernanz y cols., 2023). Este preprocesamiento de las variables mejora la eficiencia y velocidad de convergencia de los modelos de aprendizaje automático y redes neuronales profundas, mitigando el riesgo de que cierta variable domine indebidamente el modelo debido a diferencias en sus escalas.

3.3.4.1. Estandarización en la temperatura

Para la variable temperatura, se aplica la técnica de estandarización, dado a que esta variable, presenta una distribución aproximadamente normal o Lognormal. Este proceso de transformar los datos para que tengan una media $\mu = 0$ y una desviación estándar de $\sigma = 1$, siguiendo la ecuación 3.1

$$z = \frac{x - \mu}{\sigma} \quad (3.1)$$

Donde:

z es el valor estandarizado

x : valor original

μ : media de la distribución

σ : desviación estándar de la distribución

3.3.4.2. Normalización en la precipitación

Por otra parte la precipitación se somete al proceso de normalización en el intervalo $[0, 1]$. Esta elección se realiza ya que la precipitación no sigue una distribución normal en los datos, mostrando una distribución asimétrica con una concentración en valores cercanos a 0 y valores extremos dispersos. La normalización se realiza mediante la ecuación 3.2

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.2)$$

Donde:

x_{norm} es el valor normalizado

x : valor original

x_{min} : valor mínimo del conjunto de datos

x_{max} : valor máximo del conjunto de datos

3.3.5. *Downscaling* estadístico

La metodología de *downscaling* estadístico se plantea como un problema de regresión supervisado. En este enfoque se utilizan como entrada las variables de los GCM de resolución gruesa para el entrenamiento de los modelos, mientras que los datos más finos de ERA5-Land sirven como valores objetivos. El propósito es poder representar con mayor precisión las variaciones climáticas en el área de estudio, considerando sus condiciones geoespaciales de longitud, latitud y altura. De este modo permitir proyectar las variables de temperatura y precipitación con mayor

detalle para el periodo 2024 - 2035 (Mehta, Yadav, Ladavia, y Caloiero, 2023), (Van der Meer, de Roda Husman, y Lhermitte, 2023).

En este estudio se emplean arquitecturas de capas ConvLSTM2D (T. Xu, Guan, Peng, Wei, y Zhao, 2021), (Rampal y cols., 2022), (Domala y Kim, 2022) para realizar el mapeo como reducción espacial, pasando de 1° a 0.1° de tamaño de píxel. Además, estas arquitecturas se utilizan para predecir nuevos valores, reduciendo la resolución temporal de 3 horas a 1 hora para las variables climáticas de temperatura y precipitación. El resultado de estos modelos se compararon con otros métodos de regresión comúnmente utilizados en el Downscaling estadístico, como la Regresión Lineal (RL), Random Forest (RF) y Regresión de vectores de soporte (RVS) con base en diversas métricas de validación.

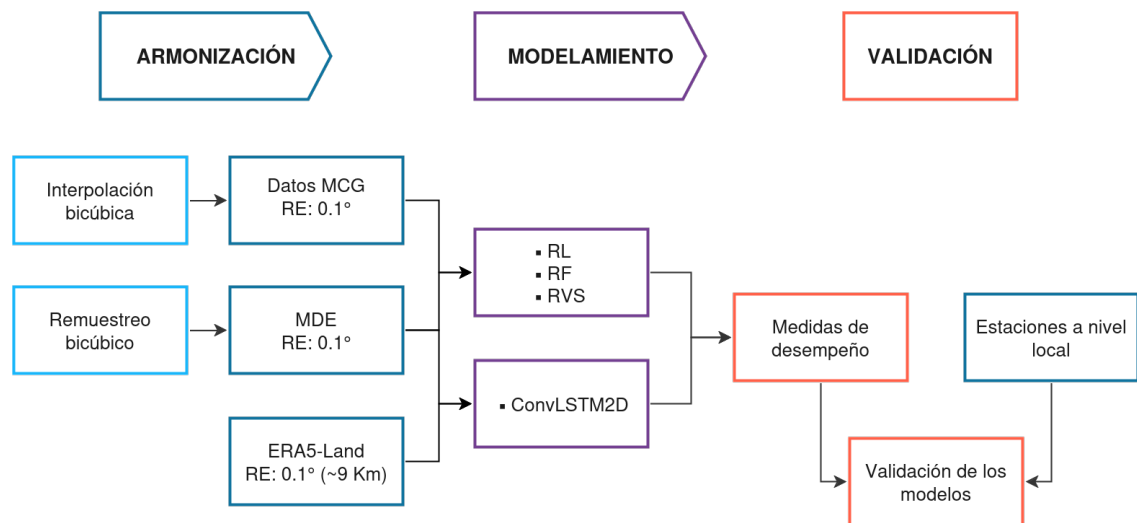


Figura 3.7: Diagrama metodológico del proceso del *Downscaling* estadístico

Fuente: elaboración propia.

3.3.5.1. Interpolación bicúbica

Debido a la gruesa escala espacial de los datos del GCM, se emplea una interpolación bicúbica para suavizar los datos a la misma dimensión de los datos de respuesta de ERA5-Land (Admasu, Grant, y Thiery, 2023). La interpolación bicúbica es un método avanzado que ofrece resultados de alta calidad en el proceso de imágenes y datos geospaciales.

Aunque la interpolación bicúbica es conocida por su precisión, también es computacionalmente intensiva en grandes cantidades de datos. Sin embargo, en este caso, dado que solo se utilizan 9 píxeles de entrada para generar cada píxel de salida, el proceso no requiere un tiempo de cómputo excesivo.

La elección del uso de este método de interpolación es crucial para mantener la integridad de los datos del GCM al cambiar su resolución, ya que preserva mejor las características de los datos originales en comparación con otros métodos bilineales o vecinos más cercanos (Y. Zhang y cols., 2021).

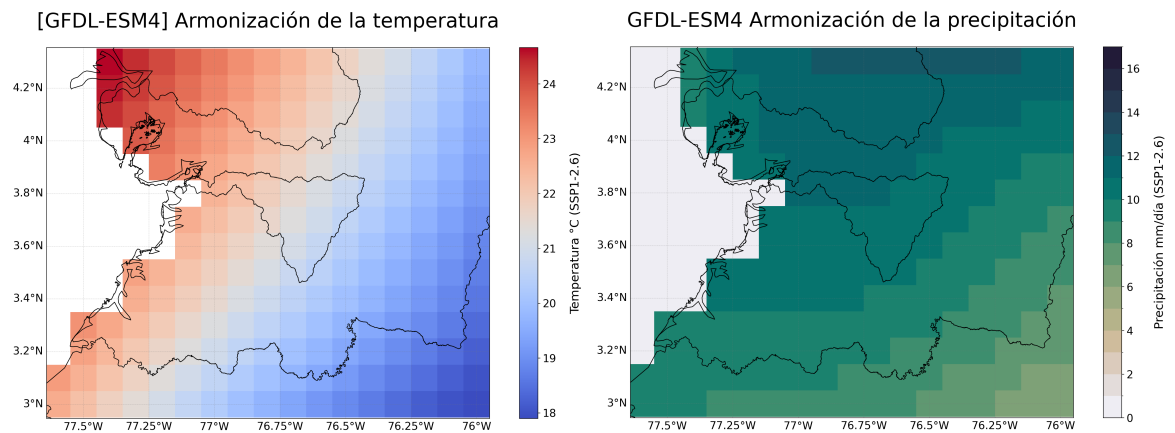


Figura 3.8: Armonización de la resolución espacial del modelo GFDL-ESM4 a resolución espacial de datos ERA5-Land utilizando el método de interpolación bicúbica.

Fuente: elaboración propia.

3.3.5.2. Partición de los datos para el entrenamiento y prueba de los modelos

La partición del conjunto de datos para ser utilizados en los modelos de RL, RF y MSV se dividieron en dos subconjuntos, uno de entrenamiento con el 80 % de los datos y el otro conjunto del 20 % para prueba para evaluar el rendimiento con datos no vistos en el entrenamiento y de este modo evitar el sobreajuste.

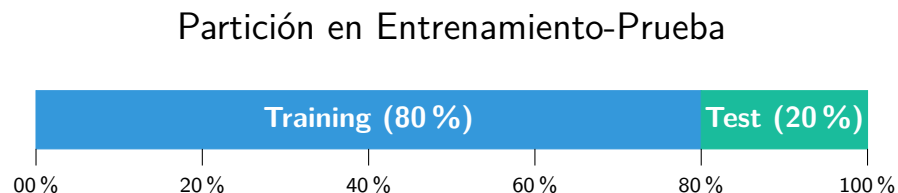


Figura 3.9: Visualización de la partición de datos en conjuntos de entrenamiento y prueba para los modelos RL, RF y MSV

Fuente: elaboración propia.

3.3.5.3. Selección de variables de entrenamiento y objetivo

En todos los modelos desarrollados se utilizaron como variables de entrada constantes las componentes geográficas: longitud, latitud y altura. Estas se etiquetan con el prefijo "Geo" para su fácil identificación (Cheng y cols., 2020), (Gomez Gonzalez, 2023). Además de estas variables geográficas, se incluyó la correspondiente variable climática (temperatura [T] o precipitación [P]) obtenida de los Modelos Climáticos Globales (GCM).

El proceso de evaluación y mejora de los modelos se realizó de manera iterativa. Inicialmente, se entrenó y evaluó el modelo base utilizando solo las variables mencionadas. Posteriormente, se fueron incorporando una a una las componentes principales adicionales, que tienen como objetivo representar la esencia del conjunto de las nueve variables climáticas originales. Tras cada adición, se realizó una nueva evaluación del modelo para determinar si la inclusión de la nueva componente resultaba en una mejora significativa del rendimiento o si, por el contrario, introducía ruido que pudiera perjudicar la precisión del modelo.

Este enfoque gradual permite identificar si cada componente principal contribuye de manera más efectiva a la calidad de las predicciones del modelo, y cuáles podrían estar introduciendo información redundante o incluso perjudicial. De esta manera, se optimiza la selección de variables para cada tipo de modelo, asegurando un equilibrio entre la complejidad del modelo y su capacidad predictiva (Teegavarapu y Goly, 2018).

Tabla 3.4: Variables utilizadas como predictoras y objetivo. Las variables latitud, longitud y altura son utilizadas como la principales variables geográficas para el entrenamiento de los modelos, las componentes principales se fueron añadiendo progresivamente al entrenamiento de los modelos.

PREDICTOR	VARIABLES PREDICTORAS				COMPONENTES PRINCIPALES					
Temperatura (ERA5)	Latitud	Longitud	Altura	Temperatura (GCM)	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
Precipitación (ERA5)	Latitud	Longitud	Altura	Precipitación (GCM)	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6

Fuente: Elaboración propia

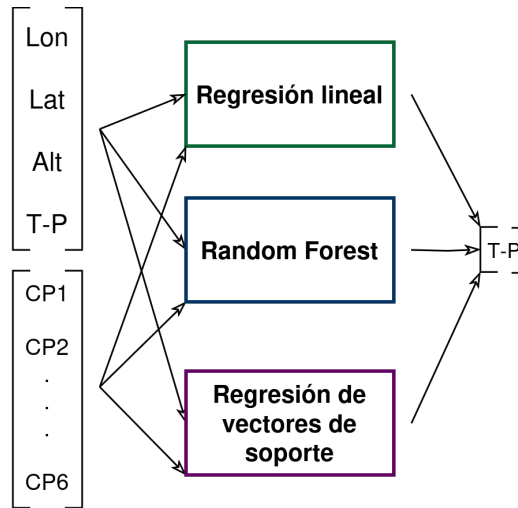


Figura 3.10: Visualización de las variables de entrada, modelos aplicados y variables respuesta.

Fuente: elaboración propia.

3.3.5.4. Regresión lineal (RL)

En este estudio se implementa un enfoque progresivo utilizando el modelo de regresión lineal con diferentes combinaciones de variables (Ebrahimi y Azadbakht, 2019). Este método permite evaluar sistemáticamente si la incorporación de cada componente principal contribuye a mejorar la precisión del modelo o por el contrario introduce ruido, buscando optimizar así su desempeño predictivo. Además, se realiza el análisis de los residuos o errores del modelo mediante la verificación de los supuestos de la regresión lineal, proporcionando una base para interpretar los resultados del modelo (Durbin y Watson, 1950), (Rosopa, Schaffer, y Schroeder, 2013). El proceso para el modelo de RL se desarrolla de la siguiente manera:

1. Modelo base: se inicia con un modelo que incorpora únicamente el componente geográfico (longitud, latitud y altura) y la variable climática de estudio (temperatura o precipitación) como variables predictoras.
2. Añadir las componentes principales: gradualmente, se añaden las componentes principales al modelo base. Después de cada adición, se evalúa el desempeño del modelo.
3. Evaluación de impacto: se analiza si las componentes principales añadidas aportan valor significativo al modelo o por el contrario generan ruido.
4. Análisis del modelo:

- Cálculo de métricas de evaluación: se calculan métricas de ECM, RECM, EMA, R y R^2 en cada entrenamiento de los modelos.
 - Monitoreo del modelo: se determina si la nueva variable aumenta la precisión y el poder predictivo del modelo.
 - Evaluación de ruido: se verifica si la variable añadida genera información irrelevante o contradictoria que podría deteriorar el rendimiento del modelo, estableciendo un criterio para determinar si las componentes principales son beneficiosas para el modelo.
5. Evaluación de los supuestos de la RL : para evaluar el comportamiento de los resultados del modelo RL obtenidos, se estudia los supuestos sobre los errores (residuos).

- Errores o residuos (ϵ_i): son la diferencia entre el valor observado (Y_i) y el valor predicho ($\hat{Y}_i$) por el modelo de regresión:

$$\epsilon_i = Y_i - \hat{Y}_i$$

- Normalidad: se asume que los residuos siguen una distribución normal con media a cero que analizamos con métodos gráficos como histogramas y Q-Q de los residuos y la prueba de estadística de normalidad de Shapiro-Wilk.
 - Distribución normal con media cero ($\epsilon \sim N(0, \sigma^2)$), donde:
 - 0 : es un vector de ceros
 - σ^2 : representa la matriz de covarianza de los residuos
 - Prueba de Shapiro-Wilk evalúa si los errores sigue una distribución normal W que se define como:

$$W = \frac{(\sum_{i=1}^n a_i X_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Donde:

X_i : observación ordenada en la muestra.

$\bar{X}$: media muestral.

a_i : coeficientes de la media de los cuantiles de una distribución normal.

n : número de observaciones en la muestra.

Se compara el valor de W con valores críticos o el $p - value$.

- si $p < 0.05$ se rechaza la hipótesis de normalidad.
- si $p \geq 0.05$ no se rechaza la normalidad.
- Gráfico Q-Q (Quantile-Quantile Plot): comparamos los cuantiles empíricos de los residuos con los cuantiles teóricos donde:
 - Cuantiles empíricos: provienen de los datos ordenados, para el conjunto de de n residuos, la posición del i -ésimo cuantil se estima cómo:

$$p_i = \frac{i - 0.5}{n}$$

Donde:

i : es el índice de orden de los datos.

n : número total de los datos.

p_i : es el porcentaje acumulado que representa el cuantil i -ésimo.

- Cuantiles teóricos: provienen de la distribución de referencia, para cada p_i , buscamos el valor $Q(p_i)$ que cumple:

$$Q(p_i) : F^{-1}(p_i)$$

Donde:

$F^{-1}(p_i)$: es la inversa de la función de la distribución acumulada, de la distribución teórica.

- Homocedasticidad: esta evalúa si la varianza de los errores debe de ser constante, la cual se verifica con la prueba de Breusch-Pagan.

- Se calculan los residuos al cuadrado $u_i = \epsilon_i^2$
- Se ajusta el nuevo modelo de regresión donde los residuos al cuadrado (u_i) son la variable dependiente.
- El estadístico de prueba de Breusch-Pagan se define como:

$$LM = nR^2$$

Donde:

n : es el número de observaciones.

R^2 : es el coeficiente de determinación de la regresión auxiliar.

- si $p < 0.05$ se rechaza la hipótesis de homocedasticidad, es decir hay heterocedasticidad.
- si $p \geq 0.05$ la varianza de los residuos es constante, es decir hay homocedasticidad.
- Independencia: este supuesto evalúa si los residuos del modelo de la regresión muestran autocorrelación entre sí y se utiliza la prueba de Durbin-Watson para verificarlo (DW).

$$DW = \frac{\sum_{i=2}^n (\epsilon_i - \epsilon_{i-1})^2}{\sum_{i=1}^n \epsilon_i^2}$$

Donde:

ϵ_i : residuos del modelo de regresión.

n : número total de observaciones.

Este estadístico DW toma valores entre 0 y 4 donde:

$DW \approx 2$: no hay autocorrelación.

$DW < 2$: autocorrelación positiva.

$DW > 2$: autocorrelación negativa.

Este primer enfoque en la metodología no solo optimiza el modelo de regresión lineal para las variables climáticas en estudio, sino que también proporciona un marco de referencia como el modelo base a comparar con los otros modelos a implementar.

3.3.5.5. Random forest regressor (RFR)

El Random Forest Regressor (RFR) es un modelo de aprendizaje automático no paramétrico que se basa en la construcción de múltiples árboles de decisión. Su funcionamiento consiste en crear estos árboles utilizando un subconjunto aleatorio de los datos, creando un bosque de múltiples árboles, donde sus predicciones finales son promediadas, generando una mayor robustez y precisión en comparación con un solo árbol de decisión (H. Zhang y cols., 2021).

En este estudio se ajustaron dos parámetros clave:

1. Número de estimadores: este parámetro define la cantidad de árboles en el bosque. Se experimentó con valores de 10, 25, 50 y 100 estimadores para encontrar el equilibrio óptimo entre precisión, tiempo de ejecución y recursos computacionales.
2. Semilla aleatoria: valor constante para garantizar la reproducibilidad de los resultados.

El proceso de optimización del modelo sigue la misma secuencia de pasos para la RL, permitiendo analizar el efecto individual de cada una de las componentes principales en el rendimiento del modelo.

Otros de los hiperparámetros del modelo se configuran de la siguiente manera según varios autores, (Bartkowiak, Castelli, y Notarnicola, 2019), (Yan y cols., 2021):

- "Estimador" : [10, 25, 50, 100]
- "max_depth": se establece como "None", permitiendo que los árboles crezcan hasta su máxima profundidad.
- "max_features": se configura como "auto", lo que significa que se considerarán todas las características en cada división.
- "bootstrap": se establece como True, habilitando el muestreo con reemplazo para la creación de los subconjuntos de datos.

Con esta configuración se busca maximizar la capacidad del modelo para capturar patrones complejos en las variables, mientras se mantiene un balance con la generalización y evitar el sobreajuste.

3.3.5.6. Regresión de vectores de soporte (RVS)

La Regresión de Vectores de Soporte (SVR) es un modelo no paramétrico basado en la optimización. Su principio fundamental consiste en encontrar un hiperplano o frontera óptima que maximice el margen entre las predicciones y los datos observados (Sachindra, Ahmed, Rashid, Shahid, y Perera, 2018). En el contexto de regresión, se utiliza un 'epsilon-tube' que es la zona de tolerancia alrededor de la función de regresión, para determinar las mejores predicciones. Si las predicciones caen dentro del tubo, estas no contribuyen al error, por el contrario los puntos por fuera del tubo se consideran como errores y estos afectan al modelo.

Matemáticamente si una predicción esta dentro del margen de 'epsilon' el error es cero, por el contrario si el esta por fuera del margen, el error es la distancia del punto al borde del tubo. Es así como este modelo nos permite un margen de error controlado por los parámetros 'epsilon' y 'C'.

Las características principales utilizadas para el modelamiento con SVR fue el Kernel "RBF" ampliamente utilizado por los autores (B. Wang y cols., 2018), (Sebbar y cols., 2023), (Sun y cols., 2023) por su eficiencia en el modelamiento de relaciones no lineales entre variables. Por otro lado se realiza la modificación de los hiperparámetros: "C", "Epsilon" y "Gamma".

Para la optimización del modelo se utilizó la siguiente estrategia:

1. Se selecciona un subconjunto inicial de 75,736 registros (equivalente a 2 meses de datos) para la búsqueda eficiente de la mejor combinación de hiperparámetros.
2. Se realizó una validación cruzada explorando las siguientes combinaciones:
 - "C" : [0.1, 1, 10]
 - "Epsilon" : [0.01, 0.1, 0.5]
 - "Gamma" : ['Scale', 'Auto']
3. Una vez identificados los mejores parámetros, se amplió el conjunto de entrenamiento a 545,789 y 1,236,541 registros para evaluar el rendimiento del modelo con un mayor volumen de datos.

3.3.6. Downscaling estadístico con RNA

Para abordar los dos desafíos principales del downscaling estadístico, la resolución espacial y temporal, se implementa una red neuronal artificial basada en una arquitectura ConvLSTM2D con bloques residuales (He, Zhang, Ren, y Sun, 2016). Esta estructura fue seleccionada por su capacidad para manejar datos espacio-temporales y capturar dependencias a largo plazo, lo cual es esencial para el modelado climático de alta precisión (Shangguan, Min, y Shi, 2023). La reducción

de la resolución espacial se realiza a través del mapeo uno a uno de los píxeles del GCM a la resolución más fina de ERA5-Land. Por otro lado, la reducción de la resolución temporal se lleva a cabo utilizando secuencias de datos de entrada del GCM para pronosticar en tres pasos de tiempo los datos horarios de ERA5-Land, como se muestra en la figura 3.11.

La capa de entrada del modelo integra varias fuentes de datos, incluyendo las variables de predicción (temperatura y precipitación), datos geoespaciales (longitud, latitud y altitud) y las componentes principales derivadas del análisis previo. Estas entradas se organizan en secuencias temporales compatibles con la estructura de la capa ConvLSTM2D. La arquitectura del modelo incluye múltiples capas ConvLSTM2D y Conv2D, así como capas de regularización, como BatchNormalization, Dropout y regularización L1L2, que ayudan a controlar el sobreentrenamiento y el ajuste excesivo del modelo.

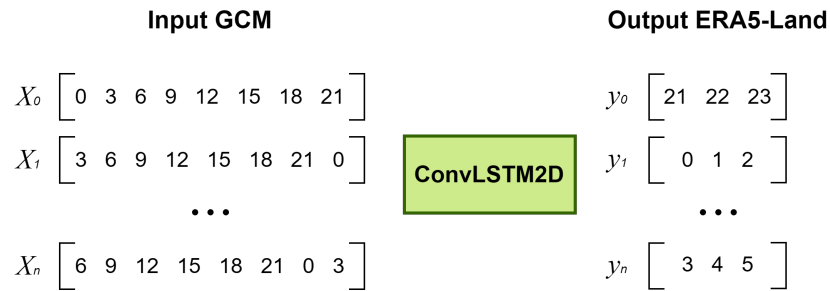


Figura 3.11: Vector de entrada con 8 secuencias de modelo GCM a la red ConvLSTM2D para realizar predicción de 3 pasos en el tiempo con los valores correspondiente a ERA5-Land.

Fuente: elaboración propia.

La capa de entrada del modelo incluye las variables de predicción (temperatura y precipitación), datos geoespaciales (altura), y las componentes principales derivadas del análisis previo. Estas entradas se organizan en secuencias temporales compatibles con la estructura de la capa ConvLSTM2D.

El diseño del modelo implica múltiples capas ConvLSTM2D y Conv2D, junto con capas de regularización como BatchNormalization, Dropout y regularización L1L2. Estas capas ayudan a controlar y mejorar la capacidad del modelo, para generalizar sobre los datos no vistos, y así evitar el sobreentrenamiento y sobreajuste. (Salman, Heryadi, Abdurahman, y Suparta, 2018), (H. Yang, Yan, Liu, y Song, 2022).

Se realizaron experimentos iterativos ajustando hiperparámetros como la profundidad de filtros, tamaños de kernel, optimizadores y tasas de aprendizaje. El objetivo fue encontrar la configuración óptima que se adapte a los recursos computacionales disponibles y maximice el

rendimiento del modelo. Este enfoque de aprendizaje supervisado aborda un problema de regresión, manejando grandes volúmenes de datos y requiriendo una capacidad de cómputo significativa.

A continuación, se detalla en cada subsección el proceso de la partición de datos, la validación cruzada, la arquitectura de la red a utilizar, el ajuste de hiperparámetros y la evaluación final de cada modelo.

3.3.6.1. Partición de los datos

Para el desarrollo y evaluación efectiva de los modelos de RNA, se realiza la partición estratégica del conjunto de datos en tres subconjuntos (Guyon, 1997), figura 3.12.

El conjunto de entrenamiento se utiliza para que la red aprenda los patrones y relaciones entre los datos, ajustando los pesos y los sesgos basándose en la minimización de la función de pérdida. La partición se realiza entre 60 % y 80 % de los datos.

La partición de validación la red lo utiliza para ajustar el proceso de entrenamiento, evaluando el rendimiento del modelo en cada época ayudando a prevenir el sobreajuste. El tamaño de la partición se realiza entre el 20 % y el 10 %.

La última partición de los datos comprende también un 20 % a 10 % como datos de prueba, no vistos durante el entrenamiento y validación. Estos nuevos datos son utilizados para proporcionar una estimación imparcial del rendimiento final del modelo, después de que el modelo ha sido completamente entrenado y ajustado.

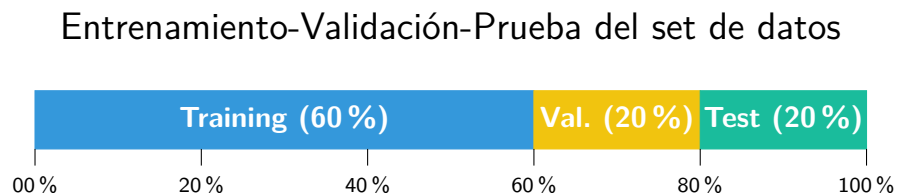


Figura 3.12: División tripartita del set completo de datos.

Fuente: elaboración propia.

Esta partición de los datos es adecuada y garantiza que el modelo de RNA implementado sea robusto, generalizable y capaz de realizar predicciones precisas en nuevos conjuntos de datos.

3.3.6.2. Validación cruzada

El modelo se evaluó mediante una estrategia de validación cruzada temporal con 5 particiones (5-fold cross-validation), complementando la división tradicional del conjunto de datos, Figura 3.13. Esta metodología se implementó preservando la estructura temporal de las

series. La validación cruzada permite una evaluación más rigurosa y realista del rendimiento del modelo, minimizando el riesgo de sobreajuste y proporcionando una estimación más confiable de su capacidad predictiva en condiciones operativas (Araya-Osses, Casanueva, Roman-Figueroa, Manuel Uribe, y Paneque, 2020), (Sanfilippo, Timm, Frazier, y Giambelluca, 2024), (Balmaceda-Huarte, 2024). La preservación de la temporalidad en cada partición resulta fundamental para mantener la integridad de los patrones climáticos subyacentes y asegurar que la evaluación refleje adecuadamente la capacidad del modelo para capturar la variabilidad climática en diferentes períodos. Esta aproximación metodológica fortalece la confiabilidad de las métricas de desempeño y proporciona una base más sólida para la validación del modelo.

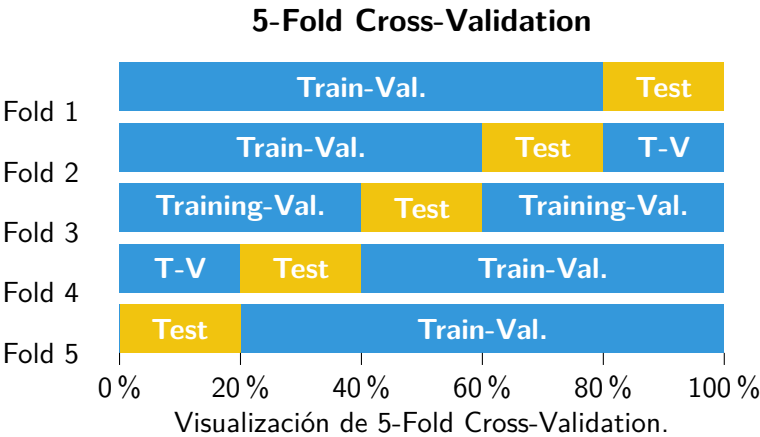


Figura 3.13: Particiones de 5-Fold para la validación cruzada en el set de datos.

Fuente: elaboración propia.

3.3.6.3. Arquitectura del Modelo RNA

Para modelar las variables de temperatura y precipitación, se implementa una arquitectura ConvLSTM2D con múltiples capas, complementadas con bloques residuales para optimizar el rendimiento. Este diseño, representado en la figura 3.14, permite capturar dependencias espaciales y temporales a largo plazo, mejorando la precisión de las predicciones y la eficiencia computacional del modelo (Hamidi y Roshani, 2023), (Han y cols., 2024).

Como se detalla en la figura 3.11, los vectores de entrada de la red incluyen múltiples características, y se emplea una capa de máscara para gestionar valores nulos, garantizando la coherencia de los datos a lo largo del modelo. El flujo de procesamiento comienza con una capa ConvLSTM2D, seguida de una capa BatchNormalization para estabilizar las activaciones y mejorar la capacidad de generalización. A continuación, los datos se dirigen a un bloque residual, que

consta de dos capas ConvLSTM2D. Este bloque facilita un aprendizaje más profundo y eficiente al permitir que las capas aprendan representaciones residuales, mejorando la optimización de los gradientes.

Después de pasar por el bloque residual, los datos atraviesan una capa adicional ConvLSTM2D, seguida de una capa Conv2D que transforma las características extraídas. Finalmente, la red se reconfigura en una secuencia de salida de tres valores, cumpliendo con los requerimientos establecidos para la capa de salida y proporcionando resultados en la resolución temporal deseada.

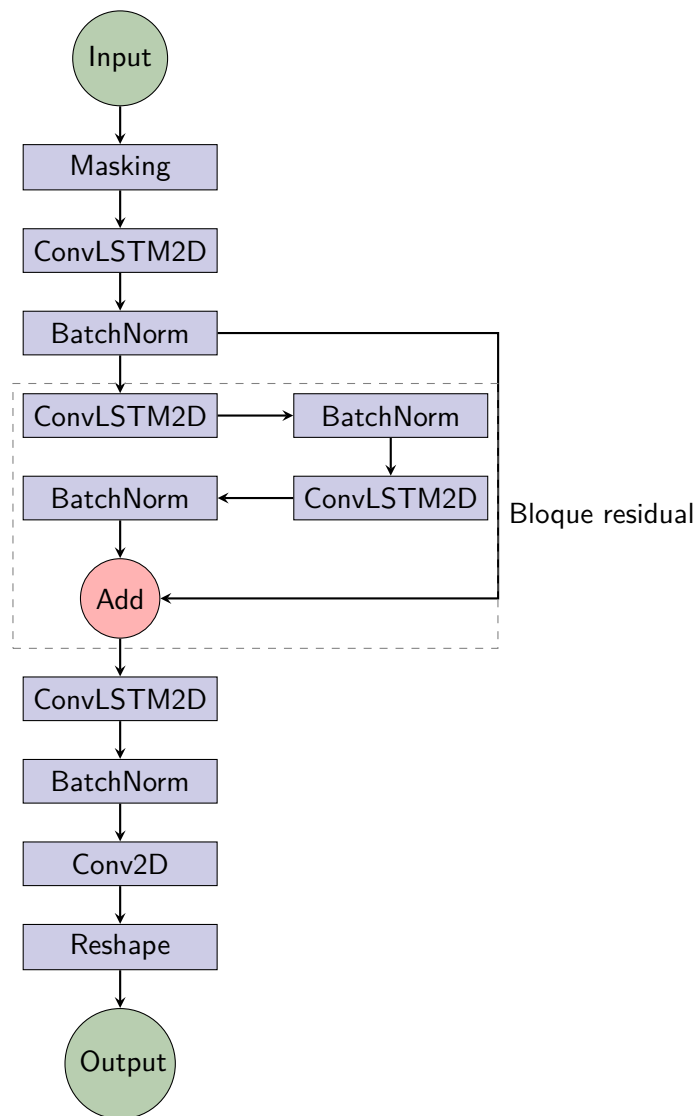


Figura 3.14: Arquitectura del modelo de Downscaling estadístico utilizando diferentes capas de RNA, y bloques residuales.

Fuente: elaboración propia.

3.3.6.4. Estrategia de compilación

Para la compilación del modelo ConvLSTM2D, se emplean los optimizadores Adam, con una tasa de aprendizaje inicial de 0.0005 (Nourani, Paknezhad, Mohammadisepasi, y Zhang, 2024). Estos optimizadores fueron seleccionados por su capacidad para ajustar automáticamente las tasas de aprendizaje de los parámetros, lo cual es particularmente útil en modelos complejos y con

grandes volúmenes de datos como los de arquitectura ConvLSTM2D. La función de pérdida se configura como el Error Cuadrático Medio (ECM), dada su capacidad para capturar la precisión de predicción en términos de la media de los errores cuadráticos. Además, se utiliza el Error Absoluto Medio (EAM) como métrica complementaria para evaluar la precisión, permitiendo una visión general del error promedio y proporcionando una evaluación de las predicciones menos susceptibles a los valores atípicos.

Para evitar el sobreentrenamiento y mejorar la convergencia del modelo, se implementa un enfoque de Early Stopping con un límite de 20 épocas sin mejoras en el conjunto de validación. Este proceso se configura para restaurar automáticamente los pesos del modelo en el punto de mejor rendimiento (es decir, aquellos que alcanzan la menor pérdida en el conjunto de validación). Además, se incluye un mecanismo de ajuste de la tasa de aprendizaje mediante ReduceLROnPlateau, que monitorea la pérdida del conjunto de validación y reduce la tasa de aprendizaje en un factor de 0.5 después de 5 épocas sin mejoras (T R y cols., 2024). Este ajuste cuenta con un periodo de cooldown de 2 épocas, permitiendo que el optimizador mantenga la tasa de aprendizaje ajustada, pero no reducible más allá de un mínimo de $1.00e-9$, evitando así la degradación del aprendizaje por una tasa de aprendizaje excesivamente baja.

Finalmente, el modelo se entrena con un tamaño de lote (Batch Size) de 64, lo cual permite un balance entre el uso eficiente de memoria y el procesamiento computacional. Este tamaño de lote facilita la estabilización de las actualizaciones de gradientes en cada iteración, lo cual es clave para un modelo de aprendizaje supervisado con un gran conjunto de datos espacio-temporales. Se ha ajustado el Batch Size para optimizar el rendimiento del modelo en función de los recursos computacionales disponibles, garantizando un entrenamiento efectivo sin comprometer la estabilidad o el tiempo de convergencia del modelo (Ioffe y Szegedy, 2015).

3.3.6.5. Tuneo de hiperparámetros

Se realiza un tuneo de hiperparámetro evaluando la capacidad computación y tiempo de ejecución, buscando los parámetros más óptimos para nuestra cantidad de datos y respuesta a la precisión de los modelos anteriores. Es así como se establecen los siguientes hiperparámetros:

- Partición de datos: Se evalúan dos configuraciones de partición para el conjunto de datos: (60-20-20) y (80-10-10). Estas combinaciones representan la división entre entrenamiento, validación y prueba. La primera configuración se enfoca en equilibrar la cantidad de datos para cada fase, asegurando suficiente información para la validación y la prueba, mientras que la segunda prioriza un mayor volumen de datos de entrenamiento, útil cuando se trabaja con conjuntos de datos más grandes y se desea un modelo bien ajustado. La elección entre estas configuraciones permite adaptar el modelo de acuerdo a la disponibilidad de datos y

los objetivos del análisis.

- Longitudes de series temporales: Para capturar dependencias temporales de diferente longitud, se prueban varias configuraciones de secuencia, incluyendo 8, 12, 16 y 24 pasos de tiempo. La elección de la longitud de la serie temporal afecta la capacidad del modelo para capturar patrones a corto y largo plazo. Secuencias más largas pueden ser útiles para detectar tendencias estacionales y patrones recurrentes, mientras que secuencias más cortas optimizan el tiempo de procesamiento, siendo eficaces en patrones de variabilidad rápida.
- Tamaño de kernel: Se evalúan diferentes tamaños de kernel en las capas convolucionales, como 2x2, 3x3, 4x4 y 5x5. Estos tamaños determinan el área de los datos de entrada que el modelo considera en cada paso de la convolución, afectando así su capacidad para captar patrones espaciales de distintas escalas. Kernels más pequeños permiten una detección de características más detallada, mientras que kernels más grandes pueden ser más eficaces para captar estructuras amplias. El ajuste de este parámetro es fundamental para optimizar el análisis espacial y la capacidad de generalización del modelo.
- Cantidad de filtros: Se prueba el modelo con diferentes cantidades de filtros en las capas convolucionales, como 16, 32, 64 y 128. La cantidad de filtros determina la cantidad de características espaciales que el modelo puede captar en cada capa. Al incrementar la cantidad de filtros, se aumenta la capacidad del modelo para detectar patrones complejos, aunque a un mayor costo computacional. Encontrar el equilibrio óptimo permite capturar la complejidad de los datos sin sobrecargar el procesamiento.
- Dropout: Se exploran tasas de dropout de 0.8, 0.7, 0.6 y 0.5 para evitar el sobreajuste durante el entrenamiento. El dropout consiste en desactivar de manera aleatoria una fracción de las neuronas en cada época, lo que reduce la dependencia del modelo en características específicas y mejora su capacidad de generalización. Tasas más altas pueden prevenir el sobreentrenamiento en conjuntos de datos pequeños o ruidosos, mientras que tasas más bajas pueden ser suficientes en modelos que ya cuentan con otras técnicas de regularización.
- Regularización L1 y L2: Se evaluaron valores de regularización L1 y L2 de 0.0001, 0.0005 y 0.001 para optimizar el rendimiento del modelo y prevenir el sobreajuste. La regularización L1 fomenta la dispersión de los pesos, es decir, promueve que muchos de ellos sean cercanos a cero, simplificando así el modelo y destacando únicamente las características más relevantes (sparsidad). Por su parte, la regularización L2 penaliza los pesos excesivamente grandes, estabilizando las conexiones y reduciendo la sensibilidad del modelo a pequeñas perturbaciones en los datos. Al combinar ambos enfoques, se logra una estructura de red más simple, robusta frente a nuevas muestras de datos y menos susceptible al ruido.

- **Optimizadores:** Se prueban dos optimizadores, Adam y RMSprop, debido a sus características de ajuste automático de tasas de aprendizaje. Adam combina las ventajas de AdaGrad y RMSprop, funcionando bien en modelos complejos y adaptándose a distintas tasas de cambio en el gradiente. RMSprop es útil para estabilizar los gradientes en problemas con datos altamente estacionales y ruidosos. El ajuste fino de estos optimizadores ayuda a alcanzar una convergencia más rápida y estable.
- **Tasa de aprendizaje:** se inicia con una tasa de aprendizaje relativamente alta para acelerar el proceso de convergencia en las primeras épocas, permitiendo que el modelo explore una gama más amplia de parámetros. Luego, si no se observan mejoras en la pérdida del conjunto de validación después de 5 épocas, se reduce drásticamente la tasa de aprendizaje para afinar el modelo. Esto ayuda a encontrar un óptimo local más preciso, reduciendo el riesgo de saltar sobre él debido a una tasa de aprendizaje elevada. Las tasas de aprendizaje que se exploran incluyen 0.005, 0.0005, 0.0001 y 0.00005, permitiendo un control más preciso en la fase final del entrenamiento.
- **Batch size:** Se prueban tamaños de lote de 32, 64 y 128, considerando el equilibrio entre la velocidad de entrenamiento y la estabilidad del modelo. Lotes más pequeños permiten actualizaciones más frecuentes de los gradientes, mientras que lotes más grandes ofrecen una representación más estable de las variaciones del conjunto de datos. El tamaño de lote óptimo se ajusta en función de los recursos computacionales y la complejidad de los datos, maximizando el rendimiento del modelo (Biniz, 2021).

3.3.6.6. Evaluación de los modelos

Se realiza una evaluación de cada modelo realizado para garantizar la precisión y generalización. Primero se visualiza la función de pérdida en entrenamiento y validación. Luego se evalúa el modelo con cada partición de los datos a través de múltiples métricas.

1. **Visualización de curvas de pérdida:** se realiza la gráfica de las curvas de pérdida para el conjunto de entrenamiento y validación con las métricas utilizadas 'ECM' y 'EMA', permitiendo observar la evolución del aprendizaje del modelo y detectar problemas como subajuste o sobreajuste, buscando la convergencia gradual de ambas curvas, sin que la pérdida de validación aumente significativamente respecto a la curva de entrenamiento.
2. **Evaluación en las tres particiones:** la evaluación se realiza en los conjuntos de entrenamiento, validación y prueba, proporcionando una visión completa del rendimiento del modelo, donde sus valores deben ser similares indicando una buena precisión y generalización entre los patrones y relaciones de los datos.

3.3.7. Desempeño de los modelos de downscaling estadístico

Para la validación del desempeño de cada uno de los modelos, determinamos su eficacia en la comparación de las predicciones de las variables y las observaciones con las estaciones a nivel local. Se desarrolla el Diagrama de Taylor con el resultado de los mejores modelos para compararlos con los datos de las observaciones realizadas por las estaciones que son consideradas como el valor real de cada variable (Kim y cols., 2022).

Basándonos en el diagrama de Taylor y el análisis complementario de las métricas de validación concluimos qué modelos funcionan mejor en comparación con los modelos de aprendizaje profundo. La selección final del modelo se implementa para el Downscaling estadístico operativo para las proyecciones planteadas.

3.4. Plataforma y publicación

La metodología empleada en esta aplicación web para la visualización y publicación de datos climáticos se basa en un enfoque integral que abarca desde el modelado de datos hasta la presentación al usuario final (Shrestha y cols., 2021), (C. Zhang y cols., 2022). Inicialmente, se selecciona el modelo más adecuado para representar con precisión los datos reales de temperatura y precipitación. Con este modelo, se realiza una proyección para el período 2024-2035, almacenando las variables resultantes en archivos NetCDF. Esta elección de formato no es arbitraria; los archivos NetCDF ofrecen una manipulación simplificada de las dimensiones de las variables climáticas y proporcionan de manera implícita los atributos y metadatos esenciales del modelado, facilitando así su uso por parte de los usuarios que deseen descargar y analizar los datos (Galiano, Migallon, Migallon, y Penades, 2010).

La arquitectura de la aplicación se ha diseñado siguiendo un paradigma desacoplado, lo que permite una mayor flexibilidad y mantenimiento (Yudin, 2020). El backend es desarrollado en Python (Zelle, 2016), utilizando el framework Django, este se encarga de la carga y publicación de los archivos NetCDF. Esta elección tecnológica permite un desarrollo ágil y robusto, capaz de manejar eficientemente los datos climáticos complejos. Por otro lado, el frontend, implementado en JavaScript, se centra en proporcionar una interfaz de usuario sencilla e intuitiva, capaz de visualizar los datos de manera clara y permitir interacción dinámica en el aumento o disminución del tiempo.

La plataforma web resultante ofrece a los usuarios la capacidad de visualizar y descargar datos de temperatura y precipitación para cada uno de los escenarios seleccionados del IPCC6 (SSP1.19, SSP2.45, SSP3.40). Todos los datos se presentan en el sistema de referencia WGS-84 (EPSG:4326), asegurando una compatibilidad global y facilitando su integración con diversos sistemas de información geográfica (Khazaei y cols., 2023). Esta elección de sistema de referencia

garantiza que los datos sean fácilmente interpretables y utilizables por una amplia gama de usuarios en diferentes contextos geográficos.

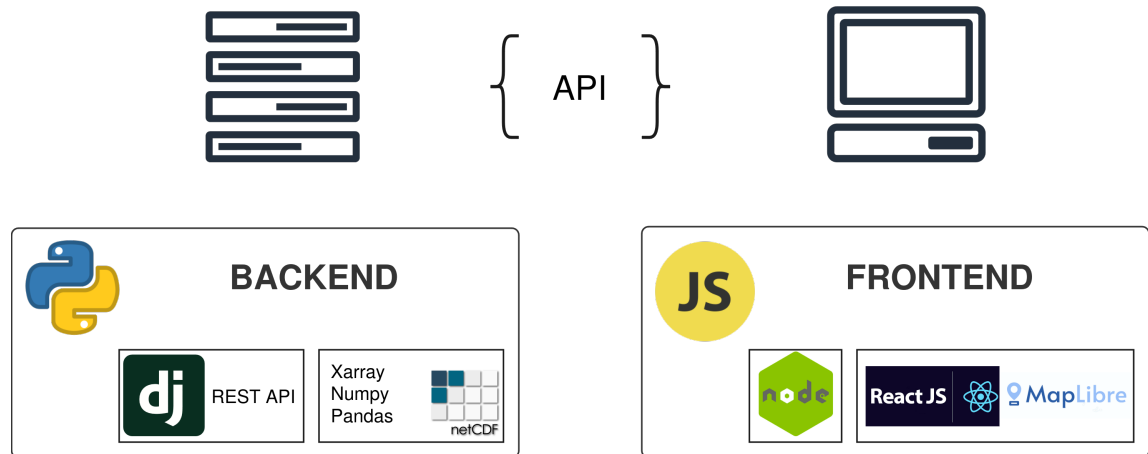


Figura 3.15: Metodología y herramientas utilizadas para la creación de la plataforma digital.

Fuente: elaboración propia.

Un aspecto fundamental de esta plataforma es su accesibilidad, diseñando un modelo de acceso libre, sin requerir registro de usuarios, dando acceso libre a los datos, permitiendo que investigadores, estudiantes y público en general puedan beneficiarse de esta información sobre el cambio climático.

3.4.1. Formato de archivo NetCDF

El formato de archivo NetCDF (Network Common Data Form) es una herramienta muy útil diseñada para el almacenamiento, procesamiento y análisis de grandes volúmenes de datos multidimensionales, utilizados en climatología y ciencias de la Tierra. Su estructura facilita el trabajo para investigadores y profesionales que trabajan con conjuntos de datos complejos y de gran tamaño (Unidata, 2024).

La versatilidad del NetCDF es su capacidad para manejar datos en múltiples dimensiones, como latitud, longitud y tiempo, de manera simultánea. Esta característica lo hace útil en el campo de la climatología, donde es primordial analizar patrones y tendencias a través de diversas escalas espaciales y temporales. Además, este formato permite realizar análisis y consultas de subconjuntos de datos de manera rápida y eficiente, ayudando a los investigadores a obtener información relevante sin perder detalles importantes (Michna y Woods, 2013).

Una de las características más destacadas del formato NetCDF es que es autodescriptivo. Cada

archivo NetCDF contiene metadatos detallados que describen cada variable y dimensión presente en el archivo. Esta característica no solo facilita la interpretación y el uso de los datos por parte de diferentes investigadores y sistemas, sino que también garantiza la preservación del contexto y la integridad de la información. Los metadatos detallados no solo mejoran la reproducibilidad de los estudios científicos, sino que también hacen que sea más fácil colaborar con otros investigadores de diversas áreas y regiones del mundo (Gaustad y cols., 2014).

La combinación de estas características hace del NetCDF una elección óptima para proyectos de investigación climática a gran escala, modelado ambiental y análisis de datos geoespaciales. Su adopción generalizada en la comunidad científica ha fomentado el desarrollo de una amplia gama de herramientas y bibliotecas de software compatibles, lo que aumenta aún más su utilidad y accesibilidad para los investigadores en todo el mundo.

La estructura del archivo se organiza en:

- Dimensiones: componente espacial y temporal en la cual están organizados los datos.
- Variables: almacenamiento de los datos de las variables climáticas distribuidas por las dimensiones.
- Atributos: metadatos que describen cada una de las variables.

3.4.2. Backend

La infraestructura del backend de nuestra aplicación se ha implementado sobre un servidor Windows, proporcionado por el laboratorio GISMODEL. En el desarrollo y estructuración del backend, se ha optado por el framework Django, reconocido por su arquitectura robusta y su filosofía de desarrollo rápido (V y cols., 2022), (Vincent, 2020). Django permite construir una aplicación web escalable y mantenible, proporcionando una serie de herramientas y funcionalidades que aceleran el proceso de desarrollo.

La carga y publicación de datos climáticos se realiza mediante un conjunto de bibliotecas Python especializadas. Xarray (Hoyer y Hamman, 2017) se utiliza como la herramienta principal para el manejo de datos multidimensionales, permitiendo una manipulación eficiente de los archivos NetCDF. Esta biblioteca se complementa con NumPy, que proporciona estructuras de datos y operaciones numéricas de alto rendimiento, esenciales para el procesamiento de grandes volúmenes de datos climáticos. Además, Pandas se emplea para la manipulación y análisis de datos tabulares, facilitando la transformación y preparación de los datos antes de su publicación (Abdollahi, Farjad, Gupta, y Hassan, 2022).

Para exponer los datos y funcionalidades de la aplicación de manera segura y eficiente, se implementa una API RESTful utilizando Django REST Framework (Christie y cols., 2024). Esta

potente biblioteca extiende las capacidades de Django, permitiéndonos crear endpoints de API robustos y bien documentados. La API facilita la comunicación entre el backend y el frontend, así como la integración con sistemas externos, proporcionando una interfaz estandarizada para el acceso y manipulación de los datos climáticos.

Esta arquitectura backend es diseñada e implementada no solo para garantizar un buen rendimiento en el manejo de datos climáticos complejos, sino que también proporciona una base sólida para futuras expansiones y mejoras en la plataforma. La combinación de tecnologías elegidas permite una gestión eficiente de los datos, asegurando al mismo tiempo la escalabilidad y mantenibilidad del sistema a largo plazo (Bao, Zhang, Ma, y Gan, 2023).

3.4.3. Frontend

El frontend de la aplicación web para la visualización de datos climáticos se desarrolla utilizando tecnologías actuales. En el núcleo de desarrollo frontend se encuentra React (Team y contributors, 2024), una biblioteca de JavaScript (Ecma International, 2024) ampliamente reconocida por su eficiencia y flexibilidad en la creación de interfaces de usuario dinámicas.

React, ejecutado en un entorno Node.js, permite construir una aplicación de página única (SPA) que ofrece una experiencia similar a una aplicación de escritorio, con transiciones suaves y actualizaciones rápidas de la interfaz. La elección de React se basa en su arquitectura de componentes, que facilita la creación de elementos de interfaz reutilizables y mantenibles. Esto no solo acelera el proceso de desarrollo, sino que también asegura una consistencia visual a lo largo de toda la aplicación.

Para la implementación del visor geográfico, pieza central de la aplicación, se utiliza MapLibre (contributors. MapLibreJS, 2024), esta potente biblioteca de mapeo de código abierto, proporciona las herramientas necesarias para crear visualizaciones geoespaciales interactivas y de alto rendimiento. MapLibre se integra perfectamente con React, permitiendo desarrollar un visor de mapas personalizado que se adapta específicamente a las necesidades de visualización de datos climáticos.

Para el despliegue de la aplicación web, se utilizó el servidor Apache, integrado en la plataforma XAMPP (Foundation, 2024), (Apache Friends, 2024), lo que proporciona un entorno confiable para la publicación de aplicaciones en entornos de producción. El proceso comienza con la preparación del proyecto React, mediante la compilación del código fuente. Este paso genera una versión optimizada y minificada de la aplicación, adecuada para su despliegue en producción. Los archivos generados, que incluyen HTML, CSS y JavaScript comprimidos, se transfieren al directorio raíz del servidor Apache, comúnmente ubicado en htdocs.

El servidor Apache se configura para manejar las solicitudes HTTP en el puerto 80, permitiendo el enrutamiento correcto hacia los recursos de la aplicación. En la aplicación React, se

asegura la correcta gestión del enrutamiento interno mediante ajustes en el archivo `.htaccess` en el directorio raíz, configurando las reglas de reescritura necesarias para manejar las rutas dinámicas sin errores. Esta configuración, combinada con la facilidad de gestión que ofrece XAMPP, permite publicar la aplicación en una URL específica, haciendo que esté accesible al público a través de la red. Al aprovechar las capacidades de Apache, se garantiza un rendimiento eficiente y una arquitectura escalable para la entrega de contenido web.

Capítulo 4

RESULTADOS Y DISCUSIONES

Este capítulo presenta los resultados obtenidos y las discusiones realizadas en cada una de las fases de la metodología y del diseño experimental propuesto en la investigación. El capítulo se estructura en tres etapas principales: análisis exploratorio de los datos, modelado de *downscaling* estadístico y publicación de la plataforma web desarrollada.

Esta primera sección se centra en el análisis exploratorio de las series temporales de temperatura y precipitación, junto con la aplicación del análisis de componentes principales sobre el conjunto de variables climáticas derivadas de los Modelos de Circulación Global. El análisis de componentes principales permite reducir la dimensionalidad de los predictores climáticos y evaluar su impacto en el desempeño del *downscaling* estadístico propuesto. Esta caracterización detallada de los datos proporciona una base para comprender la calidad, variabilidad y relaciones entre las variables climáticas utilizadas en el estudio.

La segunda sección aborda el modelado de las variables de temperatura y precipitación, empleando y evaluando los diferentes modelos de aprendizaje automático e inteligencia artificial propuestos: Regresión Lineal, Random Forest, Regresión de Vectores de Soporte y Redes Neuronales Artificiales de arquitectura ConvLSTM2D. En esta etapa, se detallan las métricas utilizadas para evaluar el desempeño de cada modelo y se destacan las fortalezas y limitaciones de los enfoques empleados en función de su capacidad para representar las dependencias espacio-temporales complejas de los datos.

En la tercera sección se describe el diseño, desarrollo y funcionamiento de la plataforma web creada para la visualización y descarga de los datos procesados. Esta plataforma permite a los usuarios acceder a proyecciones de temperatura y precipitación, ajustadas mediante *downscaling* estadístico, para diferentes escenarios climáticos proyectados entre los años 2024 y 2035, facilitando la toma de decisiones informadas a nivel local y regional.

Para finalizar, se presentan las conclusiones, que consolidan los principales hallazgos del estudio, evaluando la efectividad del *downscaling* estadístico aplicado a las variables de temperatura

y precipitación de los GCM. Las conclusiones reflejan la aplicabilidad del enfoque desarrollado para la costa del Pacífico del departamento del Valle del Cauca, destacando su capacidad para generar proyecciones climáticas de alta resolución que contribuyen a una mejor comprensión y gestión de los desafíos climáticos en la región.

4.1. Análisis exploratorio de los datos

Como se menciona en el capítulo anterior, el preprocesamiento y la calidad de los datos constituyen una de las etapas fundamentales para la generación de nueva información, influyendo de manera significativa en la relación y los resultados obtenidos al aplicar diferentes técnicas de aprendizaje automático e inteligencia artificial.

La visualización de los patrones y la distribución en espacio y tiempo de las variables analizadas facilita la identificación de datos atípicos, valores faltantes y la correlación entre las variables de entrada para la modelación climática. Este proceso es esencial para seleccionar los datos más relevantes que contribuyan a mejorar el rendimiento y la optimización del modelado espacio-temporal, así como para evaluar si el uso de múltiples variables climáticas permite generar resultados más robustos y confiables.

Del mismo modo, realizar el análisis de estadísticas descriptivas proporciona una comprensión detallada del comportamiento de las variables en la zona de estudio a lo largo de los últimos años. Mediante el uso de gráficos de dispersión, barras, series temporales y diagramas de cajas, es posible examinar las relaciones entre los datos, detectar valores faltantes y reconocer posibles datos atípicos o influyentes.

4.1.1. Resultados y discusiones para la Temperatura

La evaluación entre las diferentes fuentes de datos de temperatura se realizó mediante un análisis de correlación. Se calcularon los coeficientes de correlación de Spearman y Pearson, junto con el coeficiente de determinación (R^2), para cuantificar el grado de asociación entre los datos del GCM, ERA5-Land y las observaciones de las estaciones meteorológicas locales.

Este análisis se extendió a la comparación con las proyecciones del modelo GFDL-ESM4 bajo tres escenarios de trayectorias socioeconómicas compartidas (SSP1-1.9, SSP1-2.6 y SSP2-4.5), permitiendo evaluar la generalización climática del GCM con los patrones térmicos locales. Complementariamente, se realizó una caracterización estadística entre las tres fuentes de datos (observaciones in situ, ERA5-Land y GFDL-ESM4) mediante análisis descriptivos.

4.1.1.1. Correlación entre la temperatura de ERA5-Land y las mediciones locales

La evidencia para respaldar el proceso de reducción de escala se sustenta sobre la evaluación del coeficiente de determinación y las correlaciones entre las variables, las cuales se presentan en la tabla 4.1. En la tabla podemos encontrar los datos de correlación y coeficiente de determinación entre la temperatura de las estaciones locales y los estimaciones de temperatura del modelo ERA5-Land.

Tabla 4.1: Tabla de coeficiente de determinación, correlación de Pearson y Spearman de la variable temperatura entre los datos observados por las estaciones y las estimaciones de ERA5-Land. Se realiza la correlación Punto de cada estación a Píxel correspondiente en ERA5-Land

Estación	R ²	Pearson	Spearman
Arpto A. Bonilla	0.702	0.838	0.821
Farallones	0.733	0.856	0.877
La Cumbre	0.768	0.877	0.878
La Diana	0.711	0.815	0.884
Siloe	0.736	0.858	0.845
Unipacífico	0.651	0.807	0.835
Univalle	0.631	0.794	0.785

Para evaluar la concordancia entre los datos de temperatura de ERA5-Land y las mediciones de temperatura de las estaciones, se implementaron dos aproximaciones gráficas complementarias, Figura 4.1: series temporales por hora y diagramas de dispersión. Este análisis visual se realizó para tres estaciones meteorológicas ubicadas en diferentes zonas altitudinales de la cuenca: la estación "Unipacífico" en la zona baja, y las estaciones "La Cumbre" y "Farallones" en la zona alta. Las series temporales permiten visualizar la correspondencia en los patrones de variabilidad temporal horaria, mientras que los diagramas de dispersión facilitan la evaluación cuantitativa de la relación lineal entre ambas fuentes de datos. Esta representación dual proporciona una validación visual de la coherencia entre los datos modelados ERA5-Land y las observaciones instrumentales.

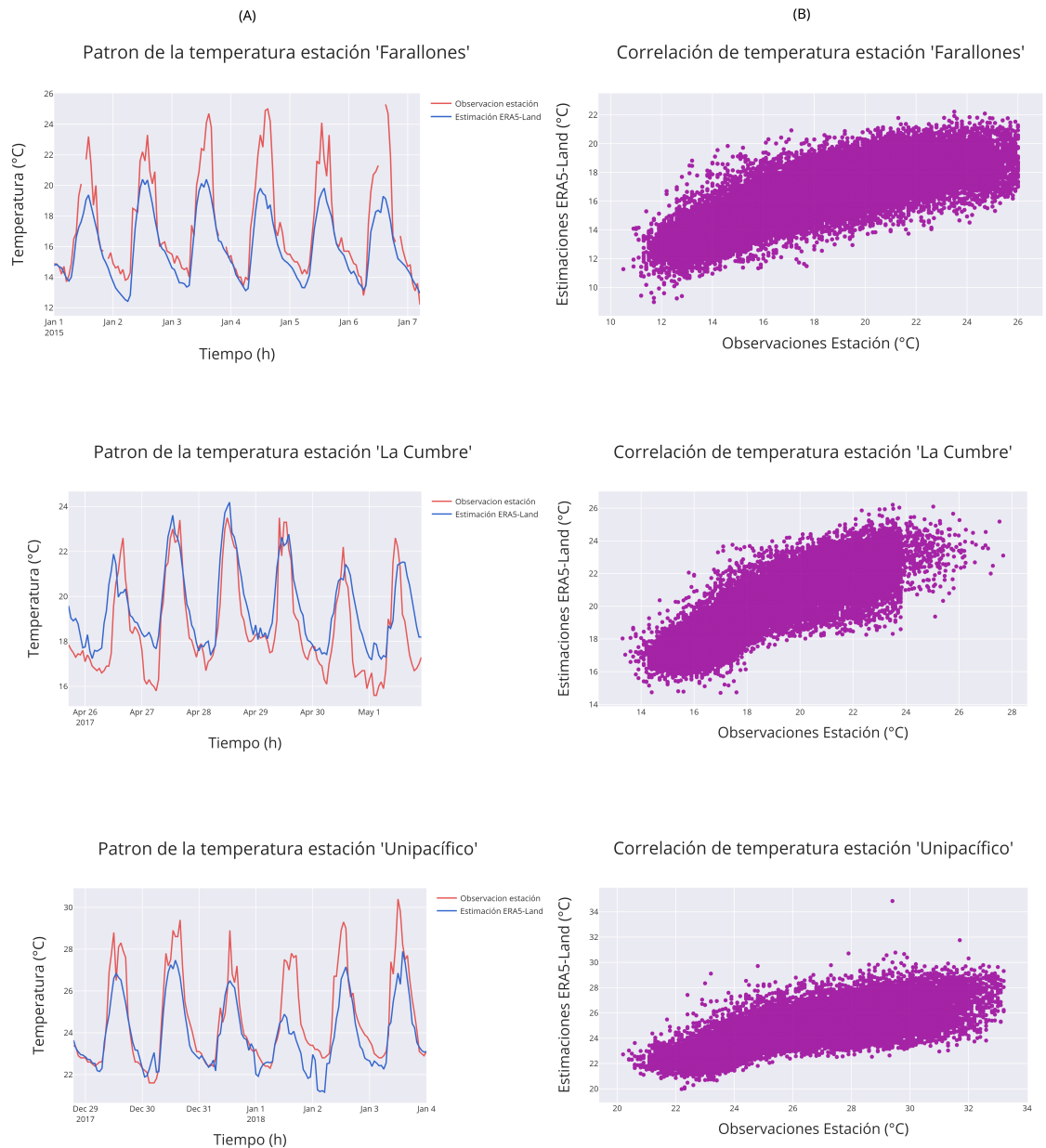


Figura 4.1: (A) Serie temporal horaria que muestra el patrón de variabilidad de la temperatura observada (rojo) entre la estación "Farallones", "La Cumbre", "Unipacífico" y el valor del píxel correspondiente del producto ERA5-Land (azul). (B) Diagrama de dispersión que ilustra la correlación entre las observaciones in situ y los valores del producto ERA5-Land.

El análisis reveló una fuerte correlación entre los datos de temperatura proporcionados por ERA5-Land y las observaciones locales registradas por las estaciones meteorológicas. Este hallazgo sugiere que los datos de reanálisis de ERA5-Land capturan adecuadamente la variabilidad térmica

a nivel regional, lo que refuerza su confiabilidad como fuente de referencia para estudios de downscaling en la región del Pacífico Colombiano.

Los valores del coeficiente de determinación R^2 oscilan entre 0.611 y 0.768, indicando que el modelo ERA-5 Land explica entre el 61.1 % y 76.8 % de la variabilidad de la temperatura observada en las estaciones.

Las correlaciones de Pearson muestran valores particularmente altos en las estaciones "La Cumbre" (0.877), "Farallones" (0.856) y "Siloe" (0.858), sugiriendo una fuerte relación lineal entre ambas fuentes de datos. La correlación de Spearman, que evalúa la relación monótona sin asumir linealidad, presenta valores consistentemente elevados, destacando "La Diana" (0.884), "Farallones" (0.877) y "La Cumbre" (0.878), lo que confirma la fuerte relación incluso en casos de posibles no linealidades.

Notablemente, la similitud entre los coeficientes de correlación de Pearson y Spearman en todas las estaciones sugiere la ausencia de valores atípicos influyentes que pudieran distorsionar la relación entre las variables, respaldando así la confiabilidad y robustez de los datos de ERA5-Land y las mediciones locales.

Por otro lado, estaciones como "Farallones", "La Cumbre" y "Siloe" muestran las correlaciones más altas con el modelo ERA5-Land (0.856, 0.877, 0.858), lo que sugiere que el modelo logra representar con mejor precisión las condiciones de temperatura en comparación con las zonas más bajas. Esta mayor precisión puede atribuirse a la estabilidad de los patrones meteorológicos en áreas de mayor altitud, donde la influencia de sistemas atmosféricos de gran escala es más dominante y menos afectada por factores locales.

En contraste, las regiones más bajas, especialmente aquellas cercanas a áreas urbanas o a la costa del Pacífico colombiano, presentan una mayor variabilidad térmica debido a la interacción de diversos factores. Entre ellos, la influencia de los vientos alisios, que transportan aire húmedo desde el océano, la proximidad al mar, cuya temperatura modula la variabilidad atmosférica, y la presencia de fenómenos urbanos que alteran el balance térmico, como las islas de calor y las diferencias en la cobertura y uso del suelo.

4.1.1.2. Descripción estadística de las fuentes de temperatura

Los resultados estadísticos descriptivos se estructura en dos niveles: inicialmente se caracteriza la distribución de la temperatura generalizada del píxel del modelo GFDL-ESM4 para toda el área de estudio, seguido de un análisis pareado entre las observaciones puntuales y los datos del píxel ERA5-Land, estos últimos extraídos de los píxeles coincidentes con la ubicación de cada estación meteorológica. Esta comparación permite evaluar la representatividad de ambas fuentes de datos estimadas y sus discrepancias en diferentes contextos topográficos y microclimáticos de la región medidos por cada estación meteorológica puntual.

Tabla 4.2: Caracterización estadística de la temperatura (°C) para diferentes fuentes de datos en la zona de estudio. Primero se presenta los estadísticos descriptivos del Modelo Climático Global GFDL-ESM4, luego el análisis comparativo entre datos observados y ERA5-Land por estación meteorológica.

Descripción estadística temperatura GCM								
Modelo	mean	std	min	25 %	50 %	75 %	max	Count
GFDL-ESM4	21.4	2.6	13.4	19.5	20.6	23.2	30.5	26279

Descripción estadística temperatura ERA5-Land y Estaciones								
Estación Arpto A. Bonilla								
Punto - Píxel	mean	std	min	25 %	50 %	75 %	max	Count
Observación	23.9	3.6	14	20.9	23.1	26.6	35.1	18276
ERA5-Land	19.9	1.9	11.7	18.6	19.5	21.2	26.9	26279
Estación Farallones								
Observación	16.6	2.9	11.2	14.4	15.8	18.7	25.9	11938
ERA5-Land	15.4	1.9	9.3	13.9	14.9	16.7	21.9	26279
Estación La Cumbre								
Observación	23.5	3.1	16.6	21.1	23.1	25.7	34.1	8315
ERA5-Land	19.7	1.9	13.1	18.2	19.2	21.1	26.8	26279
Estación Unipacífico								
Observación	25.5	2.5	20.5	23.6	24.7	27.3	33.1	10997
ERA5-Land	24.3	1.4	20.1	23.3	24.1	25.1	29.9	26279

Mediante dos visualizaciones complementarias, se presenta la distribución estadística de la temperatura. Primero el histograma de densidad con ajuste de curva normal para las tres fuentes de datos: Observaciones in situ (verde), ERA5-Land (azul) y GFDL-ESM4 (rojo); Seguido del diagrama de cajas que ilustran la media, mediana, cuartiles (25 % y 75 %) y valores atípicos para cada conjunto de datos.

Esta representación dual permite evaluar tanto la forma de la distribución como las medidas de tendencia central y dispersión, evidenciando las diferencias en el comportamiento de la temperatura captado por cada fuente de datos en las diferentes localizaciones dadas por las estaciones.

Distribución de la temperatura, estación "Farallones"

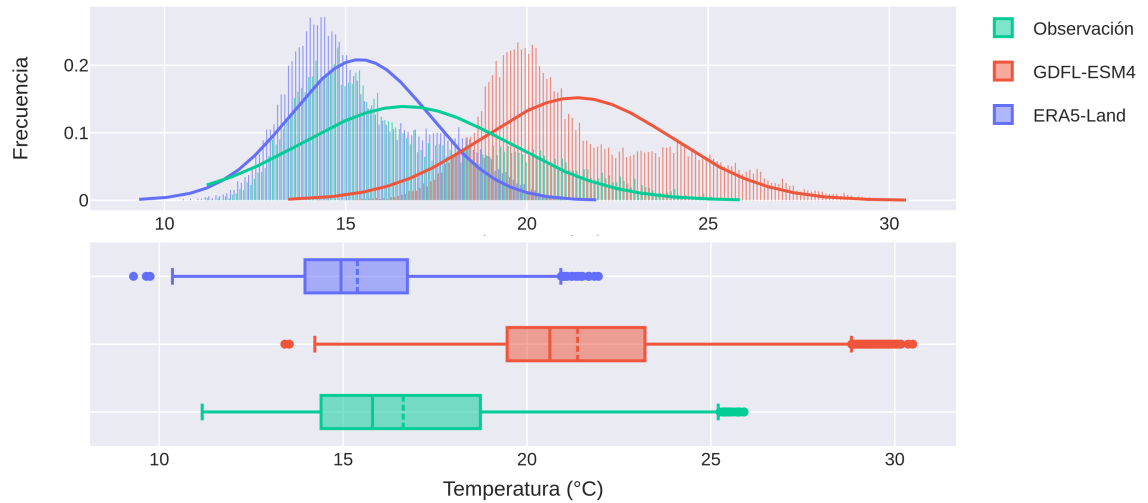


Figura 4.2: Distribución y diagrama de cajas de la temperatura entre los valores observados por la estación 'Farallones', ERA5-Land y GDFL-ESM4.

Distribución de la temperatura, estación "La Cumbre"

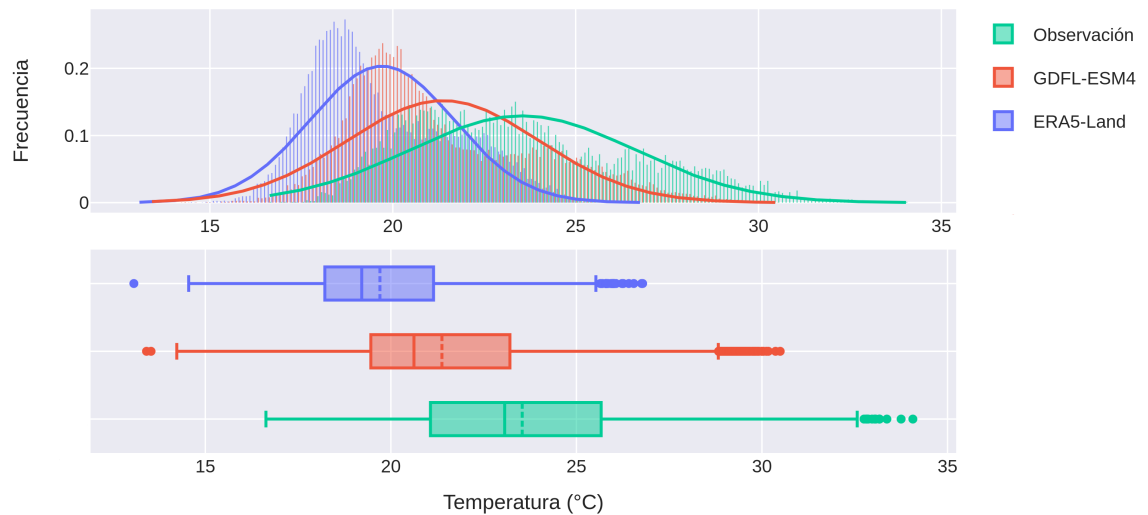


Figura 4.3: Distribución y diagrama de cajas de la temperatura entre los valores observados por la estación 'La Cumbre', ERA5-Land y GDFL-ESM4.

Distribución de la temperatura, estación "Unipacífico"

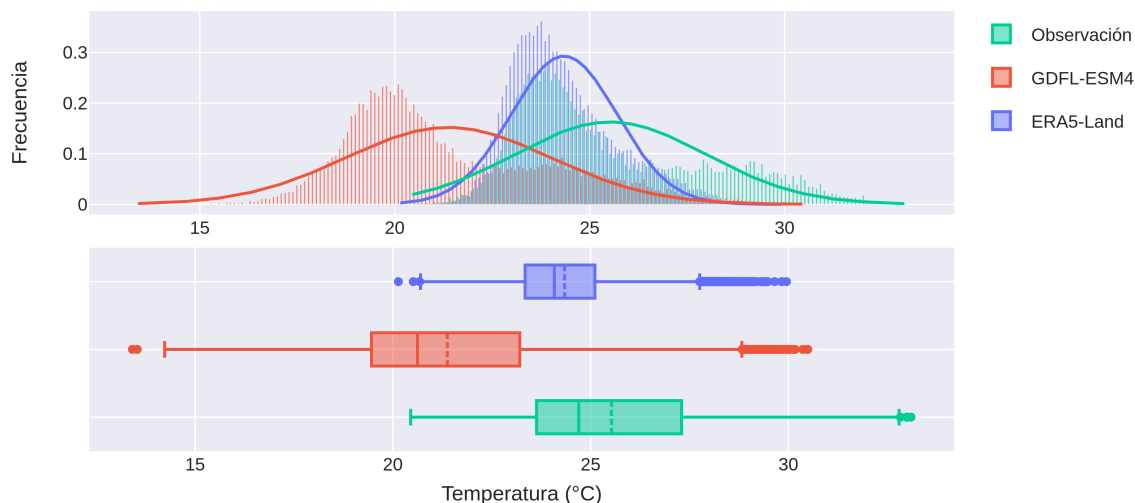


Figura 4.4: Distribución y diagrama de cajas de la temperatura entre los valores observados por la estación 'Unipacífico', ERA5-Land y GFDL-ESM4.

En el análisis exploratorio de los datos de temperatura de los datos de las estaciones meteorológicas y los modelos ERA5-Land y GFDL-ESM4, se identifica la presencia de valores atípicos a partir de los diagramas de caja, considerando aquellos que superan las 3 desviaciones estándar como se menciona en la subsección 3.3.2.3. Sin embargo, dado que estos modelos generan datos mediante parametrizaciones y procesos físicos bien definidos, dichos valores no se interpretan como errores, y pueden considerarse como representaciones de eventos climáticos extremos.

4.1.1.3. Correlación entre la temperatura de GFDL-ESM4, ERA5-Land y las observaciones locales

De igual manera se efectuó un análisis de correlación entre el GCM, los datos de reanálisis ERA5-Land y las observaciones de las estaciones meteorológicas, empleando los coeficientes de Spearman, Pearson y el coeficiente de determinación R^2 . Este análisis se realizó para tres escenarios de trayectorias socioeconómicas compartidas: SSP1-1.9 (tas119), SSP1-2.6 (tas126) y SSP2-4.5 (tas245), como se muestra en la siguiente tabla 4.3.

La similitud en los valores de correlación entre los diferentes escenarios durante el período histórico evaluado y el futuro cercano (2015-2035) es consistente con la literatura científica, que indica que las divergencias entre escenarios se magnifican principalmente después de 2040 (Nishant y cols., 2022) (Reboita, Ferreira, Ribeiro, da Rocha, y Rao, 2023). Esto se debe a que la inercia

del sistema climático y las concentraciones de gases de efecto invernadero ya presentes en la atmósfera dominan las proyecciones a corto plazo, mientras que las diferencias en las trayectorias de emisiones se vuelven más influyentes en períodos más largos.

Tabla 4.3: Se seleccionan los mejores resultados entre las correlaciones de los datos de temperatura Observados, ERA5-Land y los diferentes escenarios generados por el modelo GFDL-ESM4, lo cuales se resaltan de color verde. De este modo se selecciona el escenario que mejor se correlacione para realizar el entrenamiento del modelo.

Temperatura GFDL - ESM4									
Estación	Sperman			Pearson			R2		
Arpto A. B.	tas 119	tas 126	tas 245	tas 119	tas 126	tas 245	tas 119	tas 126	tas 245
ERA5-Land	0.587	0.605	0.635	0.608	0.631	0.649	0.370	0.398	0.421
Observación	0.499	0.516	0.517	0.527	0.551	0.546	0.278	0.304	0.298
Farallones									
ERA5-Land	0.622	0.639	0.660	0.635	0.659	0.662	0.403	0.434	0.438
Observación	0.634	0.623	0.629	0.629	0.621	0.641	0.396	0.386	0.411
La Cumbre									
ERA5-Land	0.626	0.657	0.646	0.642	0.675	0.662	0.412	0.455	0.438
Observación	0.663	0.670	0.675	0.660	0.662	0.677	0.436	0.438	0.458
Unipacífico									
ERA5-Land	0.535	0.556	0.584	0.542	0.561	0.584	0.294	0.315	0.341
Observación	0.621	0.634	0.627	0.613	0.614	0.611	0.376	0.377	0.373

El análisis comparativo de las métricas de correlación señala al escenario SSP2-4.5 como el más apropiado para la región de estudio. Este escenario representa una trayectoria intermedia de emisiones y desarrollo socioeconómico, caracterizada por un nivel moderado de mitigación climática y un desarrollo tecnológico gradual. Esta selección se alinea con las recomendaciones del IPCC para estudios de impacto regional donde se busca un equilibrio entre escenarios extremos (IPCC, 2023).

4.1.2. Resultados y discusiones para la precipitación

El análisis de la precipitación también se abordó mediante una evaluación de la correlación entre las distintas fuentes de datos, empleando los coeficientes de Spearman y Pearson y el coeficiente de determinación R^2 , para cuantificar el grado de asociación entre las observaciones in situ, los datos de reanálisis ERA5-Land y las proyecciones del modelo GFDL-ESM4 bajo tres escenarios de trayectorias socioeconómicas compartidas (SSP1-1.9, SSP1-2.6 y SSP2-4.5).

El análisis de la caracterización estadística de la precipitación se fundamenta en la evaluación

de los valores acumulados mensuales, lo cual permite identificar los patrones de estacionalidad y cuantificar la variabilidad interanual durante el período de observación establecido.

4.1.2.1. Descripción estadística de las fuentes de precipitación

El análisis estadístico descriptivo se realizó sobre la precipitación mensual acumulada, considerando los indicadores de tendencia central (media), dispersión (desviación estándar), valores extremos (mínimo y máximo) y distribución (cuartiles 25 %, 50 % y 75 %). Se analizaron las tres fuentes de datos, la precipitación generalizada del modelo climático global GFDL-ESM4 para toda el área de estudio, y las estimaciones de ERA5-Land correspondientes a los píxeles coincidentes con la ubicación de cada estación meteorológica.

La escala temporal mensual multianual es una metodología efectiva para analizar la precipitación como se mencionó en la subsección del capítulo anterior 3.3.2.2, ya que permite caracterizar patrones estacionales y ciclos anuales al filtrar la alta variabilidad de las observaciones horarias. Esta agregación mejora la detección de anomalías climáticas y fortalece los análisis estadísticos, como correlaciones entre fuentes de datos. Además, es un estándar ampliamente reconocido en la validación de modelos climáticos y productos de reanálisis, proporcionando un marco temporal óptimo para evaluaciones consistentes y comparativas.

Tabla 4.4: Estadísticas descriptivas de precipitación mensual (mm) del modelo GFDL-ESM4, ERA5-Land y estaciones meteorológicas para el período de estudio.

Descripción estadística Precipitación GCM								
Modelo	mean	std	min	25 %	50 %	75 %	max	Count
GFDL-ESM4	248.6	127.6	42.5	159.6	227.5	311.7	686.9	108

Descripción estadística Precipitación ERA5-Land y Estaciones								
Estación Arpto A. Bonilla								
Punto - Píxel	mean	std	min	25 %	50 %	75 %	max	Count
Observación	36.1	45.7	0.0	0.0	23.2	57.4	222.9	108
ERA5-Land	42.9	12.2	15.6	34.57	41.1	49.5	74.9	108
Estación Farallones								
Observación	83.6	83.9	0.0	4.7	68.9	142.5	347.6	108
ERA5-Land	32.2	8.9	15.5	24.8	30.8	37.6	56.6	108
Estación La Cumbre								
Observación	85.4	58.7	0.0	37.8	79.4	117.1	34.1	108
ERA5-Land	41.4	10.5	23.94	33.7	40.7	46.6	78.7	108
Estación Unipacífico								
Observación	354.9	268.4	0.7	176.575	264.4	498.9	1059.0	50
ERA5-Land	128.4	44.6	30.1	94.8	132.9	162.2	237.7	108

Los resultados muestran diferencias significativas entre las tres fuentes de datos. El modelo GFDL-ESM4 presenta valores medios más altos de 248.6 mm en comparación con las observaciones de las estaciones, que varían entre 85.4 mm en la estación en la parte alta de la cuenca “La Cumbre” y 354.9 mm en la parte baja en la estación “Unipacífico”.

Los datos de ERA5-Land generalmente subestiman la precipitación en comparación con las observaciones locales, especialmente notable en las estaciones “Unipacífico” (128.4 mm vs 354.9 mm) y la estación “La Cumbre” (41.4 mm vs 85.4 mm). La variabilidad temporal, representada por la desviación estándar, es particularmente alta en la estación Unipacífico (268.4 mm) y en el modelo GFDL-ESM4 (127.6 mm), sugiriendo una marcada estacionalidad en la precipitación como se describe en la región Pacífica del departamento.

Para analizar la variabilidad temporal de la precipitación mensual multianual que se generaliza por el modelo GFDL-ESM4, se generó un diagrama de cajas para los tres escenarios de cambio climático seleccionados (SSP1-1.9, SSP1-2.6 y SSP2-4.5), considerando el período 2015-2023. Este análisis permite visualizar la distribución y estacionalidad de la precipitación mensual bajo diferentes trayectorias socioeconómicas compartidas (SSP).

Precipitación total por mes, "GFDL-ESM4"



Figura 4.5: Distribución estadística mensual multianual de la precipitación según escenarios SSP del modelo GFDL-ESM4 (2015-2023)

El análisis de la precipitación mensual revela un patrón bimodal característico de la región, con máximos en abril-mayo y octubre-noviembre. El escenario SSP1-2.6 muestra mayor variabilidad interanual, especialmente en los meses más lluviosos, con valores atípicos que superan los 600 mm/mes. Los meses más secos (julio-agosto) presentan menor variabilidad entre escenarios, con medianas entre 100-200 mm/mes. El escenario SSP2-4.5, más pesimista en términos de emisiones, sugiere una ligera reducción en la precipitación durante los períodos húmedos, particularmente en noviembre-diciembre, lo que podría indicar cambios en los patrones estacionales bajo condiciones de mayor forzamiento radiativo.

Del mismo modo se analizó la distribución mensual de precipitación de ERA5-Land para los píxeles correspondientes a las siete estaciones meteorológicas estudiadas (Aeropuerto, Farallones, La Cumbre, La Diana, Siloe, Univalle y Unipacífico), permitiendo evaluar la variabilidad espacial y temporal de la precipitación en el área de estudio.

Precipitación total por mes, "ERA5-Land"

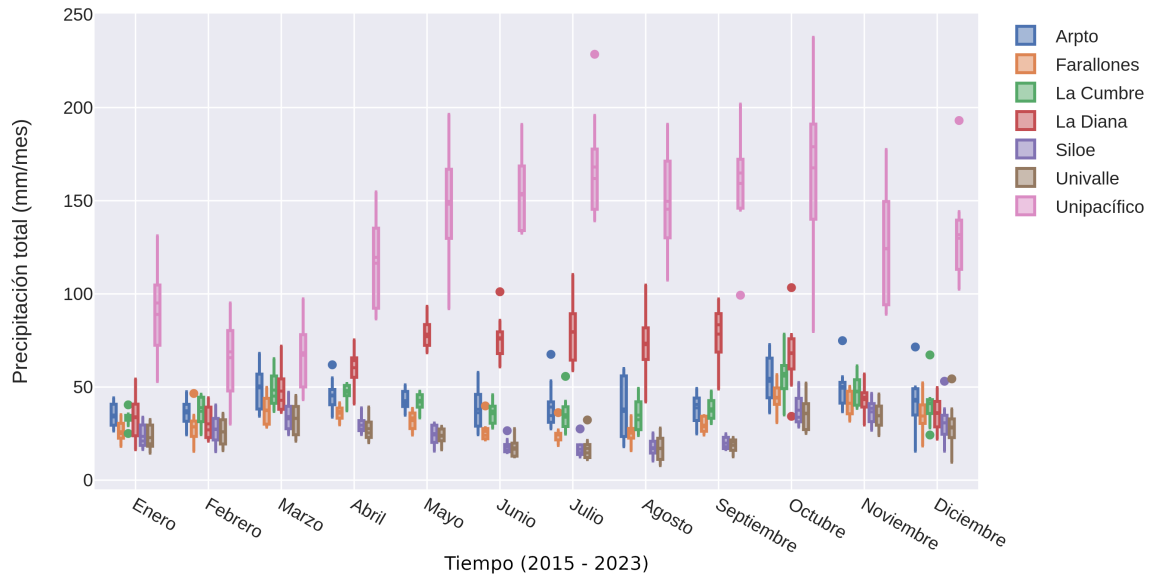


Figura 4.6: Distribución mensual de precipitación total (mm/mes) según ERA5-Land para los píxeles correspondientes a las estaciones meteorológicas (2015-2023).

Precipitación total por mes, estación "La Cumbre"

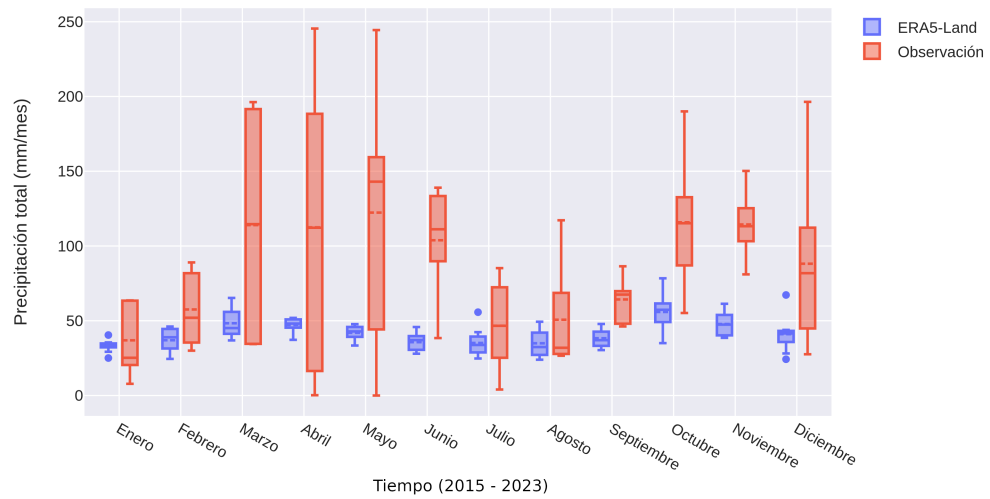


Figura 4.7: Comparación de la distribución total mensual (mm/mes) de precipitación entre datos observados y ERA5-Land para la estación "La Cumbre".

Precipitación total por mes, estación "Farallones"

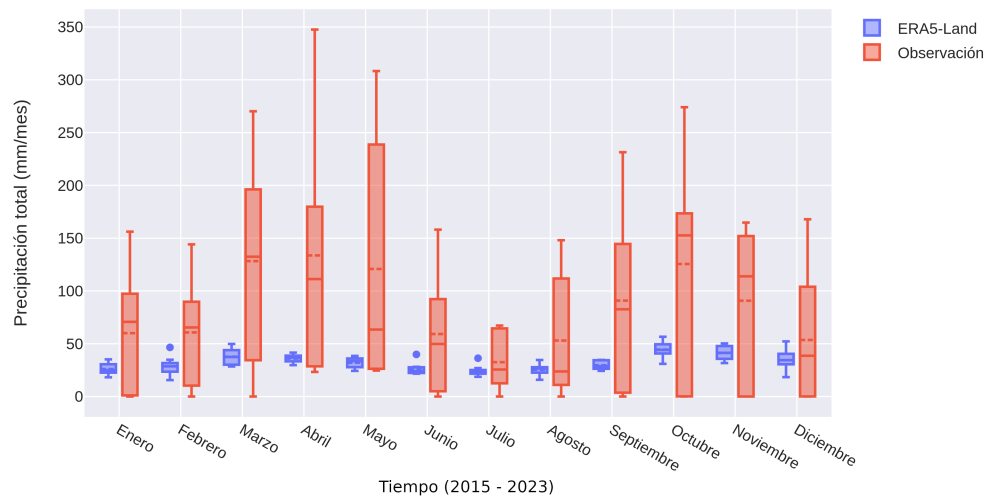


Figura 4.8: Comparación de la distribución total mensual (mm/mes) de precipitación entre datos observados y ERA5-Land para la estación "Farallones".

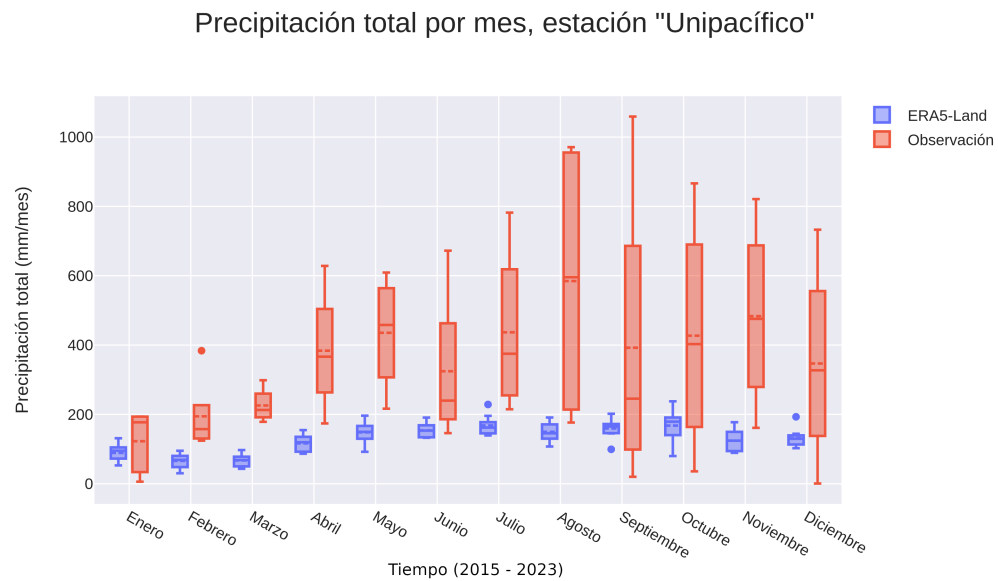


Figura 4.9: Comparación de la distribución total mensual (mm/mes) de precipitación entre datos observados y ERA5-Land para la estación "Unipacífico".

El análisis muestra una marcada diferencia en los patrones de precipitación entre las estaciones. La estación "Unipacífico" presenta los valores más altos y mayor variabilidad, con registros que superan los 150 mm/mes, particularmente durante septiembre-octubre, seguida de la estación "La Diana" que presenta un comportamiento particular de altas precipitaciones entre 50 y 100 mm/mes entre los meses julio, agosto y septiembre, mientras que en la estación "La Cumbre" los meses más húmedos alcanzan los 50 mm/mes. Del mismo modo las demás estaciones muestran valores significativamente menores, generalmente por debajo de 50 mm/mes. Se observa un patrón bimodal menos pronunciado que en el modelo GFDL-ESM4, con máximos relativos en abril-mayo y octubre-noviembre. La variabilidad interestacional es más evidente en la estación "Unipacífico", mientras que las estaciones urbanas (Arpto, Univalle, Siloe) muestran patrones más homogéneos entre sí.

En contraste, los resultados para la precipitación mostraron una correlación considerablemente más baja con los datos observados en las estaciones meteorológicas. Aunque no existe una correspondencia precisa entre los valores, los resultados del análisis mensual multianual, coinciden con lo reportado en estudios previos (González, 2014), (Aguirre y cols., 2017), donde se evidencia una mayor concentración de precipitación en las zonas bajas de la cuenca, en comparación con las zonas más altas. Esta diferencia sugiere una complejidad adicional en la modelización de la precipitación, que, aunque los modelos capturan tendencias espaciales generales, la variabilidad

puntual de la precipitación sigue siendo un desafío considerable, posiblemente debido a factores orográficos y microclimáticos no captados por los modelos de escala gruesa y en particular la naturaleza de los datos de precipitación utilizados.

4.1.2.2. Correlación entre la precipitación de ERA5-Land y las mediciones locales

Los resultados del análisis de correlación entre la precipitación horaria estimada por ERA5-Land y las observaciones de las estaciones meteorológicas revelan una limitada capacidad de los datos de ERA5-Land para representar la variabilidad pluviométrica. Se emplearon tres métricas de correlación para evaluar la concordancia entre ambas fuentes de datos, cuyos resultados se presentan en la Tabla 4.5.

Tabla 4.5: Métricas de correlación entre precipitación horaria observada en las estaciones y los valores de precipitación total/hora (mm/h) ERA5-Land para pares estación-píxel correspondientes.

Estación	R ²	Pearson	Spearman
Arpto A. Bonilla	0.003	0.057	0.102
Farallones	0.002	0.051	0.116
La Cumbre	0.002	0.042	0.112
La Diana	0.001	0.021	0.119
Siloe	0.005	0.071	0.174
Unipacífico	0.0003	0.017	0.122
Univalle	0.003	0.051	0.115

En contraste con la temperatura, el análisis de los coeficientes de correlación para la precipitación revela una baja correlación entre las estimaciones de ERA5-Land y las mediciones de la precipitación in situ. Los valores del coeficiente de determinación R² son bajos, oscilando entre 0.0003 y 0.005, indicando que el modelo ERA5-Land explica menos del 1 % de la variabilidad observada en la precipitación horaria. La estación "Siloe" presenta el R² más alto (0.005), mientras que "Unipacífico" muestra la correlación más débil (R² = 0.0003).

Las correlaciones de Pearson son consistentemente bajas (0.017 - 0.071), mientras que las correlaciones de Spearman, aunque ligeramente superiores (0.102 - 0.174), siguen siendo débiles, lo que significa que esta diferencia entre los coeficientes de Pearson y Spearman sugiere la presencia de relaciones no lineales y la influencia de valores extremos en la distribución de la precipitación.

4.1.2.3. Comparativo entre la precipitación de GFDL-ESM4, ERA5-Land y las observaciones locales

La evaluación de la correlación entre las proyecciones del modelo GFDL-ESM4 bajo diferentes escenarios SSP, los datos de reanálisis ERA5-Land y las observaciones in situ revela patrones de correlación heterogéneos a través de las diferentes estaciones de monitoreo. El análisis incluye tres métricas de correlación (Spearman, Pearson y R^2) para cada escenario (pr 119, pr 126 y pr 245), cuyos resultados se presentan en la Tabla 4.6.

Tabla 4.6: Análisis de métricas de correlación para la precipitación: comparación entre escenarios de GFDL-ESM4 con datos de precipitación de ERA5-Land y observaciones meteorológicas locales. Se seleccionan los mejores resultados de color verde, donde el escenario SSP 2-45 para la precipitación (pr 245) es el que mejor se correlaciona con las fuentes locales.

Precipitación GFDL - ESM4									
Estación	Sperman			Pearson			R2		
Arpto A. B.	pr 119	pr 126	pr 245	pr 119	pr 126	pr 245	pr 119	pr 126	pr 245
ERA5-Land	0.146	0.192	0.264	0.145	0.197	0.275	0.021	0.039	0.075
Observación	0.055	0.036	0.244	0.056	0.081	0.173	0.003	0.007	0.030
Farallones									
ERA5-Land	0.258	0.337	0.328	0.295	0.338	0.380	0.087	0.114	0.145
Observación	0.008	0.049	-0.023	0.038	0.030	0.070	0.001	0.0009	0.005
La Cumbre									
ERA5-Land	0.179	0.231	0.321	0.221	0.236	0.318	0.049	0.056	0.101
Observación	0.257	0.228	0.283	0.228	0.199	0.315	0.052	0.040	0.099
Unipacífico									
ERA5-Land	0.317	0.302	0.337	0.310	0.298	0.323	0.096	0.088	0.104
Observación	0.121	-0.059	0.105	0.106	0.012	0.146	0.011	0.0001	0.021

El análisis de los coeficientes de correlación revela varios patrones significativos, como la variabilidad entre escenarios, donde el escenario SSP2-4.5 (pr 245) muestra correlaciones más altas, con valores máximos de correlación de Pearson de 0.380 (ERA5-Land, “Farallones”) y 0.315 (Observación, “La Cumbre”). Los coeficientes de determinación R^2 son generalmente bajos, aunque mejoran en el escenario SSP2-4.5, alcanzando valores de hasta 0.145 para las observaciones en la estación “Farallones”

Por otra parte, la disparidad ERA5-Land contra las observaciones, donde ERA5-Land tiende a mostrar correlaciones más altas con el GCM que las observaciones directas, particularmente notable en las estaciones “Farallones” y “Unipacífico”.

Es importante destacar la variabilidad espacial, debido a que las correlaciones varían significativamente entre estaciones, indicando una fuerte influencia de factores locales en la representatividad de las simulaciones. Es así como para esta variable las estaciones "La Cumbre", "Farallones" y "Unipacífico" muestran las correlaciones más consistentes a través de los diferentes escenarios.

La selección del escenario SSP2-4.5 para el entrenamiento de los modelos posteriores se fundamenta en su mejor desempeño estadístico, aunque las correlaciones generalmente bajas sugieren la necesidad de considerar técnicas de corrección de sesgos o métodos de regionalización adicionales para mejorar la representatividad local de las proyecciones.

4.1.3. Variabilidad climática por hora

Se analizó el ciclo horario de precipitación total y temperatura promedio entre las estaciones meteorológicas representando el gradiente altitudinal en la región de estudio: "Unipacífico" (zona baja: 16 m.s.n.m), "La Cumbre" (zona alta 1613 m.s.n.m) y Farallones (zona montañosa alta: 2275 m.s.n.m). Este análisis compara los registros de las observaciones de las estaciones con estimaciones ERA5-Land durante el período 2015 a 2023, permitiendo caracterizar la variabilidad espacio-temporal de los patrones hidrometeorológicos a escala horaria (IDEAM, Consultado en el años 2023).

El análisis revela patrones climáticos distintivos asociados al gradiente altitudinal. Para la estación "La Cumbre" se observa un régimen térmico moderado con temperaturas promedio que oscilan entre 17 °C y 21 °C según los datos observados. Estas condiciones de la temperatura son representadas de manera adecuada por los valores de ERA5-Land, que estima temperaturas entre 18.3 °C y 22.6 °C, aunque con una ligera sobreestimación. Por otro lado la precipitación presenta un patrón unimodal, con acumulados máximos entre 400 mm y 600 mm durante las horas del mediodía (13:00 a 15:00). Este comportamiento sugiere la influencia predominante de procesos convectivos locales impulsados por el calentamiento diurno.

Precipitación acumulada y temperatura promedio por hora, estación "La Cumbre"

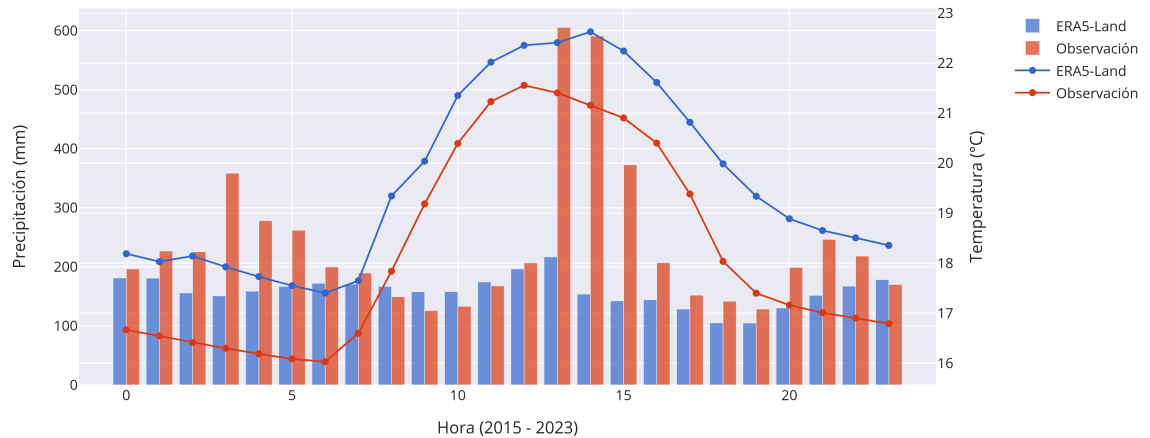


Figura 4.10: Precipitación total y temperatura promedio por hora entre el año 2015 y 2023. Datos estimados por ERA5-Land y las mediciones realizadas por la estación "La Cumbre"

Mientras que en la estación "Farallones" en la zona montañosa alta, se registra un régimen térmico más frío, con temperaturas entre 14 °C a 18 °C promedio registrado. Las precipitaciones presentan un patrón bimodal con acumulados que varían entre 100 mm y 450 mm cada hora. Estas precipitaciones pueden estar asociadas tanto al calentamiento diurno como a la interacción de sistemas orográficos, reflejando la complejidad de los procesos climatológicos en áreas montañosas. Aunque ERA5-Land sigue el patrón general observado de la temperatura, subestima las temperaturas en horas de la tarde, donde las estaciones reportan temperaturas máximas promedio de 21 °C en comparación con 18 °C máximos estimados por ERA5-Land.

Precipitación acumulada y temperatura promedio por hora, estación "La Farallones"

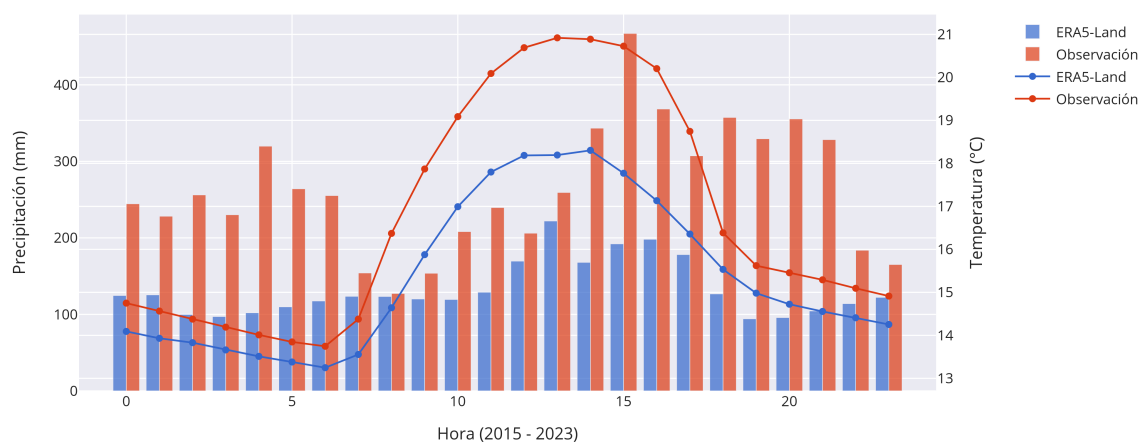


Figura 4.11: Precipitación total y temperatura promedio por hora entre el año 2015 y 2023. Datos estimados por ERA5-Land y las mediciones realizadas por la estación "Farallones"

Finalmente, la estación "Unipacífico" en la parte baja y costera, exhibe el régimen térmico más cálido, con valores de temperatura mínimos promedio de 23 °C y promedios máximos de 29 °C, con una marcada amplitud térmica diaria. Las precipitaciones intensas son características de esta estación, con acumulados que superan los 600 mm presentes en horas del mediodía, hasta alcanzar valores máximos de 1150 mm acumulados en horas vespertinas. Este comportamiento está asociado a sistemas convectivos costeros, característicos de la región.

Precipitación acumulada y temperatura promedio por hora, estación "Unipacífico"

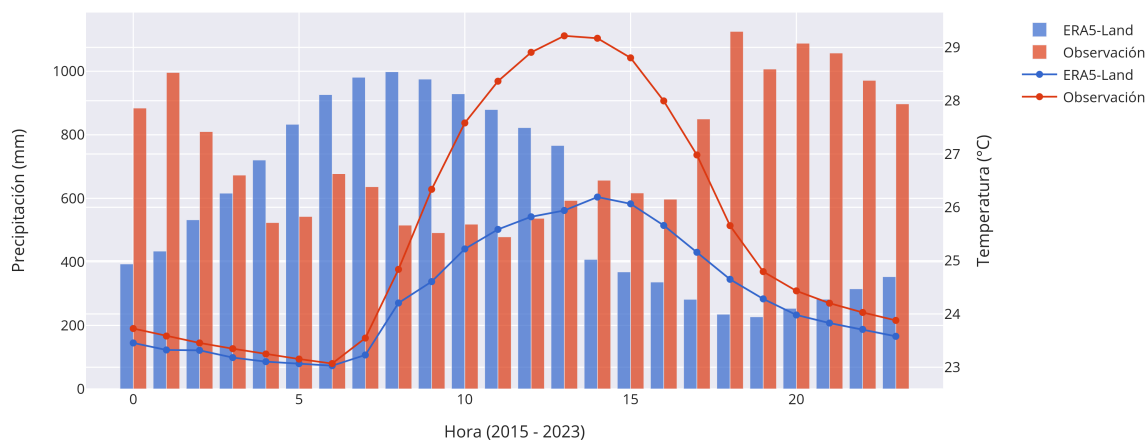


Figura 4.12: Precipitación total y temperatura promedio por hora entre el año 2015 y 2023. Datos estimados por ERA5-Land y las mediciones realizadas por la estación "Unipacífico".

4.1.4. Componentes principales

En la caracterización del clima de la cuenca a partir de la construcción de las componentes principales, reveló que las primeras dos componentes (CP1 y CP2) explican el 70.72 % de la variabilidad total de las variables climáticas del GCM ('uas', 'vas', 'huss', 'mrsos', 'sp', 'rlds', 'rlus', 'rsds', 'rsus') ver tabla 3.1, donde CP1 contribuye con 40.30 % y CP2 con 30.42 %. Las siguientes cuatro componentes (CP3-CP6) aportan conjuntamente un 24.85 % adicional, con contribuciones individuales que van desde 8.82 % hasta 3.95 %. En conjunto, las seis componentes principales seleccionadas logran capturar el 95.57 % de la varianza total del sistema, proporcionando una representación eficiente de la variabilidad de los datos originales.

Tabla 4.7: Porcentaje de varianza explicada por cada componente principal establecida.

Componentes principales	
Dimensión	Varianza explicada (%)
CP1	40.30
CP2	30.42
CP3	8.82
CP4	6.91
CP5	5.17
CP6	3.95

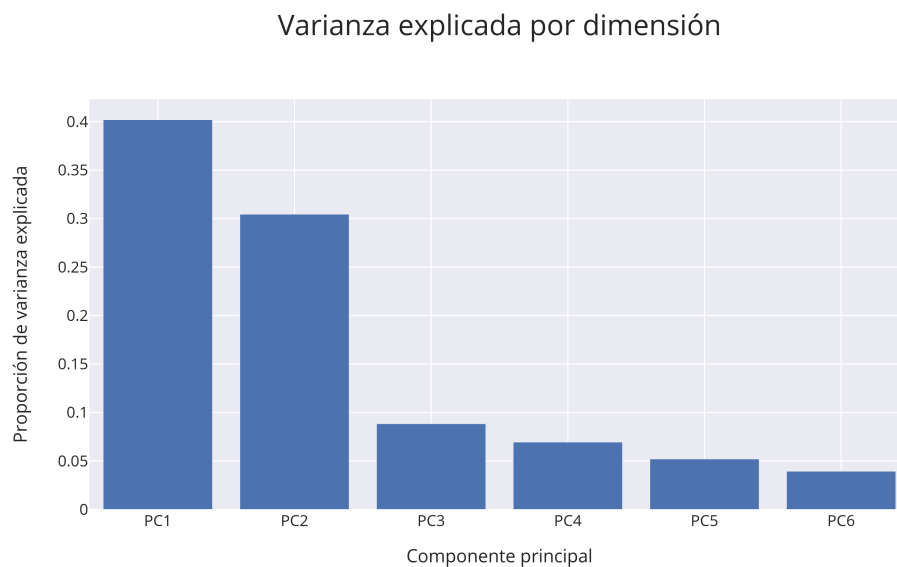


Figura 4.13: Distribución de la varianza explicada por componentes principales de las variables climáticas del GCM.

El marcado dominio de las dos primeras componentes principales, que explican más del 70 % de la varianza, sugiere una fuerte estructura de correlación en los datos originales. Este patrón indica que gran parte de la información climática puede ser capturada eficientemente mediante un número reducido de variables transformadas.

La notable diferencia entre la varianza explicada por las dos primeras componentes (CP1 y CP2) y las componentes subsiguientes (CP3-CP6) señala una clara jerarquía en la importancia de los patrones climáticos subyacentes. Esta reducción dimensional a seis componentes principales

resulta particularmente ventajosa para el desarrollo del modelo de downscaling estadístico, ya que permite mantener la mayor parte de la información relevante mientras se reduce significativamente la complejidad del sistema.

Además, considerando que los datos climáticos a resolución horaria generan matrices de gran dimensionalidad, esta reducción de variables representa una estrategia crucial para optimizar el rendimiento computacional de los modelos, disminuyendo los tiempos de procesamiento y los requerimientos de memoria sin comprometer significativamente la calidad de la información climática utilizada.

4.2. *Downscaling* estadístico para datos de temperatura y precipitación

En esta sección se presentan los resultados del downscaling estadístico a datos de temperatura y precipitación del GCM, implementado mediante cuatro enfoques de aprendizaje automático supervisado: Regresión Lineal (RL) como línea base paramétrica, Random Forest (RF) y Regresión de Vectores de Soporte (SVR) y Redes Neuronales Artificiales con arquitectura ConvLSTM2D. Esta selección abarca desde técnicas clásicas hasta métodos avanzados de deep learning, permitiendo evaluar el compromiso entre complejidad computacional y capacidad predictiva en el contexto de reducción de escala de las variables temperatura y precipitación. Estos modelos han demostrado su eficacia en estudios previos de downscaling estadístico, destacando particularmente en el tratamiento de series temporales climáticas.

La evaluación del desempeño de los modelos de downscaling estadístico se realiza mediante un conjunto complementario de métricas de validación que permiten caracterizar diferentes aspectos de la calidad predictiva. El Error Cuadrático Medio (ECM) y su raíz (RECM) se utilizan para cuantificar la magnitud de los errores de predicción, penalizando especialmente las desviaciones grandes. El Error Absoluto Medio (EAM) proporciona una medida más fuerte de la precisión general del modelo, mientras que el Error Porcentual Absoluto Medio (EPAM) permite comparar el desempeño entre diferentes escalas y variables climáticas modeladas. Complementariamente, el coeficiente de determinación (R^2) evalúa la proporción de varianza explicada por el modelo, y los coeficiente de correlación de Pearson (R) y Spearman (S) mide la fuerza y dirección de la relación lineal entre los valores observados y predichos. Esta combinación de métricas ofrece una evaluación integral que abarca tanto la precisión absoluta como la capacidad de los modelos para capturar la variabilidad espacio-temporal de las variables climáticas.

El desarrollo de los modelos se realizó utilizando datos del período 2015-2023, implementando una estrategia de validación que permite evaluar la precisión y generalización de los modelos. El conjunto de predictores base incluye las variables geográficas fundamentales (longitud, latitud

y altura), denotadas como "Geo", que capturan la variabilidad espacial del área de estudio. Posteriormente, se adoptó un enfoque incremental, incorporando secuencialmente las componentes principales (CP) derivadas del ACP previo. Esta metodología permite evaluar la contribución marginal de cada componente principal en el desempeño del modelo, cuantificado mediante las métricas de evaluación establecidas.

4.2.1. Regresión Lineal

El modelo de regresión lineal muestra comportamientos contrastantes entre las variables modeladas. Los resultados del modelamiento mediante regresión lineal se presentan separadamente para precipitación y temperatura. Primero las Tablas 4.8 y 4.10 muestran las métricas de validación para las variables temperatura y precipitación respectivamente, considerando diferentes configuraciones de predictores basadas en el ACP, desde el uso de las variables geoespaciales ("Geo": Longitud, Latitud y altura) hasta la incorporación progresiva de componentes principales (1CP hasta 6CP).

Luego se presentan los resúmenes de los modelos de RL para temperatura (Figura: 4.14) y precipitación (Figura: 4.17). Se realiza un análisis inferencial sobre los coeficientes del modelo, evaluando su significancia estadística y el cumplimiento de los supuestos de la regresión lineal mediante el estudio de los residuos.

Para la predicción de la temperatura, el modelo muestra un desempeño sólido, con coeficientes de determinación R^2 superiores a 0.831 en todas las configuraciones evaluadas. El mejor resultado se obtiene con la configuración de 6CP, alcanzando un $R^2 = 0.848$, lo que indica que el 84.8 % de la variabilidad observada en la temperatura es explicada por las 10 variables incluidas en el modelo de regresión. Además, el p – *valor* del modelo es 0.000, lo que sugiere que al menos uno de las variables predictoras es estadísticamente significativa.

Tabla 4.8: Métricas de evaluación del modelo de regresión lineal para temperatura, utilizando la componente geoespacial "Geo" y las componentes principales generadas.

Regresión lineal Temperatura						
Variables	ECM	RECM	EAM	EPAM	R	R^2
Geo	3.781	1.944	1.488	0.51 %	0.912	0.831
Geo, 1CP	3.769	1.94	1.478	0.51 %	0.914	0.835
Geo, 2CP	3.731	1.931	1.467	0.50 %	0.916	0.839
Geo, 3CP	3.661	1.913	1.458	0.50 %	0.916	0.839
Geo, 6CP	3.426	1.850	1.423	0.49 %	0.921	0.848

En términos de error, la Raíz del Error Cuadrático Medio (RECM) revela que las predicciones del modelo, en promedio, se desvían en ± 1.85 °C con respecto a los valores observados, mientras que el Error Absoluto Medio (EAM) indica una desviación promedio de 1.42 °C. Además, el Error Porcentual Absoluto Medio (EPAM) se mantiene consistentemente bajo, alrededor de 0.50 %, lo que refuerza la capacidad del modelo para realizar predicciones precisas y confiables de la temperatura, con un rendimiento general adecuado.

La imagen 4.14 muestra el resumen del modelo RL aplicado a la variable de temperatura, evaluando la relación entre el set completo de datos con 5282079 registros y las 10 variables predictoras seleccionadas, permitiendo identificar el impacto de cada predictor sobre la variable dependiente, proporcionando una visión detallada del comportamiento de la temperatura en función de las variables explicativas consideradas.

OLS Regression Results						
Dep. Variable:	y	R-squared (uncentered):	0.857			
Model:	OLS	Adj. R-squared (uncentered):	0.857			
Method:	Least Squares	F-statistic:	3.165e+06			
Date:	Tue, 01 Apr 2025	Prob (F-statistic):	0.00			
Time:	15:41:00	Log-Likelihood:	-2.3590e+06			
No. Observations:	5282079	AIC:	4.718e+06			
Df Residuals:	5282069	BIC:	4.718e+06			
Df Model:	10					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
x1	0.2159	0.001	331.262	0.000	0.215	0.217
x2	-0.5193	0.001	-476.642	0.000	-0.521	-0.517
x3	0.1728	0.001	260.505	0.000	0.172	0.174
x4	0.3036	0.001	326.258	0.000	0.302	0.305
x5	0.2437	0.001	457.701	0.000	0.243	0.245
x6	-0.0932	0.000	-208.962	0.000	-0.094	-0.092
x7	0.1319	0.000	619.315	0.000	0.132	0.132
x8	-0.1875	0.001	-362.832	0.000	-0.189	-0.186
x9	-0.1159	0.000	-346.579	0.000	-0.117	-0.115
x10	-0.7061	0.000	-2433.228	0.000	-0.707	-0.705
Omnibus:	90982.564	Durbin-Watson:	0.875			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	175648.161			
Skew:	0.085	Prob(JB):	0.00			
Kurtosis:	3.877	Cond. No.	17.9			

Figura 4.14: Resultados del modelo RLM para los valores de temperatura. El resumen detalla el ajuste del modelo, coeficientes y significancia estadística.

Fuente: elaboración propia.

El análisis de los coeficientes del modelo muestra que todas las variables predictoras son estadísticamente significativas con $p - \text{valor} < 0.05$. En particular, la variable x_{10} (altura) tiene el mayor impacto negativo en la variable objetivo con un coeficiente de -0.706, mientras que x_4 (CP1) presenta el mayor impacto positivo con un coeficiente de 0.303.

Para la evaluación del supuesto de normalidad de los residuos, la prueba Jarque-Bera presentada en la Figura: 4.14 y Shapiro-Wilk en la tabla: 4.9, cuyos resultados de $p - \text{valor} = 0.000$ sugieren que los residuos del modelo RL no siguen una distribución normal, lo que puede estar asociado con la presencia de valores atípicos, que son mejor observados en el gráfico Q-Q,

Tabla 4.9: Prueba de supuestos del modelo de RL para la temperatura.

Pruebas para la temperatura		
Prueba	Estadístico	P-Value
Shapiro-Wilk	0.993	0.000
Breusch-Pagan	567215	0.000
Durbin-Watson	0.85	

Figura 4.16.

En el histograma, Figura: 4.15 se observa que la distribución es aproximadamente normal, aunque se identifican pequeñas desviaciones en las colas, lo que sugiere una ligera asimetría en los residuos. La línea de densidad refleja esta tendencia, indicando que, si bien no se evidencian problemas graves, es recomendable realizar una revisión más detallada.

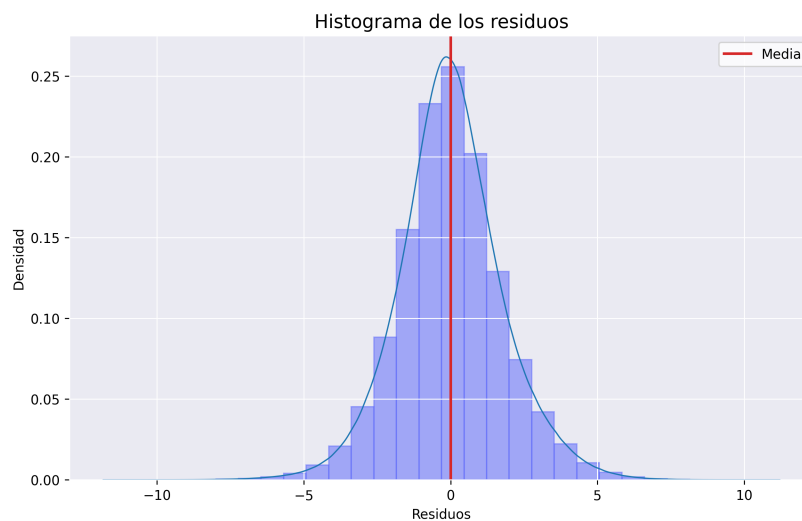


Figura 4.15: Histograma de los residuos de la temperatura.

Fuente: elaboración propia.

Por otro lado, el gráfico Q-Q muestra que la mayoría de los puntos siguen la línea de referencia, lo que sugiere que la normalidad se cumple en la parte central de la distribución. Sin embargo, se observa una curvatura en los extremos, lo que indica la presencia de colas más gruesas de lo esperado en una distribución normal. Esto sugiere que los residuos presentan valores extremos con mayor frecuencia de la esperada, lo que afecta la validez de las inferencias estadísticas basadas en la normalidad.

La homocedasticidad se analizó mediante la prueba de Breusch-Pagan, obteniendo un

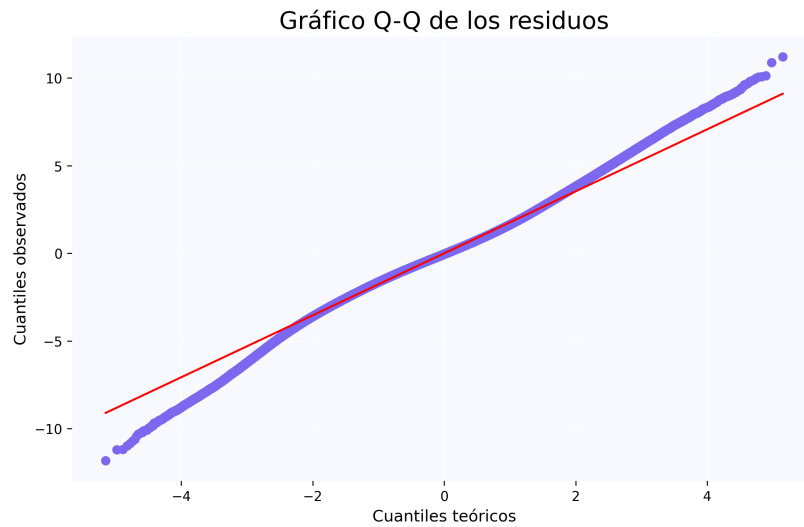


Figura 4.16: Q-Q Plot de los residuos del modelo RL para la temperatura.

Fuente: elaboración propia.

$p - valor = 0.000$, lo que indica la presencia de heterocedasticidad en los residuos.

Finalmente, la independencia de los residuos se evaluó utilizando la estadística de Durbin-Watson ($DW = 0.85$). Dado que este valor es cercano a 0, sugiere la presencia de autocorrelación positiva en los residuos, lo que podría afectar la validez de las inferencias del modelo.

En contraste, para la precipitación el modelo de regresión lineal presenta limitaciones, con coeficientes de determinación con R^2 menores a 0.04 en todas las configuraciones evaluadas. Esto implica que el modelo únicamente logra explicar el 3.8 % de la variabilidad observada, lo cual es consistente con la naturaleza altamente variable y no lineal de la precipitación presente en los datos, analizados en la sección anterior.

La RECM de 0.61 mm/h, indica que las predicciones presentan, en promedio, una desviación de ± 0.61 mm/h respecto a los valores observados. Asimismo, el ECM cercano a 0.38 mm, muestra una falta de mejora significativa incluso con la incorporación de componentes principales adicionales, lo que refuerza las limitaciones del modelo en capturar patrones más complejos de precipitación.

De las métricas de correlación se observa que el coeficiente de Spearman de 0.22 sugiere una correlación débil basada en el orden de los datos, mientras que el coeficiente de Pearson de 0.19 confirma una correlación lineal igualmente débil. El hecho de que Spearman supere a Pearson indica la posible existencia de una relación monótona entre las variables, aunque no necesariamente lineal, lo que es coherente con la distribución no normal de la precipitación y la prevalencia de

numerosos valores nulos o cercanos a cero que se presentan en esta variable a una resolución temporal horaria.

Tabla 4.10: Métricas de evaluación del modelo de regresión lineal para precipitación utilizando datos geospaciales "Geo" y las componentes principales generadas.

Regresión lineal Precipitación						
Variables	ECM	RECM	EAM	S	R	R ²
Geo	0.386	0.621	0.3487	0.207	0.174	0.029
Geo, 1CP	0.385	0.621	0.3482	0.212	0.179	0.032
Geo, 2CP	0.385	0.621	0.3447	0.219	0.182	0.033
Geo, 3CP	0.383	0.619	0.3447	0.226	0.196	0.037
Geo, 6CP	0.383	0.619	0.3421	0.229	0.196	0.038

En la Imagen 4.17 se presenta el resumen del RL aplicado a la precipitación, donde se analiza la relación entre la variable dependiente y un conjunto de 10 variables predictoras, utilizando un total de 5281540 registros, cuantificando la influencia de cada predictor sobre la precipitación.

OLS Regression Results						
Dep. Variable:	y	R-squared (uncentered):	0.250			
Model:	OLS	Adj. R-squared (uncentered):	0.250			
Method:	Least Squares	F-statistic:	1.760e+05			
Date:	Thu, 03 Apr 2025	Prob (F-statistic):	0.00			
Time:	01:54:55	Log-Likelihood:	8.5841e+06			
No. Observations:	5281550	AIC:	-1.717e+07			
Df Residuals:	5281540	BIC:	-1.717e+07			
Df Model:	10					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
x1	-0.0115	0.000	-77.446	0.000	-0.012	-0.011
x2	-0.0078	0.000	-28.385	0.000	-0.008	-0.007
x3	0.0274	0.000	68.170	0.000	0.027	0.028
x4	-0.0220	0.001	-41.203	0.000	-0.023	-0.021
x5	0.0065	0.000	29.470	0.000	0.006	0.007
x6	0.0785	0.000	225.686	0.000	0.078	0.079
x7	0.0017	0.000	6.805	0.000	0.001	0.002
x8	0.0299	0.000	85.874	0.000	0.029	0.031
x9	-0.0005	0.000	-1.947	0.051	-0.001	3.1e-06
x10	-0.0108	0.000	-82.607	0.000	-0.011	-0.011
Omnibus:	5098726.287	Durbin-Watson:	0.319			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	266820121.918			
Skew:	4.743	Prob(JB):	0.00			
Kurtosis:	36.503	Cond. No.	50.5			

Figura 4.17: Resultados del modelo RLM para los valores de precipitación. El resumen detalla el ajuste del modelo, coeficientes y significancia estadística.

Fuente: elaboración propia.

El análisis de los coeficientes del modelo para la precipitación muestra que todas las variables predictoras son estadísticamente significativas con $p - valor = 0.00$, excepto $x9$ (CP6) con un

$p - valor = 0.051$ que indica que esta ultima componente no influye significativamente en la predicción del modelo. Del mismo modo, la variable x_6 (CP3) tiene el mayor impacto positivo en la variable objetivo con un coeficiente de 0.078, mientras que x_4 (CP1) presenta el mayor impacto negativo con un coeficiente de -0.022.

Tabla 4.11: Prueba de supuestos del modelo de RL para la precipitación.

Pruebas para la temperatura		
Prueba	Estadístico	P-Value
Shapiro-Wilk	0.579	0.000
Breusch-Pagan	109606	0.000
Durbin-Watson	0.318	

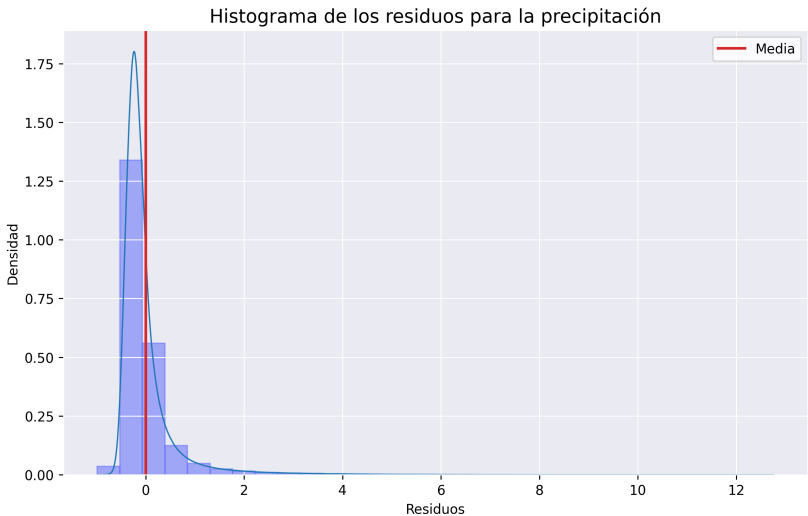


Figura 4.18: Histograma de los residuos de la precipitación.

Fuente: elaboración propia.

El histograma de los residuos de la regresión para la precipitación revela una distribución asimétrica a la derecha, con una alta concentración de valores cerca de cero y una cola extendida hacia valores positivos. Esto sugiere que, en general, los errores de predicción son pequeños, aunque algunos valores atípicos pueden estar afectando la normalidad de los residuos. La media de los residuos se encuentra cercana a cero; sin embargo, la mayoría de los valores son negativos, lo cual no es ideal para la variable precipitación. Este comportamiento está fuertemente influenciado por la escala temporal horaria de los datos, lo que representa un desafío significativo para la modelación de la precipitación a una resolución tan detallada.

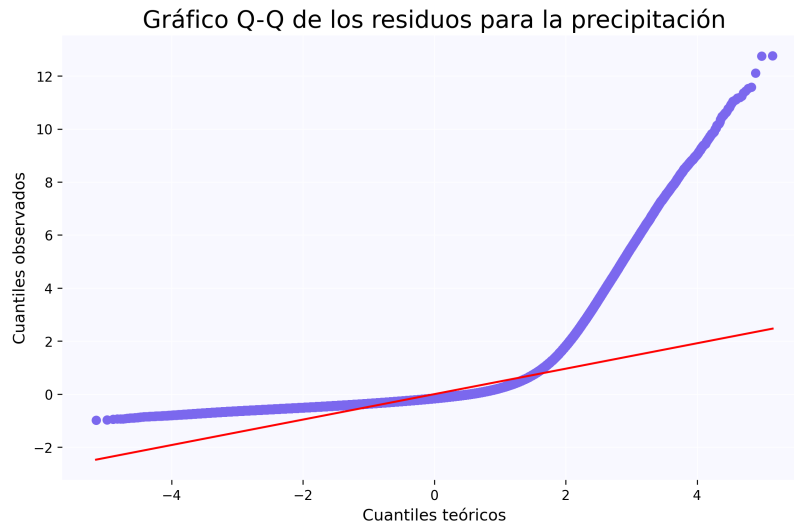


Figura 4.19: Q-Q Plot de los residuos del modelo RL para la precipitación.

Fuente: elaboración propia.

El gráfico de cuantiles teóricos (Q-Q plot) presentado en la Figura 4.19 permite evaluar la normalidad de los residuos del modelo de precipitación de una manera visual. En la región central de la distribución cercana a 0, los residuos se alinean aproximadamente con la línea de referencia, lo que indica que la mayoría de los valores siguen la distribución teórica esperada. No obstante, en los extremos o colas, se observan desviaciones significativas, expresando que los residuos para la precipitación no presentan el supuesto de normalidad.

En la cola derecha, los residuos exhiben una curvatura pronunciada hacia arriba, lo que sugiere la presencia de valores atípicos o errores de predicción más elevados en el rango positivo. Por otro lado, en la cola izquierda, los residuos muestran una menor dispersión y un comportamiento más aplanado, lo que indica una menor presencia de valores atípicos en la dirección negativa. Estas observaciones sugieren que la distribución de los residuos presenta colas más pesadas de lo esperado bajo una distribución normal.

La normalidad de los residuos por medio de las pruebas de Jarque-Bera ($p\text{-valor} = 0.000$) y Shapiro-Wilk ($p\text{-valor} = 0.000$) indican que los residuos no siguen una distribución normal. Esto sugiere la presencia de valores atípicos o una estructura no capturada por el modelo, lo que puede afectar la validez de las inferencias estadísticas.

La prueba de Breusch-Pagan ($p\text{-valor} = 0.000$) confirma la presencia de heterocedasticidad en los residuos, lo que implica que la variabilidad de los errores no es constante a lo largo de las predicciones. Esto puede afectar la eficiencia de los estimadores para la variable precipitación.

El estadístico Durbin-Watson (valor = 0.318) sugiere una fuerte autocorrelación positiva en los residuos, lo que indica que los errores de predicción en la precipitación están correlacionados en el tiempo o el espacio. Esto puede ser problemático en el modelamiento de series temporales y datos espaciales implícitos en la precipitación.

Los resultados muestran que la regresión lineal es un buen modelo base para el *downscaling* estadístico de temperatura por hora, pero deficiente para el modelamiento de la precipitación. Esta disparidad con la precipitación se atribuye a la naturaleza de ambas variables con un comportamiento y patrones diferentes en espacio y tiempo: mientras la temperatura presenta patrones de variación más suaves y relaciones más lineales con las variables predictoras, la precipitación se caracteriza por su comportamiento no lineal, intermitente y altamente variable.

La incorporación progresiva de componentes principales muestra una mejora en el desempeño del modelo para temperatura, confirmando que las primeras componentes capturan efectivamente la información más relevante. Sin embargo, para precipitación, ni siquiera la inclusión de componentes adicionales logra mejorar significativamente el desempeño.

4.2.2. Random Forest

Los resultados del modelo Random Forest para *downscaling* de temperatura y precipitación se realizó evaluando diferentes configuraciones del número de estimadores (E10, E25, E50 y E100) combinado con la selección de predictores basadas en las variables espacial "Geo" (Longitud, Latitud y Altura) y en la incorporación de las componentes principales. Este proceso de calibración busca encontrar el equilibrio óptimo entre capacidad predictiva y eficiencia computacional.

4.2.2.1. Resultados Random forest para la temperatura

Los resultados muestran un patrón consistente de mejora en el desempeño del modelo RF para la predicción de la temperatura al incrementar tanto el número de estimadores como el número de componentes principales. Para la configuración con 10 estimadores (E10), el R^2 mejora progresivamente desde 0.862 utilizando solo datos geoespaciales "Geo", hasta alcanzar 0.918 al incluir 6CP.

Del mismo modo, la RECM se reduce de 1.75 °C a 1.35 °C, mientras que el EMA disminuye de 1.32 °C a 1.04 °C. El Error Porcentual Absoluto Medio (EPAM) también refleja una mejora significativa, reduciéndose de 0.45 % a 0.36 %.

Tabla 4.12: Métricas de validación del modelo Random Forest para temperatura con 10 estimadores, evaluando el impacto progresivo de las componentes principales.

Random forest Temperatura E10						
Variables	ECM	RECM	EAM	EPAM	R	R ²
Geo, 6CP	1.832	1.353	1.040	0.36 %	0.958	0.918

Random forest Temperatura E10						
Variables	ECM	RECM	EAM	EPAM	R	R ²
Geo, 6CP	1.832	1.353	1.040	0.36 %	0.958	0.918

Las configuraciones con mayor número de estimadores (E25, E50 y E100) muestran un desempeño notablemente, alcanzando valores de R² de 0.924, con un EPAM constante de 0.34 % indicando la capacidad general del modelo para realizar buenas predicciones. Sin embargo, las diferencia en el desempeño entre las configuraciones E25, E50 y E100 es marginal, lo que sugiere una convergencia en la capacidad predictiva del modelo al duplicar el número de estimadores.

Tabla 4.13: Métricas de validación del modelo Random Forest para temperatura con configuraciones de 25, 50 y 100 estimadores utilizando 6 componentes principales.

Random forest Temperatura E25						
Variables	ECM	RECM	EAM	EPAM	R	R ²
Geo, 6CP	1.710	1.307	1.005	0.34 %	0.961	0.923

Random forest Temperatura E50						
Variables	ECM	RECM	EAM	EPAM	R	R ²
Geo, 6CP	1.710	1.307	1.005	0.34 %	0.961	0.923

Random forest Temperatura E100						
Variables	ECM	RECM	EAM	EPAM	R	R ²
Geo, 6CP	1.694	1.301	1.001	0.34 %	0.962	0.924

El análisis de la evaluación revela varios aspectos significativos del modelamiento con Random Forest para downscaling estadístico de temperatura. En primer lugar, el modelo RF mejora en el desempeño respecto al modelo de RL, evidenciado por valores más altos de R² (RL: 0.848 frente a RF: 0.925) y menores errores registrados, con un ECM reducido de 3.426 (RL) a 1.694 (RF) y

un EPAM que pasa de 0.49 % (RL) a 0.34 % (RF). Esto confirma la capacidad del Random Forest para capturar relaciones espacio-temporales presentes en los datos.

La convergencia en el desempeño observada a partir de 25 estimadores sugiere un punto óptimo en la arquitectura del modelo. El incremento marginal en la precisión con 50 y 100 estimadores no justifica el aumento significativo en los requisitos computacionales, considerando que el uso de memoria RAM supera los 50 GB para E100. Esta observación es particularmente relevante en el contexto de modelamiento operacional, donde la eficiencia computacional es crucial.

La incorporación de componentes principales tiene un impacto positivo más notable en configuraciones con menos estimadores (E25), mientras que en configuraciones más complejas (E50-E100) el beneficio es menos pronunciado. Este comportamiento sugiere que, a medida que se incrementa el número de estimadores, el modelo se vuelve menos dependiente de las nuevas variables predictoras para mejorar su desempeño.

Desde una perspectiva práctica, estos resultados indican que una configuración del modelo Random Forest con 25 estimadores y 6 componentes principales representa un equilibrio apropiado entre precisión predictiva y eficiencia computacional para el downscaling estadístico de temperatura.

4.2.2.2. Resultados Random forest para la precipitación

Los resultados también revelan un desempeño limitado del modelo Random Forest para la precipitación. En la configuración con 10 estimadores (E10), el R^2 muestra valores muy bajos, aumentando de 0.025 (utilizando "Geo") a 0.067 (con 6CP). Las configuraciones con mayor número de estimadores (E25, E50 y E100) muestran un comportamiento similar, con R^2 bajos (0.076, 0.081 y 0.085 respectivamente).

El Error Cuadrático Medio (ECM) muestra una ligera mejora al aumentar el número de estimadores, reduciéndose de 0.514 a 0.417. Sin embargo, los coeficientes de correlación Pearson y Spearman se mantienen consistentemente bajos, no superando el 0.259 y 0.296 respectivamente.

Tabla 4.14: Métricas de evaluación modelo Random Forest para precipitación con 10 estimadores, evaluando la incorporación progresiva de componentes principales.

Random forest Precipitación E10						
Variables	ECM	RECM	EAM	S	R	R ²
Geo	0.514	0.717	0.382	0.190	0.158	0.025
Geo, 1CP	0.445	0.667	0.362	0.249	0.210	0.044
Geo, 2CP	0.429	0.655	0.357	0.272	0.239	0.057
Geo, 3CP	0.426	0.652	0.356	0.292	0.255	0.065
Geo, 6CP	0.417	0.646	0.356	0.296	0.259	0.067

Tabla 4.15: Métricas de evaluación del modelo Random Forest para precipitación con 25, 50 y 100 estimadores utilizando 6 componentes principales.

Random forest Precipitación E25						
Variables	ECM	RECM	EAM	S	R	R ²
Geo, 6CP	0.395	0.629	0.347	0.281	0.277	0.076

Random forest Precipitación E50						
Variables	ECM	RECM	EAM	S	R	R ²
Geo, 6CP	0.388	0.623	0.344	0.289	0.286	0.081

Random forest Precipitación E100						
Variables	ECM	RECM	EAM	S	R	R ²
Geo, 6CP	0.384	0.619	0.343	0.298	0.292	0.085

Los resultados del modelamiento de precipitación mediante RF revelan varios aspectos críticos. Primero, a pesar de la capacidad del Random Forest para capturar relaciones no lineales, el modelo no logra mejorar significativamente el desempeño respecto al modelo de RL para esta variable, lo que sugiere que la complejidad del problema va más allá de la no linealidad.

El poco aumento del R² al incorporar componentes principales, contrario a lo observado en temperatura, indica que la reducción dimensional podría no estar proporcionando información crucial para la predicción de precipitación. Esto sugiere que los patrones de precipitación podrían estar relacionados con interacciones más complejas entre las variables originales que no son capturadas eficientemente por las componentes principales.

El coeficiente de correlación de Spearman (S) muestra un leve incremento desde 0.1901 hasta 0.2958 conforme se incorporan más componentes principales, en contraste con el comportamiento débil observado en el R^2 . Esta tendencia es particularmente relevante dado que el coeficiente de Spearman es más apropiado para evaluar relaciones monótonas no lineales y es menos sensible a valores extremos, características inherentes a los datos de precipitación.

4.2.3. Regresión de vectores de soporte

Los resultados del modelo de Regresión de Vectores de Soporte (RVS) para el *downscaling* estadístico, siguiendo un proceso de dos etapas. En la primera etapa, se realizó una validación cruzada para optimizar los hiperparámetros del modelo utilizando el registro de un año de datos a una resolución temporal de 3 horas (545,720 registros), obteniendo los valores óptimos de "C", "epsilon" y "gamma". Posteriormente, el resultado de estos parámetros optimizados se utilizaron para entrenar el modelo con un conjunto de datos extendido a un registro de dos años (1,093,440 registros horarios), permitiendo evaluar la robustez y escalabilidad del modelo.

4.2.3.1. Resultados Regresión de vectores de soporte para temperatura

El análisis del modelo RVS revela varios aspectos significativos en el contexto del *downscaling* estadístico para los datos de temperatura. Los hiperparámetros optimizados ($C=10$, $\epsilon=0.1$, $\gamma='auto'$) sugieren un equilibrio efectivo entre la capacidad de generalización y la precisión del modelo.

Los resultados muestran un buen desempeño y consistente del modelo RVS en ambos escenarios temporales. Para el conjunto de un año, el modelo alcanza un R^2 de 0.919 con un EPAM de 0.46 %, mientras que para dos años de datos, mantiene un desempeño similar con un R^2 de 0.925 y un EPAM de 0.43 %. Los errores de predicción, medidos por el ECM, muestran una ligera mejoría al aumentar el tamaño del conjunto de datos, reduciéndose de 1.892 °C a 1.656 °C.

La ligera mejora en todas las métricas al aumentar el tamaño del conjunto de datos indica que el modelo se beneficia del mayor volumen de información sin mostrar signos de sobreajuste, validando tanto la elección de los hiperparámetros como la capacidad del RVS para manejar eficientemente un gran volumen de datos horarios.

Tabla 4.16: Métricas de evaluación del modelo de Regresión de Vectores de Soporte para temperatura utilizando un año de datos (545720 registros)

Temperatura 545720 Registros						
Variables	ECM	RECM	EAM	EPAM	R	R^2
Geo	1.892	1.375	1.052	0.46 %	0.959	0.919

Tabla 4.17: Métricas de evaluación del modelo de Regresión de Vectores de Soporte para temperatura utilizando dos año de datos (1093440 registros)

Temperatura 1093440 Registros						
Variables	ECM	RECM	EAM	EPAM	R	R ²
Geo	1.656	1.287	0.958	0.43 %	0.962	0.925

El análisis del modelo RVS destaca aspectos significativos en el contexto del *downscaling* de temperatura. En primer lugar, la consistencia en las métricas de desempeño entre ambos períodos de tiempo sugiere una robustez del modelo, capaz de mantener su capacidad predictiva al duplicar el volumen de datos de entrenamiento.

La ligera mejora observada en todas las métricas al aumentar el tamaño del conjunto de datos (reducción del RECM de 1.375 °C a 1.287 °C, reducción del EAM 1.052 °C a 0.958 °C, reducción de EPAM de 0.46 % a 0.43 %, y aumento del R² de 0.919 a 0.925) indica que el modelo se beneficia moderadamente de la información adicional proporcionada por un conjunto de datos más extenso. Esta mejora, aunque modesta, sugiere que el modelo logra capturar patrones más detallados de la variabilidad espacio-temporal de la temperatura.

Es notable la capacidad del RVS para mantener un alto nivel de precisión con conjuntos de datos extensos, considerando que se trata de registros horarios que capturan la variabilidad diurna y estacional completa. Los valores de correlación superiores a 0.95 en ambos casos confirman la capacidad del modelo para reproducir en forma apropiada los patrones de temperatura a escala horaria. Cabe mencionar que la ejecución del modelo con esta gran cantidad de datos no requiere de muchos recursos computacionales, pero su tiempo de ejecución fue de aproximadamente 2 días para el entrenamiento con dos años de registros.

4.2.3.2. Resultados Regresión de vectores de soporte para precipitación

Los resultados de la validación entre los hiperparámetros del modelo de RVS para el *downscaling* estadístico para la precipitación, arrojó los siguientes valores: C=1, epsilon=0.01, gamma='Scale'. Los resultados para esta partición que incluye 545720 registros, el modelo mostró un desempeño limitado para la predicción de los valores de precipitación, obteniendo un valor de R² de 0.039, un coeficiente de correlación de Pearson: 0.198 y Spearman: 0.213. El resultado de las métricas de los errores fueron muy consistentes con los resultados obtenido en los modelos previos (RL y RF), con un ECM: 0.397, RECM: 0.630 y EAM: 0.371, tabla: 4.18.

El análisis extendido a un periodo de dos años con los hiperparametros obtenidos, el modelo RVS mantuvo un desempeño igualmente limitado. Los resultados muestran de R² de 0.039 y los coeficientes de Person:0.199 y Spearman: 0.220 apenas logran un leve incremento. De manera

similar las métricas de error muestran una disminución con poca importancia al aumentar el número de registros, ver tabla: 4.19.

Tabla 4.18: Métricas de evaluación del modelo de Regresión de vectores de soporte con el número de registros de un año.

Precipitación 545720 Registros						
Variables	ECM	RECM	EAM	S	R	R ²
Geo	0.394	0.628	0.341	0.213	0.198	0.039

Tabla 4.19: Métricas de evaluación del modelo de Regresión de vectores de soporte con el número de registros de dos años

Precipitación 1093440 Registros						
Variables	ECM	RECM	EAM	S	R	R ²
Geo	0.388	0.611	0.334	0.215	0.199	0.039

El análisis del modelo RVS para precipitación revela varios aspectos en el contexto del downscaling. A pesar de utilizar los hiperparámetros que generaron los mejores resultados, el modelo también muestra limitaciones significativas para capturar los patrones de precipitación como en los modelos anteriores de RL y RF, evidenciado por los bajos valores de R² menores a 0.04 en ambos períodos temporales evaluados.

Sin embargo, es notable que el coeficiente de correlación de Spearman muestra valores ligeramente superiores (0.213 y 0.215) en comparación con el coeficiente de correlación lineal Pearson, sugiriendo que el modelo captura mejor las relaciones monótonas y el orden de magnitud de los eventos de precipitación que sus valores exactos. Esta característica es particularmente relevante en datos horarios, donde la naturaleza intermitente y no lineal de la precipitación se hace más evidente.

La estabilidad en las métricas al duplicar el período de análisis indica que las limitaciones del modelo no están relacionadas con el tamaño del conjunto de datos, sino con la complejidad inherente de la precipitación a escala horaria. La ligera mejora en el ECM (de 0.3938 a 0.3881) sugiere que el modelo mejora ligeramente de una mayor representación de datos, aunque sin lograr una mejora sustancial en su capacidad predictiva general.

4.2.4. Modelamiento con redes neuronales ConvLSTM2D

Para abordar el problema de *downscaling* estadístico se desarrolló una red neuronal con arquitectura ConvLSTM2D aplicada a los datos de temperatura y precipitación del modelo global GFDL-ESM4 y datos de ERA5-Land. Esta arquitectura integra las capacidades de las redes convolucionales para identificar patrones espaciales complejos con la habilidad de las redes LSTM para capturar dependencias temporales. De este modo, el modelo no solo reduce la resolución espacial, sino que también mejora la resolución temporal, permitiendo realizar predicciones de temperatura y precipitación al pasar de una resolución temporal de 3 horas a 1 hora.

El desempeño del entrenamiento de los modelos se analizó mediante la evaluación de la función de pérdida (loss) en cada partición del conjunto de datos, así como el cálculo de las métricas aplicadas a los conjuntos de entrenamiento, validación y prueba. Adicionalmente, se examina la distribución espacial del Error Cuadrático Medio (ECM) y la correlación para el conjunto de datos de prueba, que corresponde a datos no observados por el modelo durante el entrenamiento. Este análisis permite una evaluación de la capacidad predictiva y generalización del modelo.

Además se generaron gráficos de series temporales para analizar el comportamiento de la temperatura y la precipitación modelada por las RNA ConvLSTM2D, en comparación con las estimaciones del modelo ERA5-Land y los datos observados en estaciones meteorológicas de estudio. Además se genera la tabla con las mejores métricas de evaluación de los modelos planteados en esta investigación para cada uno de las variables y se elaboraron los diagramas de Taylor para evaluar el desempeño del modelo en la representación de la temperatura y la precipitación con las mediciones de las estaciones locales.

4.2.4.1. Entrenamiento del modelo ConvLSTM2D para la temperatura

Los resultados de la evaluación del modelo se presentan en términos de la función de error (loss) utilizando las métricas de Error Cuadrático Medio (ECM) y el Error Absoluto Medio (EAM). Estas métricas son evaluadas para cada partición del conjunto de datos: entrenamiento, validación y prueba. Durante el entrenamiento, el ECM y EAM alcanzaron valores de 0.1017 y 0.2252, respectivamente, mientras que para la partición de validación los valores fueron ligeramente superiores 0.1035 y 0.2284. En el conjunto de prueba, los valores del ECM y EAM fueron 0.0992 y 0.2225, indicando un desempeño consistente y una adecuada generalización del modelo en datos no observados, tabla 4.20. Adicionalmente, el resultado final tras la optimización de la pérdida es el promedio del total de las "pérdidas" por lote, ya que el entrenamiento se realiza por muestras de "Batchsize" indicando que los valores de ECM y EAM se mantuvieron similares, reforzando la estabilidad del proceso de entrenamiento (Abadi y cols., 2016), (Gong, Wu, y Zhou, 2023).

Tabla 4.20: Métricas de evaluación de la pérdida (loss) de entrenamiento del modelo de ConvLSTM2D para la variable temperatura, aplicado a la partición de los datos.

Evaluación	Loss		Final	
	ECM	EAM	ECM	EAM
Entrenamiento	0.1017	0.2252	0.1012	0.2247
Validación	0.1035	0.2284	0.1027	0.2271
Prueba	0.0992	0.2225	0.1017	0.2251

La convergencia observada en los valores de ECM y EAM de la pérdida entre las particiones de entrenamiento, validación y prueba sugiere que el modelo logra una buena generalización, evitando problemas como el sobreajuste, donde la curva de la función de pérdida en el conjunto de datos de validación mantiene por debajo de la función de pérdida de entrenamiento, como se muestra en la figura 4.20. Los resultados finales confirman que la fase de optimización no introdujo cambios significativos en el desempeño del modelo, lo que se atribuye a una adecuada configuración de los hiperparámetros y regularización.

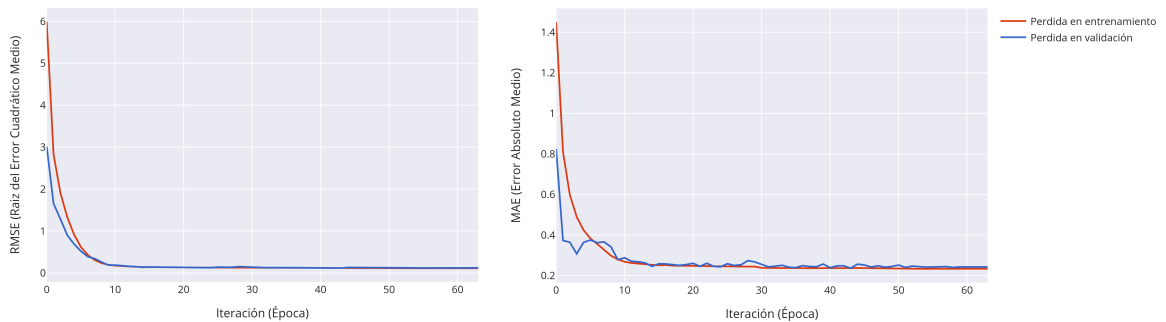


Figura 4.20: Evolución de las curvas de error durante el entrenamiento de la red ConvLSTM2D para la predicción de la temperatura. Métricas de evaluación RMSE a la izquierda y EAM a la derecha.

4.2.4.2. Evaluación del modelo ConvLSTM2D para la temperatura

El modelo ConvLSTM2D muestra un desempeño sobresaliente para el *downscaling* estadístico de temperatura, con resultados consistentes a través de las diferentes particiones de datos. Los valores de R^2 superiores a 0.95, la RECM de 1.01, un EAM de 0.76 de EPAM de 0.26 % y los valores de correlación de R de 0.97 en todas las particiones indican una excelente capacidad predictiva y generalización del modelo.

Tabla 4.21: Métricas de evaluación en el desempeño del modelo ConvLSTM2D en la predicción de la temperatura: métricas de evaluación por partición del set de datos.

Métricas de evaluación para la Temperatura						
Partición	ECM	RECM	EAM	EPAM	R	R ²
Entrenamiento	1.015	1.007	0.764	0.26 %	0.977	0.954
Validación	1.030	1.014	0.773	0.26 %	0.976	0.953
Prueba	1.020	1.001	0.766	0.26 %	0.976	0.954

Los resultados de la evaluación del modelo para la predicción de la temperatura se presentan en términos de métricas de error y desempeño estadístico, desglosadas por las particiones de entrenamiento, validación y prueba. El Error Cuadrático Medio (ECM) y el Error Absoluto Medio (EAM) muestran valores consistentes a lo largo de las particiones, con un ECM promedio cercano a 1.02 y un EAM de aproximadamente 0.765. El error porcentual absoluto medio (EPAM) se mantuvo constante en todas las particiones con un valor de 0.26 %, evidenciando un bajo nivel de error relativo. Las métricas de correlación y coeficiente de determinación representadas por R y R², también reflejan un alto desempeño del modelo, con valores superiores a 0.976 y 0.953, respectivamente, indicando una fuerte capacidad del modelo para capturar las variaciones de la temperatura en las diferentes particiones de los datos.

El análisis de las métricas sugiere que el modelo logra una representación bastante apropiada y consistente de la temperatura, con valores bajos de error y altos coeficientes de correlación. La estabilidad del EPAM en todas las particiones destaca una robusta capacidad del modelo para generalizar a datos no observados. Aunque las diferencias en el ECM y el EAM entre las particiones son bajas, esto podrían indicar ligeras particularidades en la distribución de los datos de validación y prueba. Por otro lado, los altos valores de R² y R reflejan que el modelo captura de manera eficiente la estructura de los datos, aunque futuras mejoras podrían enfocarse en reducir aún más el ECM para lograr predicciones aún más precisas en escenarios más complejos.

La distribución espacial del ECM por píxel revela un patrón de error claramente diferenciado en el área de estudio. En la región de la costa del Pacífico y la parte alta montañosa, los valores de ECM oscilan entre 1.0 y 1.25, indicando una menor magnitud de error en estas áreas. En contraste, la región suroriental registra los valores de ECM más altos, con magnitudes que varían entre 1.5 y 2.0, y puntos específicos donde el ECM supera 2.0, reflejando una mayor dificultad en las predicciones en esta zona.

Se observa un gradiente de error progresivo hacia el sureste de la región estudiada. Las áreas del norte y centro-occidental presentan errores más bajos y distribuciones homogéneas, mientras que las regiones sur y suroriental exhiben una variabilidad considerable y una mayor magnitud

en los errores de predicción. Este gradiente destaca la influencia de características geográficas y climáticas locales en el desempeño del modelo.

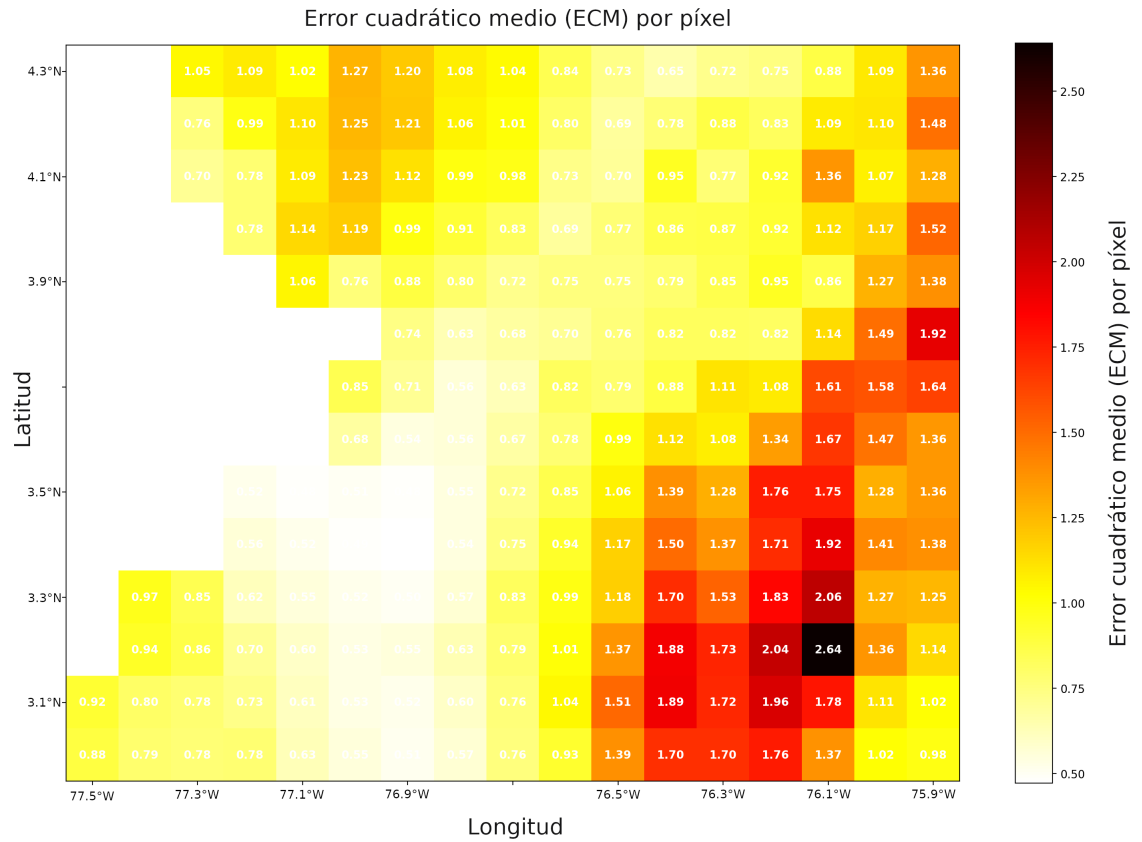


Figura 4.21: Mapa de distribución espacial del error cuadrático medio (ECM) por píxel para las predicciones del modelo ConvLSTM2D para el *downscaling* estadístico de la temperatura con la partición de prueba.

El modelo ConvLSTM2D revela un patrón distintivo en la correlación espacial a lo largo del dominio de estudio. En la región Pacífica, las correlaciones son elevadas, superando 0.85, con picos de entre 0.90 y 0.92 en las áreas montañosas. Por otro lado, en la región occidental se observan correlaciones más bajas, con valores que fluctúan entre 0.70 y 0.75, siendo la zona suroccidental la de menor rendimiento, con valores mínimos cercanos a 0.73.

El gradiente espacial de correlación es evidente, mostrando un aumento desde las áreas costeras del Pacífico hacia las zonas montañosas. En estas últimas, las correlaciones oscilan entre 0.68 y 0.91, reflejando un desempeño más consistente en las áreas de altitud elevada. Además, la distribución espacial de las correlaciones a nivel de píxel también evidencia una marcada

consistencia, lo que sugiere patrones recurrentes en el comportamiento del modelo en respuesta a las características geográficas del área de estudio.

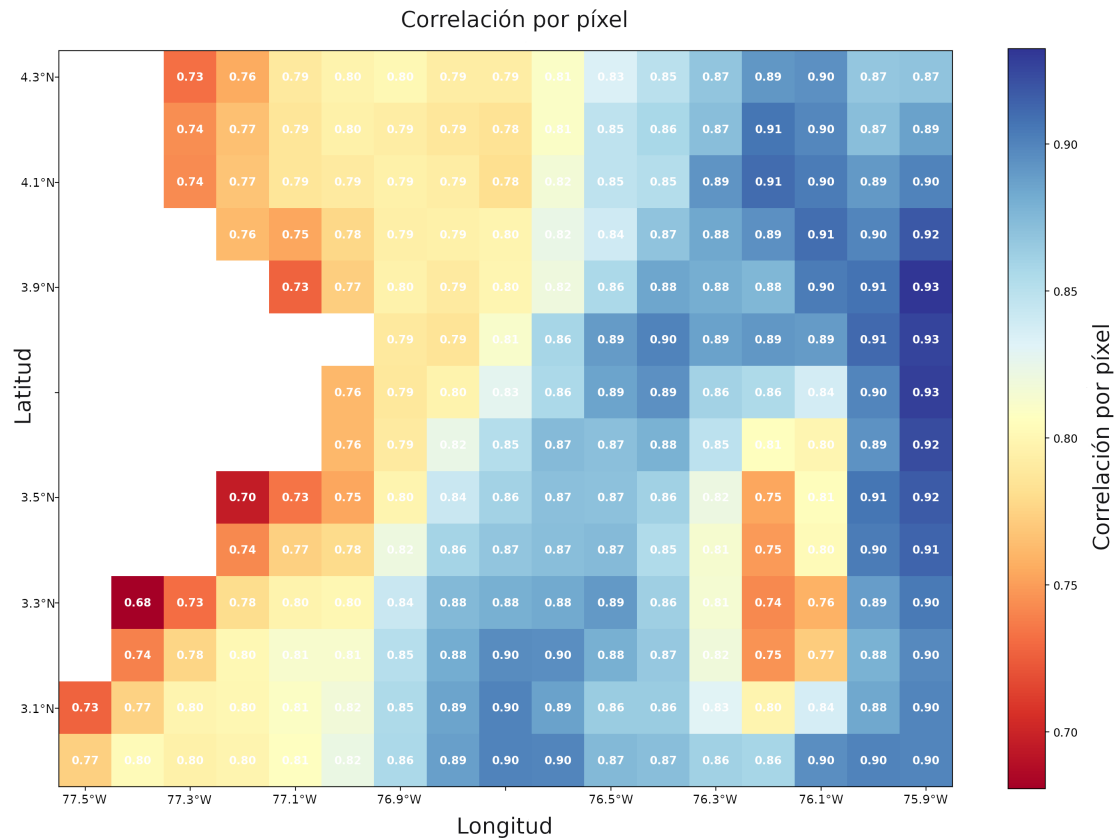


Figura 4.22: Mapa de distribución de correlación de Pearson por píxel para las predicciones del modelo ConvLSTM2D para el *downscaling* estadístico de la temperatura con la partición de prueba.

La variabilidad espacial observada en las correlaciones sugiere una marcada influencia de los patrones topográficos y la proximidad al océano en el desempeño del modelo. Las correlaciones más elevadas en la región oriental indican que el modelo tiene una mayor capacidad para capturar los patrones térmicos en áreas continentales, donde la variabilidad de la temperatura está predominantemente modulada por factores terrestres, como la altitud y la cobertura vegetal. Sin embargo, las correlaciones más bajas detectadas en áreas urbanizadas y en zonas dedicadas a cultivos transitorios presentes en la región, podrían estar relacionadas con una mayor variabilidad térmica local, derivada de las actividades humanas y los cambios en el uso del suelo.

Por otro lado, la disminución de las correlaciones hacia el oeste parece asociarse con la complejidad de los sistemas costeros y la interacción dinámica entre el océano y la atmósfera. Este

gradiente de correlación, que se intensifica desde las zonas costeras hacia el interior continental, sugiere que la arquitectura ConvLSTM2D responde mejor en regiones donde los forzantes atmosféricos presentan mayor homogeneidad. Este hallazgo pone en evidencia la sensibilidad del modelo a las características topográficas y climáticas locales, así como sus limitaciones en áreas donde la interacción entre los forzantes oceánicos y costeros introduce una alta variabilidad en los patrones climáticos.

4.2.4.3. Entrenamiento del modelo ConvLSTM2D para la precipitación

La evaluación del modelo convLSTM2D para el *downscaling* estadístico de precipitación muestran valores de pérdida ECM bajos, de específicamente 3.849e-05 para entrenamiento, 3.786e-05 para validación y 4.012e-05 para prueba. Por su parte, el EAM se mantiene constante en 0.0036 para todas las particiones de datos. Esto nos indica que el modelo es capaz de generalizar el patrón espaciotemporal de la variable precipitación.

Tabla 4.22: Métricas de evaluación de la pérdida (loss) de entrenamiento del modelo ConvLSTM2D para la variable precipitación, aplicado a la partición de los datos.

Evaluación	Loss		Final	
	ECM	EAM	ECM	EAM
Entrenamiento	3.849e-05	0.0036	3.895e-05	0.0036
Validación	3.786e-05	0.0036	3.873e-05	0.0036
Prueba	4.012e-05	0.0036	4.034e-05	0.0036

Los resultados obtenidos sugieren un rendimiento consistente del modelo ConvLSTM2D para la precipitación. La similitud entre los valores de pérdida en los conjuntos de entrenamiento, validación y prueba indica que el modelo no presenta problemas de sobreajuste o subajuste, lo cual es particularmente destacable en aplicaciones de predicción meteorológica. La estabilidad del EAM en 0.0036 a través de todas las particiones refuerza la confiabilidad del modelo, mientras que los valores bajos del ECM sugieren que las predicciones del modelo se mantienen muy cercanas a los valores reales. Esta consistencia en el rendimiento entre las diferentes particiones de datos sugiere que el modelo ha logrado capturar efectivamente los patrones subyacentes en los datos de precipitación.

El entrenamiento de la red ConvLSTM2D para la precipitación tiene una convergencia satisfactoria, evidenciada por la disminución significativa de las métricas de error del valor de la pérdida en las primeras 5 épocas. La Raíz del Error Cuadrático Medio (RECM) muestra una reducción drástica inicial desde aproximadamente 1200 hasta estabilizarse por debajo de 50 después de la época 10, tanto para el conjunto de entrenamiento como de validación. Paralelamente, el

Error Absoluto Medio (EAM) sigue un patrón similar, convergiendo desde 15 hasta valores cercanos a 1 después de 25 épocas.

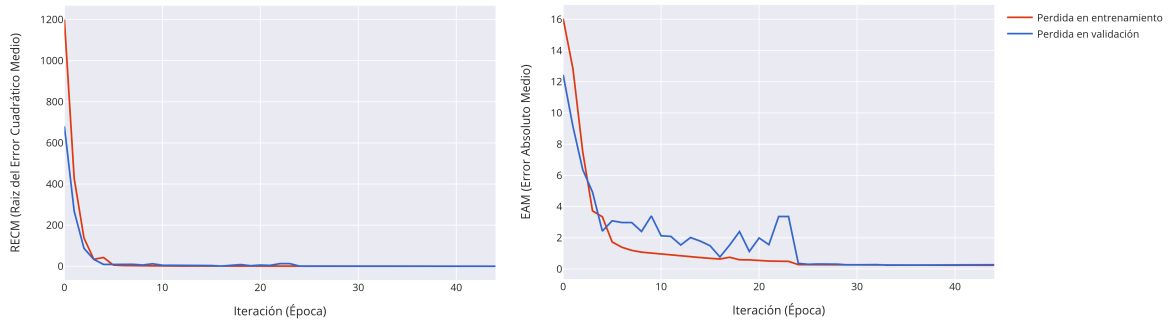


Figura 4.23: Evolución de las curvas de error durante el entrenamiento de la red ConvLSTM2D para la predicción de la precipitación. Métricas de evaluación RMSE a la izquierda y EAM a la derecha.

La similitud en las curvas de pérdida entre entrenamiento y validación sugiere que el modelo no presenta sobreajuste significativo. Sin embargo, se observa una ligera fluctuación en la validación del EAM entre las épocas 05 - 25, indicando cierta inestabilidad en la generalización del modelo durante esta fase. La estabilización final de ambas métricas después de la época 30 sugiere que el modelo ha alcanzado un punto óptimo de convergencia, donde incrementar el número de épocas no aportó mejoras significativas en el rendimiento predictivo.

4.2.4.4. Evaluación del modelo ConvLSTM2D para la precipitación

Para la evaluación del modelo para la predicción de la precipitación por hora se añadió la métrica de correlación de Spearman (S) debido a que esta representa mejor la naturaleza de los datos de precipitación. Los valores obtenidos muestran un ECM cercano a 0.066 para todas las particiones, un RECM aproximado de 0.256, y un EAM alrededor de 0.15. La correlación de Spearman (S) se mantiene cercano al -0.198, mientras que el coeficiente de correlación de Pearson (R) oscila alrededor de -0.15. El R^2 presenta valores aproximadamente de 0.024 para entrenamiento, 0.024 para validación y 0.020 para prueba.

Tabla 4.23: Métricas de evaluación de desempeño del modelo de ConvLSTM2D con cada partición del set de datos.

Métricas de evaluación para la Precipitación						
Partición	ECM	RECM	EAM	S	R	R ²
Entrenamiento	0.066	0.257	0.150	-0.198	-0.155	0.024
Validación	0.065	0.255	0.149	-0.199	-0.156	0.024
Prueba	0.070	0.264	0.150	-0.199	-0.144	0.020

Los resultados revelan aspectos interesantes sobre el comportamiento del modelo con los datos de precipitación. La consistencia en los valores de ECM, RECM y EAM entre las diferentes particiones sugiere que el modelo mantiene un rendimiento estable, lo cual es deseable en términos de generalización. Sin embargo, los bajos valores de R² indican que el modelo tiene limitaciones significativas en su capacidad predictiva, sugiriendo que las predicciones del modelo son menos precisas que simplemente usar la media de los datos observados. El coeficiente de correlación de Pearson relativamente bajo (R: -0.15 aproximadamente) también indica una relación débil entre las predicciones y los valores observados. Esto podría atribuirse a la naturaleza altamente variable y compleja de la precipitación, que es notoriamente difícil de predecir con precisión, posiblemente característico para esta región del globo. El coeficiente de correlación de Spearman cercano al -0.198 sugiere una mejora muy modesta sobre un modelo de referencia. Estos resultados sugieren que, aunque el modelo muestra consistencia en sus predicciones, hay espacio considerable para mejoras en su capacidad predictiva.

La distribución espacial del ECM revela patrones que sugieren una variabilidad significativa en la precisión de las predicciones según la ubicación geográfica. La concentración de errores más altos en la franja diagonal podría estar relacionada con características topográficas específicas de la región y con patrones atmosféricos particulares que hacen más desafiante la predicción de precipitación en estas áreas. El gradiente de error observado, que disminuye hacia las regiones orientales y continentales del mapa, sugiere que el modelo tiene un mejor rendimiento en estas zonas, posiblemente debido a patrones de precipitación más estables o predecibles. Esta heterogeneidad espacial en el rendimiento del modelo proporciona información valiosa para futuras mejoras al modelar las precipitaciones con redes neuronales, indicando que podrían necesitarse ajustes específicos para las regiones con mayor error o la incorporación de variables adicionales específicas que ayuden a capturar mejor la variabilidad de la precipitación en las zonas más problemáticas.

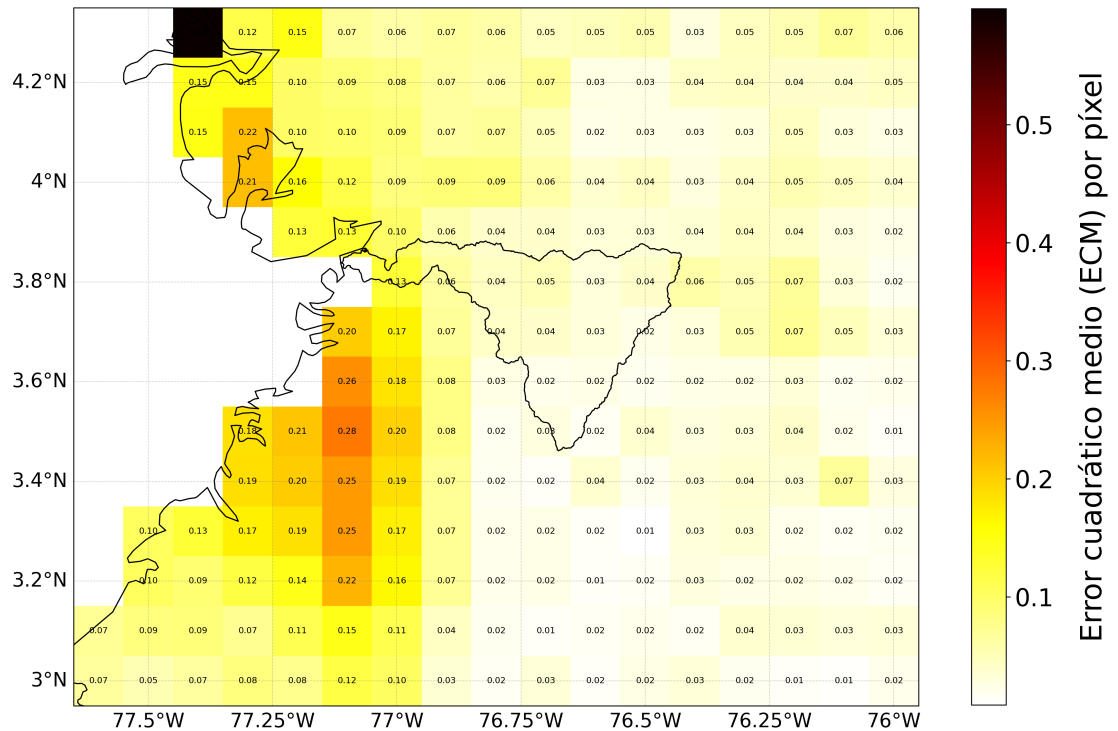


Figura 4.24: Mapa de la distribución del error cuadrático medio de la precipitación con el set de prueba.

El mapa de distribución espacial del coeficiente de correlación de Spearman muestra un patrón heterogéneo de correlaciones que varían entre valores positivos y negativos, con una escala que va desde aproximadamente -0.06 hasta 0.04. Esta distribución espacial de la correlación presenta una estructura irregular, donde se alternan zonas de correlación positiva y negativa, con valores predominantemente cercanos a cero. Se identifican concentraciones específicas de correlaciones positivas ligeramente más altas, especialmente en la zona costera, mientras que las correlaciones negativas más acentuadas se distribuyen de forma dispersa, destacándose en píxeles cercanos a las áreas urbanas.

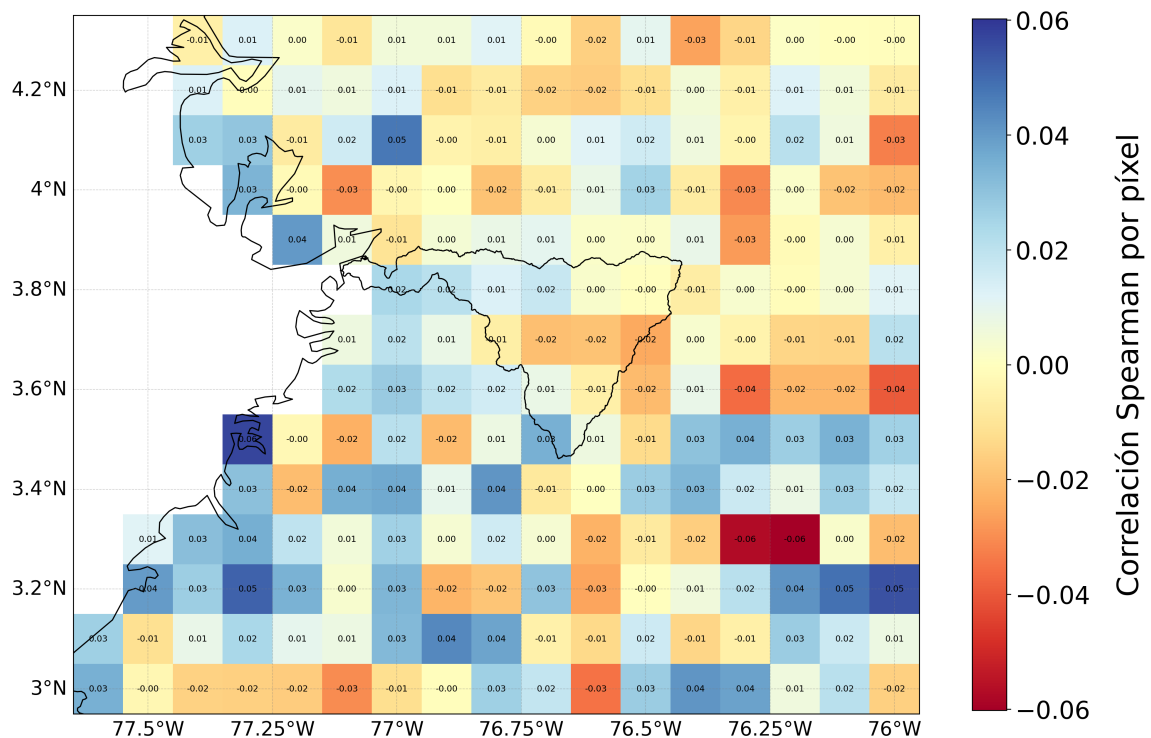


Figura 4.25: Mapa de la distribución de correlación de Pearson para la precipitación con el set de prueba.

Los resultados de la correlación de Spearman reflejan una relación compleja entre las predicciones del modelo y los valores observados a lo largo del dominio espacial. El dominio de valores cercanos a cero indica que el modelo enfrenta dificultades para capturar adecuadamente la variabilidad de la precipitación en la mayoría del área de estudio. Las regiones con correlaciones positivas más destacadas, aunque limitadas, señalan áreas donde el modelo tiene un mejor desempeño al replicar la dirección de cambio de la precipitación, aunque estas correlaciones aún son relativamente débiles. Por otro lado, la presencia de correlaciones negativas en ciertas zonas resulta preocupante, ya que sugiere que en estas áreas las predicciones del modelo tienden a ir en dirección opuesta a los valores observados, reflejando inconsistencias significativas en el rendimiento.

4.2.4.5. Comparación de la modelación ConvLSTM2D de los datos de temperatura y precipitación

Para tener mejor representación de los datos modelados para el *downscaling* estadístico por la red ConvLSTM2D sobre los datos de temperatura y precipitación del GCM, se presentan las gráficas para la comparación temporal entre datos de ERA5-Land, las observaciones de las estaciones y los resultados modelados para la estación "La Cumbre", "Farallones" y "Unipacífico". Las figuras

4.26, 4.27, 4.28 ilustran el comportamiento de la temperatura durante un corto período de aproximadamente una semana para mejor detalle de los resultados. Por otra parte las figuras 4.29, 4.30 4.31, muestran el patrón mensual de precipitación durante un período de aproximadamente tres años.

Patrón de temperatura estación "La Cumbre"

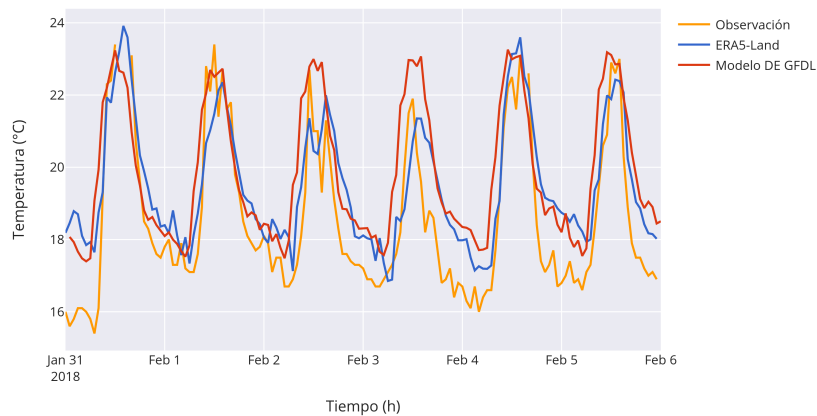


Figura 4.26: Patrón de la temperatura entre las observaciones de la estación "La cumbre", la estimación de ERA5-Land y las predicciones de *Downscaling* estadístico desarrollado.

Patrón de temperatura estación "Farallones"

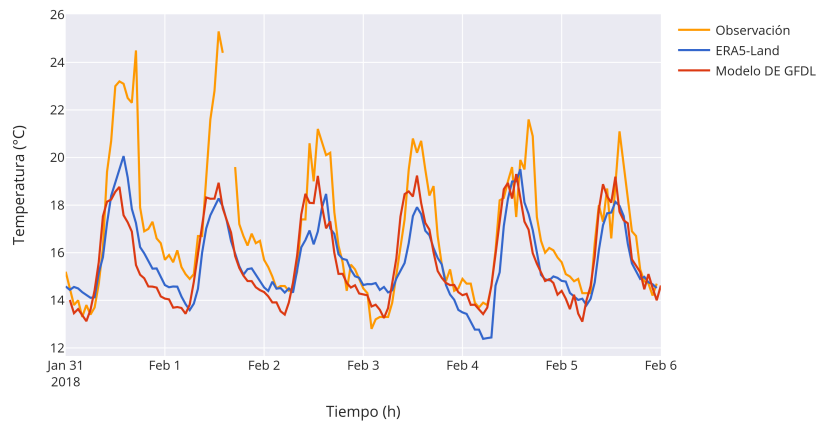


Figura 4.27: Patrón de la temperatura entre las observaciones de la estación "Farallones", la estimación de ERA5-Land y las predicciones de *Downscaling* estadístico desarrollado.

Patrón de temperatura estación "Unipacífico"

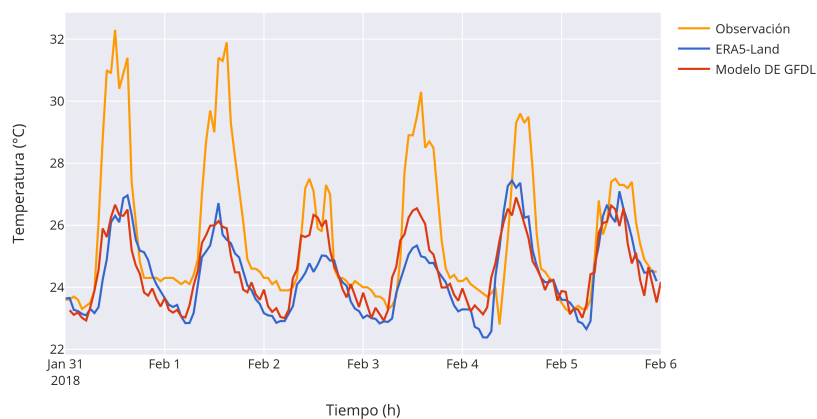


Figura 4.28: Patrón de la temperatura entre las observaciones de la estación "Unipacífico", la estimación de ERA5-Land y las predicciones de *Downscaling* estadístico desarrollado.

Patrón mensual de precipitación estación "La Cumbre"

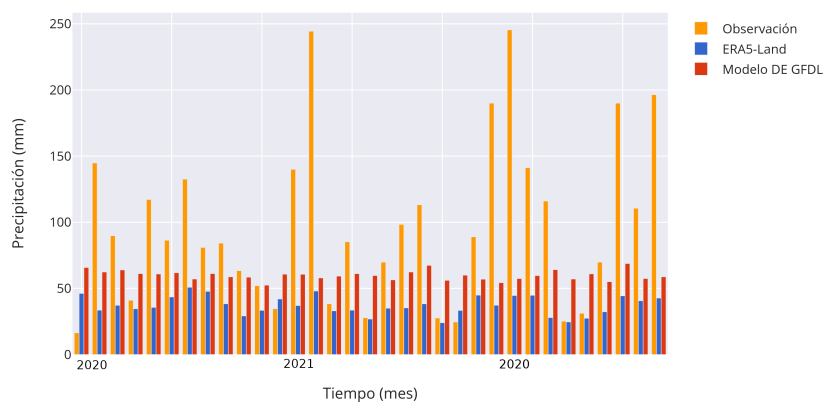


Figura 4.29: Patrón de la precipitación acumulada por mes entre las observaciones de la estación "La Cumbre", la estimación de ERA5-Land y las predicciones de *Downscaling* estadístico desarrollado.

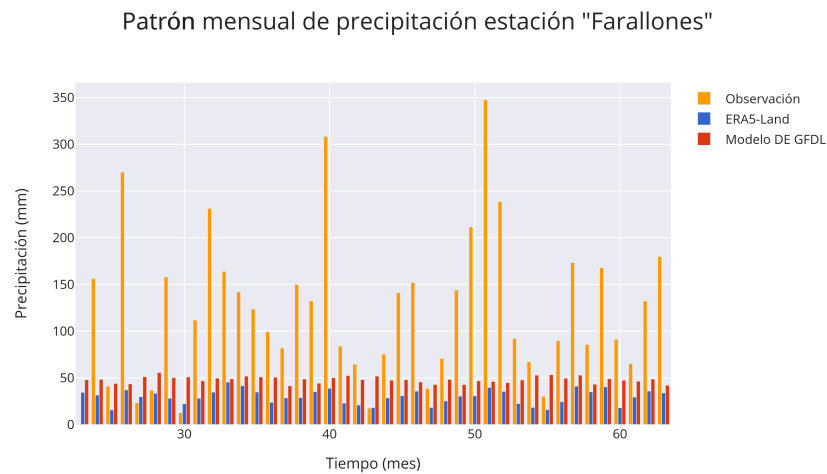


Figura 4.30: Patrón de la precipitación acumulada por mes entre las observaciones de la estación "Farallones", la estimación de ERA5-Land y las predicciones de *Downscaling* estadístico desarrollado.

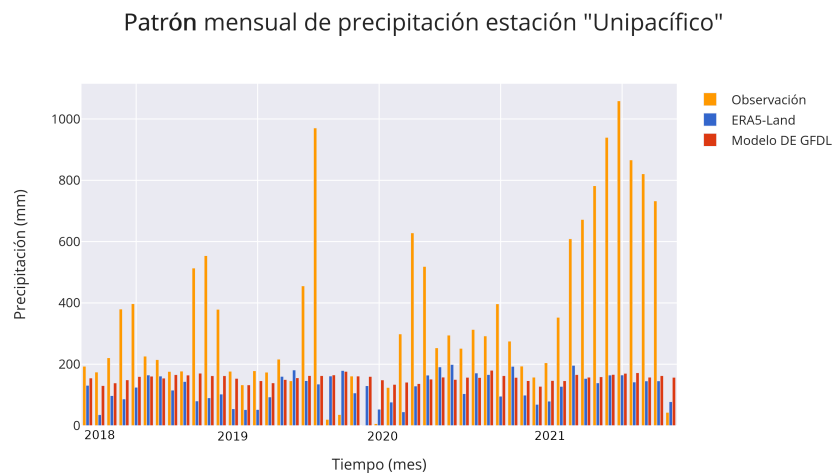


Figura 4.31: Patrón de la precipitación acumulada por mes entre las observaciones de la estación "Unipacífico", la estimación de ERA5-Land y las predicciones de *Downscaling* estadístico desarrollado.

La comparación entre los datos observados y modelados por la red ConvLSTM2D revela aspectos importantes sobre el rendimiento del modelo para las variables temperatura y precipitación. En el caso de la precipitación, tanto ERA5-Land como el modelo GFDL muestran

una clara limitación para capturar la magnitud de los eventos de precipitación intensa, lo cual es un desafío común en la modelación de esta variable debido a su alta variabilidad espacial y temporal.

En contraste, para la temperatura, el modelo demuestra una capacidad significativamente mejor para reproducir los patrones diarios observados a una resolución temporal horaria, capturando adecuadamente tanto la fase como la amplitud del ciclo diurno, aunque con ligeras diferencias en la representación de los valores máximos y mínimos. Esta discrepancia en el rendimiento entre variables meteorológicas sugiere que mientras los modelos son robustos para variables más estables como la temperatura, aún requieren mejoras significativas para la predicción precisa de eventos de precipitación, especialmente durante eventos extremos.

4.3. Validación del modelo ConvLSTM2D para el *Downscaling* estadístico

Para facilitar una interpretación más clara y estructurada de los resultados obtenidos con la red ConvLSTM2D para cada variable analizada, las tablas 4.24 y 4.25 presentan los mejores resultados de evaluación de los modelos: RL, RF y RVS. Este planteamiento permite comparar la precisión del modelo ConvLSTM2D en la generación de nuevos valores de temperatura y precipitación frente a las metodologías estadísticas convencionales utilizadas en el *downscaling* estadístico, destacando sus fortalezas y limitaciones relativas.

En la Tabla 4.24, los resultados evidencian que el modelo ConvLSTM2D supera consistentemente a los modelos estadísticos tradicionales (RL, RF y RVS) en todas las métricas de error analizadas para la variable temperatura. La arquitectura ConvLSTM2D utilizada, obtiene el menor ECM de 1.020 y EMA de 0.766, junto con un RECM significativamente bajo (1.001). Además, el modelo logra el valor más alto de correlación con $R = 0.976$ y un coeficiente de determinación de $R^2 = 0.954$, lo que indica una excelente capacidad para representar tanto la magnitud como la variabilidad temporal de los datos de temperatura.

Tabla 4.24: Mejores resultados de los modelos evaluados para el *downscaling* estadístico para la variable temperatura.

EVALUACIONES DE LA TEMPERATURA						
MODELO	ECM	RECM	EMA	EPAM	R	R2
RL	3.426	1.850	1.423	0.49 %	0.921	0.848
RF	1.694	1.301	1.001	0.34 %	0.962	0.924
RVS	1.656	1.287	0.958	0.43 %	0.962	0.925
ConvLSTM2D	1.020	1.001	0.766	0.26 %	0.976	0.954

En comparación, los modelos estadísticos muestran un desempeño inferior, con ECM más altos (por ejemplo, RL: 3.426) y correlaciones más bajas (R para RL: 0.921). Este resultado resalta la ventaja de las arquitecturas basadas en redes neuronales convolucionales recurrentes, como ConvLSTM2D, para capturar patrones temporales y espaciales complejos en variables con menor variabilidad estocástica como la temperatura. Esto demuestra que, en el contexto de la modelación de *downscaling* estadístico para la cuenca del río Dagua, ConvLSTM2D es una herramienta más robusta para predecir la temperatura frente a los enfoques tradicionales, consolidando su aplicabilidad en regiones con condiciones topográficas y climáticas complejas.

Por otro lado la Tabla 4.25 muestra los valores de las métricas de evaluación para la modelación de la precipitación utilizadas (modelos RL, RF, RVS y ConvLSTM2D). Los resultados destacan la notable reducción del ECM y el EMA alcanzada por el modelo ConvLSTM2D con un ECM: 0.070 y un EMA: 0.1504, en comparación con los modelos tradicionales, lo que indica una mejor precisión en la predicción de los valores de precipitación. Asimismo, el modelo ConvLSTM2D obtiene un RECM de 0.264, significativamente menor al de los modelos RL, RF y RVS, que se mantienen alrededor de 0.619.

Tabla 4.25: Mejores resultados de los modelos evaluados para el *downscaling* estadístico para la variable precipitación.

EVALUACIONES DE LA PRECIPITACIÓN						
MODELO	ECM	RECM	EMA	S	R	R2
RL	0.383	0.619	0.3421	0.229	0.196	0.038
RF	0.384	0.619	0.3436	0.298	0.292	0.085
RVS	0.388	0.611	0.3343	0.215	0.199	0.039
ConvLSTM2D	0.070	0.264	0.1504	-0.199	-0.144	0.020

Sin embargo, es importante señalar que, a pesar de su capacidad para reducir los errores absolutos, ConvLSTM2D muestra correlaciones negativas (R: -0.144) y un R² bajo (0.020). Esto sugiere que, aunque el modelo captura adecuadamente la magnitud de la precipitación, presenta limitaciones para representar la variabilidad temporal y espacial de esta variable.

Por otro lado, el diagrama de Taylor se emplea como una herramienta gráfica para ilustrar el desempeño de los diferentes modelos estadísticos o simulaciones, frente a un conjunto de observaciones de referencia. Este diagrama sintetiza tres métricas clave en un único gráfico: la correlación, la desviación estándar y el error cuadrático medio, proporcionando una visión general del ajuste entre los modelos y los datos medidos por las estaciones.

Mediante el uso diagrama de Taylor, se evaluó el rendimiento del modelo ConvLSTM2D en la predicción de nuevos valores, comparándolo con los modelos RL, RF y RVS. El análisis se realizó

utilizando de las estaciones meteorológicas "Farallones", "La Cumbre" y "Unipacífico". El diagrama de Taylor demuestra ser particularmente útil para validar modelos climáticos y meteorológicos, como los empleados en el *downscaling* estadístico, al ofrecer una representación clara y concisa del desempeño relativo de cada modelo.

4.3.1. Validación estación "La Cumbre"

En la estación "La Cumbre", los modelos evaluados presentan correlaciones que varían entre 0.77 y 0.83, destacándose la arquitectura ConvLSTM2D como el modelo con la mayor correlación de 0.83. El modelo RF también muestra un desempeño competitivo, con una correlación de 0.82. En cuanto a la desviación estándar, los valores estimados oscilan entre 1.68 °C y 1.83 °C, siendo ConvLSTM2D el modelo con la desviación más próxima a los datos observados. Sin embargo, al evaluar la Raíz del Error Cuadrático Medio (RECM), el modelo RF presenta un valor menor (1.69 °C) en comparación con ConvLSTM2D (1.93 °C), lo que indica ligeras diferencias en el ajuste entre los modelos.

Tabla 4.26: Métricas de validación de los modelos de Downscaling estadístico para la variable temperatura sobre la estación "La Cumbre".

Temperatura estación "La Cumbre"				
Modelo	C. Pearson	DS Pred	DS Real	RECM
Regresión Lineal	0.7775	1.2927	2.1811	2.2712
Random Forest	0.8262	1.6819	2.1811	1.6964
Regresión VS	0.7697	1.3053	2.1811	1.8570
ConvLSTM2D	0.8350	1.8359	2.2363	1.9357

En el caso de la precipitación para la estación "La Cumbre", el modelo RF alcanzó la correlación más alta con un coeficiente $R = 0.0292$, mientras que el modelo ConvLSTM2D mostró la correlación más baja $R = 0.0035$. Sin embargo, es importante destacar que el modelo ConvLSTM2D obtuvo la mejor aproximación en términos de desviación estándar de 0.09 mm/h, así como una RECM de 0.89mm/h, lo que indica un mejor ajuste a los valores observados en esta métrica específica.

Tabla 4.27: Métricas de validación de los modelos de DE para la variable precipitación sobre la estación "La Cumbre".

Precipitación estación "La Cumbre"				
Modelo	C. Spearman	DS Pred	DS Real	RECM
Regresión Lineal	-0.0048	0.1043	1.6558	1.6990
Random Forest	0.0292	0.3218	1.6558	1.7249
Regresión VS	0.0084	0.2034	1.6558	1.6119
ConvLSTM2D	0.0035	0.0933	0.9412	0.8952

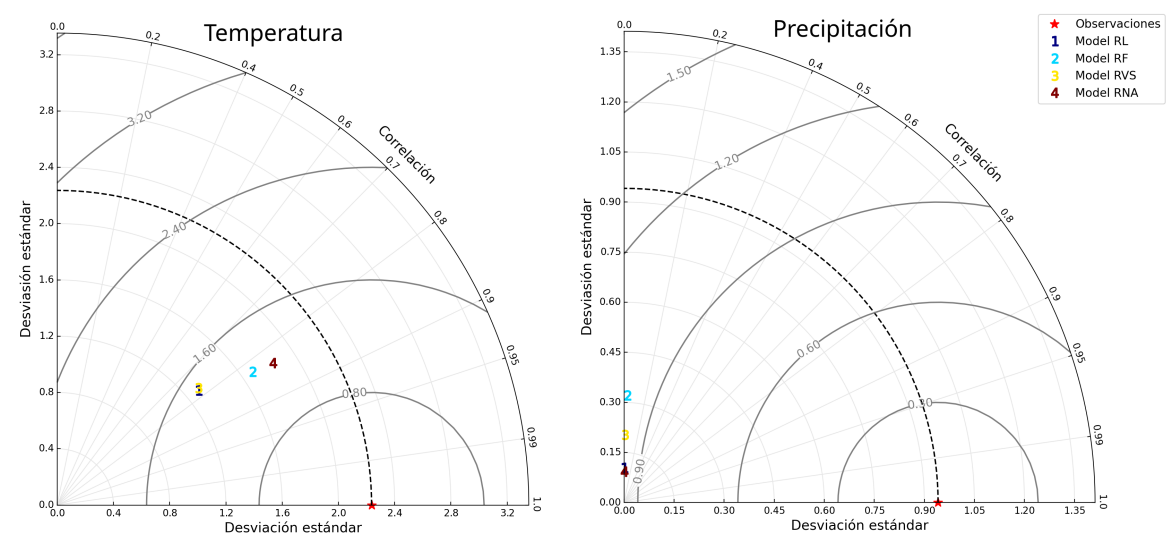


Figura 4.32: Diagrama de Taylor de los modelos de temperatura, evaluado con los datos medidos por la estación meteorológica "La Cumbre".

4.3.2. Validación estación "Farallones"

Para la estación "Farallones" la correlación es muy similar en los cuatro modelos, donde los valores más destacados son los de RF con 0.74 y ConvLSTM2D con 0.78. Del mismo modo las desviaciones estándar tiene una diferencia de 1.67 para RF y 0.79 para el modelo ConvLSTM2D. Por otro lado el ECM es mejor representado para esta estación por el modelo ConvLSTM2D de 2.3 °C.

Tabla 4.28: Métricas de validación de los modelo de DE para la variable temperatura sobre la estación "Farallones".

Temperatura estación "Farallones"				
Modelo	C. Pearson	DS Pred	DS Real	RECM
Regresión Lineal	0.7497	1.3025	3.0118	3.8797
Random Forest	0.7815	1.6760	3.0118	3.2439
Regresión VS	0.7304	1.3043	3.0118	3.3805
ConvLSTM2D	0.7894	1.7960	3.1241	2.3798

La precipitación en esta estación, reflejan resultados de correlación que oscilan entre 0.0052 para el modelo ConvLSTM2D y 0.15 para el modelo de RL, evidenciando el desempeño más limitado para el modelamiento de esta variable a una resolución horaria. El modelo ConvLSMT2D presenta una menor desviación estándar de 0.07 y un ECM de 0.86 aproximándose mejor a los datos de referencia, mientras que el modelo RF muestra un ajuste ligeramente inferior.

Tabla 4.29: Métricas de validación de los modelos de DE para la variable precipitación sobre la estación "Farallones".

Precipitación estación "Farallones"				
Modelo	C. Spearman	DS Pred	DS Real	RECM
Regresión Lineal	0.0153	0.0999	1.6390	1.6716
Random Forest	0.0146	0.3217	1.6390	1.6992
Regresión VS	0.0101	0.2171	1.6390	1.6801
ConvLSTM2D	0.0052	0.0700	0.9238	0.8607

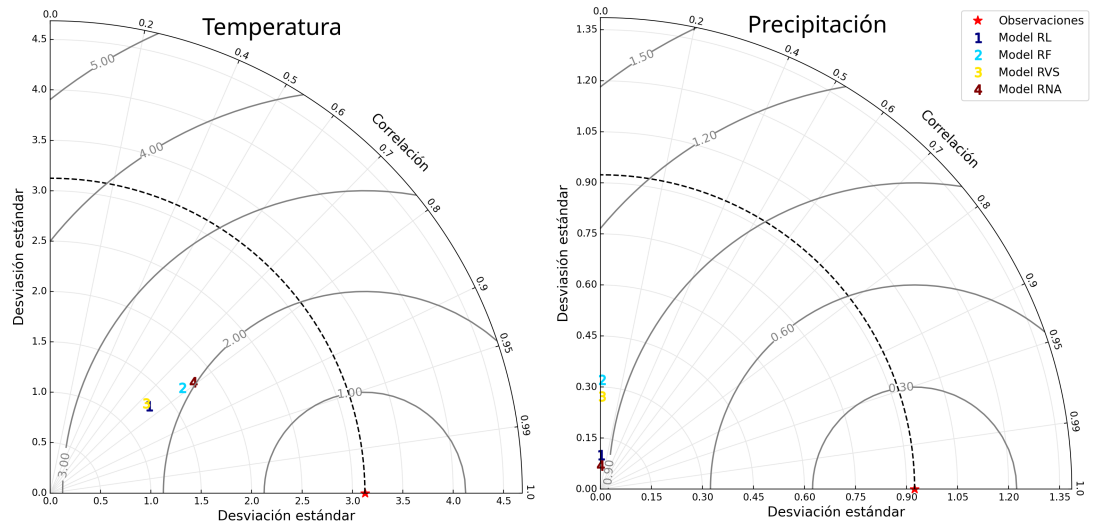


Figura 4.33: Diagrama de Taylor de los modelos de temperatura, evaluado con los datos medidos por la estación meteorológica "Farallones".

4.3.3. Validación estación "Unipacífico"

Del mismo modo en la estación "Unipacífico", el modelo ConvLSTM2D se destaca como el de mejor desempeño, con una correlación de $R = 0.79$, seguido del modelo RF, que obtuvo un valor de $R = 0.78$. En cuanto a la desviación estándar, las predicciones del modelo ConvLSTM2D (1.16) muestran una mayor proximidad a los valores observados en comparación con RF (1.59). Además, la RECM se encuentra en un rango de 3.21 a 2.06, siendo ConvLSTM2D el modelo con el menor error (2.06). Esto indica que las predicciones del modelo ConvLSTM2D para la temperatura son más precisas y confiables en esta estación.

Tabla 4.30: Métricas de validación de los modelos de DE para la variable temperatura sobre la estación "Unipacífico".

Temperatura estación "Unipacífico"				
Modelo	C. Pearson	DS Pred	DS Real	RECM
Regresión Lineal	0.7107	1.2864	2.4662	5.7893
Random Forest	0.7805	1.5967	2.4662	6.2622
Regresión VS	0.6834	1.3032	2.4662	6.4067
ConvLSTM2D	0.7925	1.1656	2.5640	2.1298

El resultado de los valores de correlación modelada en la zona baja de la cuenca refleja las mismas dificultades que en las otras estaciones. El modelo RVS mostró el valor más bajo de

correlación $R = 0.0052$, mientras que el modelo de RL alcanzó la correlación más alta, aunque aún limitada $R = 0.0340$. El modelo ConvLSTM2D logró una correlación intermedia con un valor de $R = 0.0194$, junto con una RECM de 2.06 mm/h, que, aunque relativamente elevado, es inferior al de los otros modelos. Esto sugiere que ConvLSTM2D presenta ventajas relativas en comparación con los métodos tradicionales, aunque persisten desafíos significativos para modelar la precipitación en toda la región.

Tabla 4.31: Métricas de validación de los modelos de DE para la variable precipitación sobre la estación "Unipacífico".

Precipitación estación "Unipacífico"				
Modelo	C. Spearman	DS Pred	DS Real	ECM
Regresión Lineal	0.0340	0.0863	3.2116	3.2123
Random Forest	0.0062	0.3021	3.2116	3.2230
Regresión VS	0.0052	0.4981	3.2116	3.8187
ConvLSTM2D	0.0194	0.1060	2.0463	2.0670

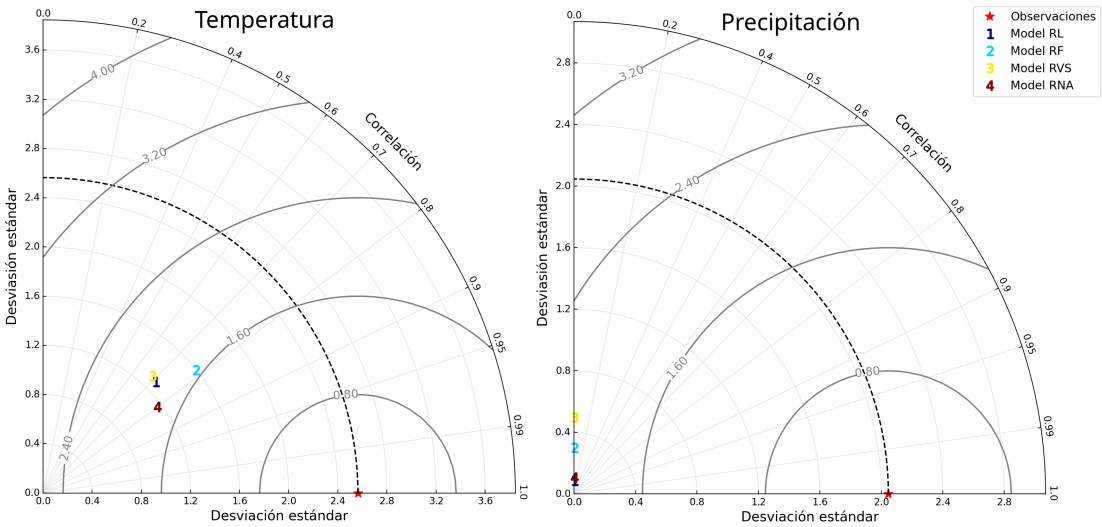


Figura 4.34: Diagrama de Taylor de los modelos de temperatura, evaluado con los datos medidos por la estación meteorológica "Unipacífico".

El análisis del diagrama de Taylor destaca la capacidad del modelo ConvLSTM2D para competir y en algunos casos superar a los modelos tradicionales para el *downscaling* estadístico, especialmente en la predicción de la temperatura, en relación con los valores de referencia medidos por las estaciones, a pesar de estar generando nuevos valores de temperatura y precipitación, fue

igual o superior ya que el modelo RF en métricas de correlación y error es mínima. Si bien las diferencias entre ConvLSTM2D y RF son mínimas en términos de correlación y error, la arquitectura ConvLSTM2D logra destacarse al pronosticar nuevos valores para reducir la resolución temporal, manteniendo una representación precisa de los patrones espacio-temporales.

En el caso de la temperatura, los resultados muestran una tendencia generalizada de los modelos a subestimar la variabilidad de los datos reales de las estaciones, evidenciada por desviaciones estándar inferiores en comparación con las observaciones. Esto es particularmente evidente en estaciones como "Farallones" a 2275 m.s.n.m., donde las condiciones climáticas son más complejas debido a la altitud. Sin embargo, los modelos mantienen correlaciones moderadas y bajos ECM en todas las estaciones, lo que indica que logran capturar razonablemente bien las dinámicas térmicas, aunque con mayor precisión en estaciones de menor altitud, como "Unipacífico" a 16 m.s.n.m., donde las condiciones térmicas son más homogéneas.

Por otro lado, la precipitación presenta desafíos significativos en su modelamiento a escala horaria, donde todos los modelos subestiman marcadamente la variabilidad observada, mostrando correlaciones bajas y una limitada capacidad para reproducir los valores reales, lo que es más evidente en estaciones de la zona alta e intermedia como "La Cumbre" a 1613 m.s.n.m. En este contexto, el modelo RF se posiciona como el de mejor desempeño, aunque la ventaja sobre ConvLSTM2D es modesta. Estos resultados reflejan la complejidad de los procesos asociados con la precipitación, caracterizados por alta variabilidad espacial y temporal y una fuerte no linealidad, particularmente en regiones con alta heterogeneidad topográfica marcada como el Pacífico colombiano.

4.3.4. Análisis de proyecciones climáticas mediante el *downscaling* estadístico con redes ConvLSTM2D

Se analizan las proyecciones climáticas de las variables temperatura y precipitación generadas mediante el *downscaling* estadístico basado en redes neuronales ConvLSTM2D para el período 2024-2035, bajo los escenarios SSP1-1.9, SSP1-2.6 y SSP2-4.5. El análisis incluye estadísticas descriptivas y visualizaciones gráficas que permiten identificar los patrones modelados en las estaciones seleccionadas: "La Cumbre", "Farallones" y "Unipacífico". Estas herramientas facilitan el reconocimiento del comportamiento de las variables en diferentes subregiones de la cuenca, considerando la variabilidad espacial y temporal en el caso de estudio.

4.3.4.1. Proyecciones climáticas para el escenario SSP2-4.5

Bajo el escenario SSP2-4.5 de las proyecciones para el período 2024-2035, el análisis revela patrones distintivos en los regímenes de precipitación mensual en las estaciones "Unipacífico",

"Farallones" y "La Cumbre". Los resultados muestran una marcada variabilidad espacial y temporal en las precipitaciones, caracterizada mediante visualizaciones temporales y análisis estadísticos descriptivos.

Tabla 4.32: Estadísticos Descriptivos de Precipitación Mensual multianual (mm/mes) por Estación bajo Escenario SSP2-4.5 proyectos desde el año 2024 hasta el año 2035.

Descripción estadística escenario SSP2-4.5								
Estación	mean	std	min	25 %	50 %	75 %	max	count
Farallones	176.18	10.31	135.73	170.74	178.73	182.89	197.16	132
La Cumbre	189.07	7.20	165.04	185.37	190.29	194.26	208.78	132
Unipacífico	293.07	13.78	256.01	285.27	294.71	303.70	318.88	132

Los valores modelados indican que la estación "Unipacífico", ubicada en la zona baja de la cuenca, presenta consistentemente las precipitaciones más altas, con una media de 293.07 mm/mes, lo que coincide con los patrones identificados en estudios previos y el análisis exploratorio realizado. En contraste, las estaciones "La Cumbre" (media de 189.07 mm/mes) y "Farallones" (media de 176.18 mm/mes), situadas en la parte alta de la cuenca, registran valores medios más bajos, reflejando la influencia de la topografía en los regímenes de precipitación.

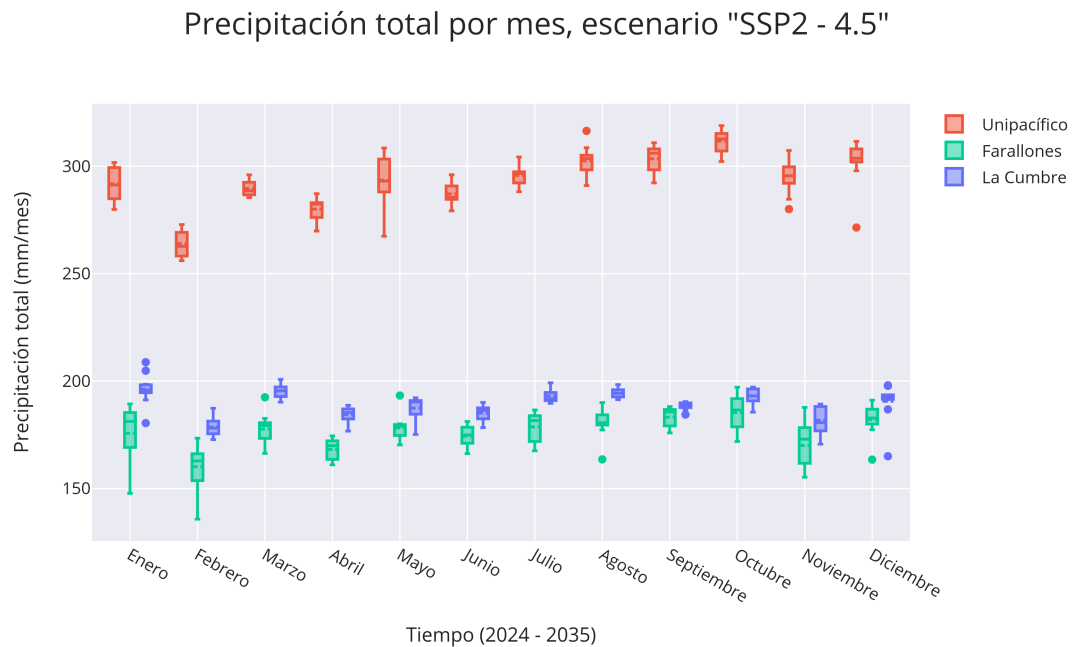


Figura 4.35: Distribución Temporal de la Precipitación Mensual Multianual Proyectada (2024-2035) bajo Escenario SSP2-4.5.

La estación "Unipacífico" también muestra mayor variabilidad, con una desviación estándar de 13.78 mm, lo que sugiere una mayor susceptibilidad a eventos extremos de precipitación. Los datos temporales modelados muestran levemente el patrón bimodal, donde los períodos de mayor precipitación ocurren entre agosto y octubre, mientras que los valores mínimos se registran en febrero. Este comportamiento es consistente con los patrones climáticos históricos observados en la región previamente, lo que refuerza la capacidad del modelo para capturar las dinámicas climáticas locales.

En general, los resultados evidencian que, aunque la arquitectura ConvLSTM2D ofrece ventajas claras en la generación de nueva información con mayor resolución espacial y temporal, estas son más pronunciadas para la temperatura que para la precipitación. Esto sugiere que el modelo es más efectivo en captar patrones más predecibles y estables, como los de la temperatura, mientras enfrenta mayores dificultades en variables más dinámicas como la precipitación.

Los patrones espaciales identificados en los resultados ofrecen evidencia de la variabilidad climática regional. En áreas como la región costera del Pacífico, el modelo evidencia un mejor desempeño, lo que subraya la importancia de la proximidad al océano y los forzantes locales. En contraste, en regiones urbanas y áreas agrícolas transitorias, donde las correlaciones fueron más bajas, se observa cómo los cambios abruptos en el uso del suelo y la variabilidad térmica

local desafían la capacidad del modelo. Estos hallazgos destacan la necesidad de regionalizar las respuestas: mientras que el modelo es eficaz para representar las condiciones climáticas en zonas montañosas y costeras, su desempeño podría mejorarse en áreas urbanas mediante el ajuste de parámetros y la inclusión de variables adicionales como el tipo de cobertura terrestre o tipo de índice de vegetación.

Del mismo modo al añadir las componentes principales más significativas mejoran la precisión de cada uno de los modelos. Esto resalta la importancia de una adecuada selección de características, ya que las componentes principales que explican menos varianza no contribuyen de manera significativa a la mejora de las predicciones. Este resultado es consistente con la teoría estadística y estudios previos (Schoof y Pryor, 2001), (Tadić y cols., 2019) que enfatizan el impacto de la dimensionalidad reducida en modelos de predicción climática.

Las métricas de validación utilizadas, tales como la correlación, el coeficiente de determinación (R^2), el RMSE y MAE, revelaron que los modelos basados en aprendizaje profundo, como ConvLSTM2D, superaron consistentemente a los métodos tradicionales, como la regresión lineal múltiple, Random Forest y la regresión de vectores de soporte. En particular, el aumento en la correlación entre las predicciones y los valores observados, así como la reducción del error, demuestran la capacidad superior de las redes neuronales profundas para capturar relaciones no lineales complejas en los datos, un aspecto en el que los métodos tradicionales mostraron limitaciones, ya que al aumentar datos o variables el modelo no presentaba mejoras importantes.

Por ejemplo, aunque el modelo de regresión lineal múltiple estableció una línea base sólida y presentó un bajo consumo computacional, sus capacidades predictivas se vieron limitadas por su naturaleza lineal, lo que dificultó la captura de interacciones más complejas entre las variables climáticas. Este hallazgo coincide con estudios previos que muestran que los modelos lineales son eficaces como una primera aproximación, pero no son adecuados para abordar la complejidad intrínseca de los procesos atmosféricos no lineales (Ballarin y cols., 2023). Random Forest mostró una mejora considerable respecto a la RL, con mejores métricas de rendimiento, gracias a su capacidad para manejar relaciones no lineales y su robustez frente a datos ruidosos (H. Zhang y cols., 2021). Sin embargo, la elevada demanda de memoria RAM durante su entrenamiento con grandes volúmenes de datos limitó el número de estimadores que se podían utilizar, lo que podría haber afectado su capacidad para generalizar en todo el conjunto de datos. A pesar de ello, Random Forest demostró ser un modelo robusto y de alto rendimiento, aunque computacionalmente más exigente que la RL.

La Regresión de Vectores de Soporte (SVR) también presenta limitaciones cuando se aplica un gran volumen de datos, lo que afectó su escalabilidad. No obstante, en los experimentos realizados con un subconjunto anual de los datos, SVR demostró un rendimiento similar al de la RLM, con resultados de validación superiores pero sin reconocer el patrón y tendencia de las variables.

Esto sugiere que, si bien SVR puede ser eficaz en conjuntos de datos más pequeños o medianos, su aplicabilidad para downscaling a gran escala puede estar limitada por su elevada complejidad computacional para encontrar los parámetros más óptimos del modelo.

Los resultados se alinean con los hallazgos de (Jing y cols., 2022), (Yuan y cols., 2022), quienes también reportaron mejoras significativas en el Downscaling estadístico utilizando redes neuronales profundas Conv2D y LSTM. Sin embargo, el presente estudio va más allá al demostrar la eficacia específica de las arquitecturas ConvLSTM2D en la captura simultánea de patrones espaciales y temporales en datos climáticos, sobresalió en comparación con todos los demás modelos evaluados, para la creación de nueva información con la reducción temporal de 3 horas a 1 hora para ambas variables, lo que mejoró la resolución temporal de las predicciones y permitió una mayor granularidad en los resultados. Este resultado es coherente con otros estudios recientes que destacan la capacidad de las redes neuronales convolucionales y las LSTM para modelar conjuntamente la dimensión espacial y temporal de los datos meteorológicos (Chou y cols., 2021), (Domala y Kim, 2022), (T. Xu y cols., 2021).

Sin embargo, uno de los desafíos principales observados fue la captura de los valores extremos de temperatura, tanto máximos como mínimos. Esto se debe a la ausencia de datos del GCM en horas críticas como las 04:00-05:00 para las temperaturas mínimas y entre las 13:00-14:00 para las máximas, lo que introduce incertidumbre en las predicciones de estos extremos. Este es un aspecto importante a considerar en futuras investigaciones, donde podría explorarse el uso de técnicas de “prestar atención” o ajuste para focalizar los valores máximos y mínimos de datos horarios.

En cuanto a la precipitación, los resultados mostraron que, a nivel horario, la correlación entre las predicciones y los valores observados es débil, lo que sugiere una alta variabilidad y falta de patrones definidos. Del mismo modo, al agrupar los datos por acumulaciones mensuales, no se logró identificar patrones más claros asociados a lo largo de los meses del año de los datos modelados, lo que indica que la precipitación presentó comportamientos pocos predecibles en escalas temporales más amplias.

Si bien los modelos ConvLSTM2D ofrecieron resultados prometedores, es importante señalar que su implementación es significativamente más compleja y requiere una mayor capacidad computacional que los métodos tradicionales. El tiempo de entrenamiento fue considerablemente mayor y la necesidad de hardware especializado, como GPU, se hizo evidente durante el proceso. Este desafío computacional es un factor limitante, especialmente cuando se pretende implementar estos modelos en estudios climáticos con gran volumen de datos.

4.4. Plataforma web

Una vez identificado el modelo más preciso de downscaling estadístico adaptado al área de estudio, se generaron archivos en formato NetCDF que contienen las proyecciones de temperatura y precipitación para los escenarios climáticos SSP 1-1.9, SSP 1-2.45 y SSP 2-4.0. Estos datos se han dispuesto en una plataforma web diseñada para su visualización y descarga.

La plataforma web ofrece una interfaz intuitiva que muestra los valores de temperatura y precipitación correspondientes a la hora actual de acceso. La aplicación cuenta con dos funcionalidades principales, la primera un control temporal y un panel de control de proyecciones.

La plataforma esta disponible por medio de la dirección <http://18.222.250.4/climatepro/>

4.4.1. Visor geográfico

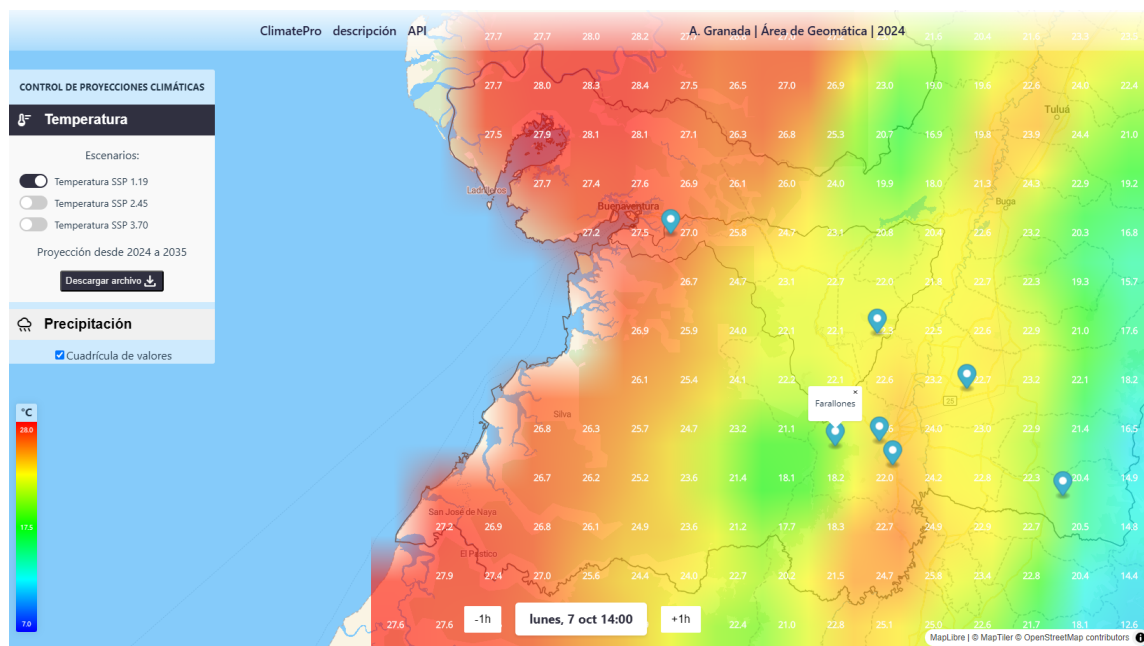


Figura 4.36: Lienzo de visualización de los datos sobre el Pacífico del departamento de Valle del Cauca.

4.4.1.1. Control de proyecciones

Esta función permite a los usuarios seleccionar y visualizar las proyecciones de temperatura y precipitación para cada uno de los escenarios modelados. Además, incluye un botón para la descarga automática del archivo NetCDF correspondiente a la variable y el escenario seleccionado.



Figura 4.37: Control de cada una de las variables en los diferentes escenarios presentados.

4.4.1.2. Control temporal

Ubicado en la parte inferior de la interfaz, este elemento muestra la hora actual y proporciona botones para avanzar o retroceder en el tiempo. Esto permite a los usuarios explorar las proyecciones climáticas en diferentes horas.

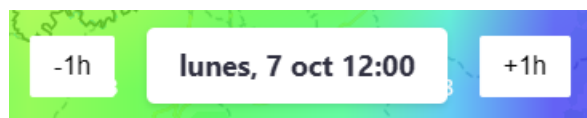


Figura 4.38: Control que permite avanzar y retroceder en el tiempo.

4.4.2. Descarga de archivos NetCDF

El formato de descarga de los archivos climáticos es tipo NetCDF con la extensión (.nc), y esta organizado con las siguientes características:

- **Dimensiones:** ordenamiento de los datos en tiempo, longitud y latitud
 1. Tiempo: 35000 horas
 2. Longitud: 17 valores (-76.2 a -75.5)
 3. Latitud: 14 valores (3.0 a 4.4)

- **Variables:** almacenamiento de los datos de las variables temperatura y precipitación. Matriz tridimensional de la forma [(35000), (17), (14)]
- **Atributos:** metadatos que describen cada una las variables
 1. Temperatura: "Centígrados"
 2. Precipitación: "mm/hora"
 3. Longitud: grados
 4. Latitud: grados
 5. Tiempo: hora

4.4.3. API

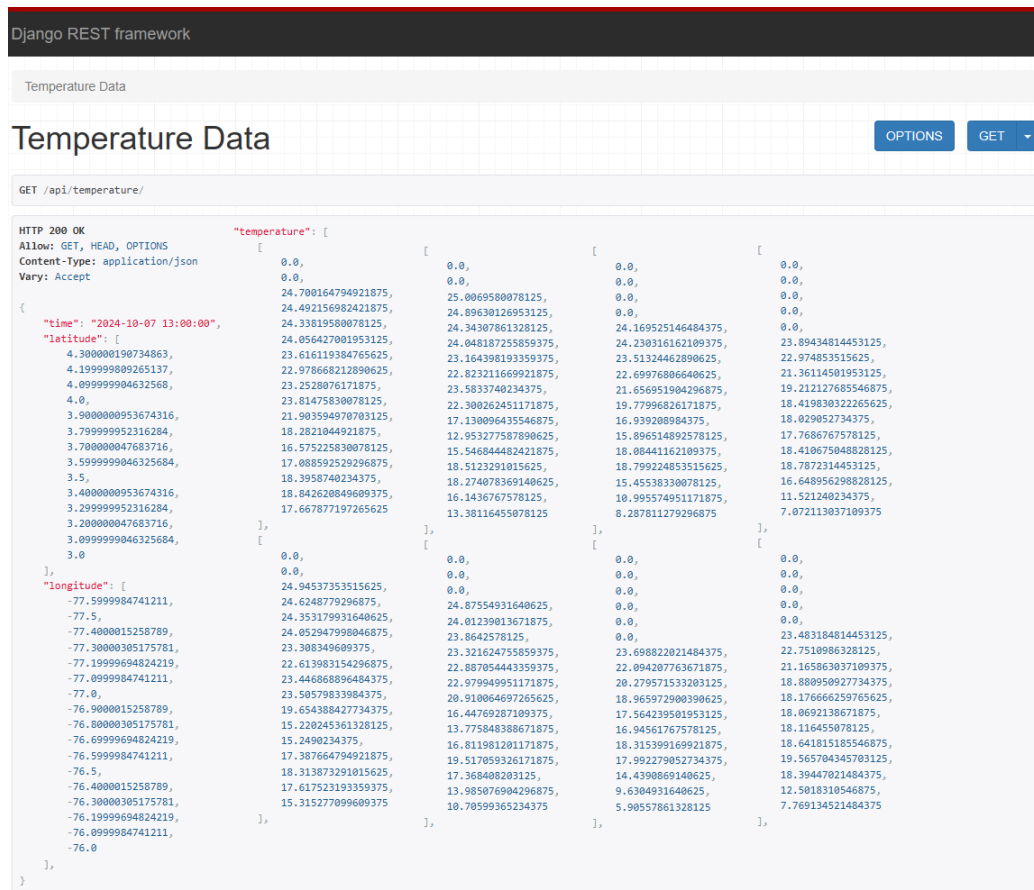
La API RESTful del sistema proporciona acceso programático a los datos climáticos procesados. Está diseñada siguiendo las mejores prácticas de diseño de API y ofrece endpoints específicos para consulta y descarga de datos. A continuación, se detallan los principales endpoints y sus funcionalidades:

1. Consulta de datos en tiempo real:
 - Temperatura: GET <http://18.222.250.4:8000/api/temperatura>
 - Precipitación: GET <http://18.222.250.4:8000/api/precipitacion>
2. Descarga de archivos de datos:
 - Temperatura: GET <http://18.222.250.4:8000/api/ssp1/temperatura>
 - Precipitación: GET <http://18.222.250.4:8000/api/ssp1/precipitacion>

Estos endpoints devuelven un objeto JSON con las coordenadas geográficas y los valores correspondientes a la hora de la petición. La respuesta incluye metadatos como la unidad de medida y la precisión de los datos.

Estos endpoints facilitan la descarga directa de archivos NetCDF que contienen series temporales completas de los datos modelados. Los archivos incluyen metadatos detallados sobre el modelo utilizado, el escenario climático y el período de proyección.

La documentación completa de la API, incluyendo ejemplos de uso, formatos de respuesta y manejo de errores, está disponible en <http://18.222.250.4/api/docs>.



Los resultados obtenidos en el desarrollo de la aplicación web para la publicación y descarga de las proyecciones climáticas demuestran la viabilidad de integrar tecnologías modernas como Python en el backend y React en el frontend para la difusión en tiempo real de proyecciones de alta resolución (C. Zhang y cols., 2022) . El uso de una API para gestionar las solicitudes de datos y permitir la descarga automática de los resultados proyectados facilita una interacción fluida y eficiente entre los usuarios y las proyecciones generadas, lo que representa un avance importante en términos de accesibilidad y manejo de grandes volúmenes de datos climáticos.

espacial y temporal, facilitada por la plataforma, es un recurso valioso para la evaluación de riesgos climáticos y la planificación de medidas de mitigación y adaptación.

Desde una perspectiva técnica, los resultados también evidencian que la combinación de un backend robusto en Python con un frontend React optimizado, gestionando grandes volúmenes de datos climáticos en tiempo real, no solo mejora la eficiencia en la entrega de información, sino que también reduce los tiempos de latencia en la solicitud y descarga de datos proyectados. Esto permite que los usuarios pudieran acceder a datos precisos y actualizados, incrementando la confianza en las proyecciones generadas a partir de los modelos de *Downscaling* estadístico.

Un aspecto clave de la implementación es la capacidad de la plataforma para gestionar grandes cantidades de datos proyectados de manera eficiente y escalable. Los resultados indicaron que la descarga automatizada de los datos seleccionados según las necesidades del usuario resultó ser una herramienta esencial para reducir la carga computacional y optimizar el uso de los recursos del servidor. Además, el sistema fue capaz de proporcionar acceso a datos climáticos proyectados en formatos estandarizados como NetCDF, facilitando su integración en modelos adicionales o en estudios de impacto climático local.

La publicación de proyecciones climáticas detalladas en una aplicación web ha demostrado ser una herramienta efectiva para la difusión de información climática a escala local y regional (Khazaei y cols., 2023). Los resultados sugieren que este enfoque no solo mejora la accesibilidad y gestión de los resultados de la reducción de escala, sino que también permite una interacción personalizada y en tiempo real con los pronósticos de las proyecciones de los escenarios climáticos simulados, lo cual es esencial para sectores que dependen de información climática precisa para la planificación y mitigación de riesgos. Además, el desarrollo de este sistema destaca la importancia de seguir mejorando la tecnología subyacente para garantizar la integración de proyecciones más expensas, precisas y fiables, especialmente en regiones vulnerables como la costa del Pacífico Colombiano.

Capítulo 5

Conclusiones

El análisis exploratorio de datos mostró una fuerte correlación entre los datos de temperatura a resolución horaria proporcionados por ERA5-Land y las observaciones locales de las estaciones meteorológicas. La estación "La Cumbre" presentó los valores más altos, con un coeficiente de determinación $R^2 = 0.78$ y una correlación de Pearson de 0.877, lo que destaca la capacidad de ERA5-Land para capturar la variabilidad térmica en esta ubicación específica. Por otro lado, la estación "Univalle" registró los valores más bajos, con $R^2 = 0.63$ y una correlación de Pearson de 0.79. Estos resultados evidencian que, aunque existen diferencias entre estaciones, las estimaciones de ERA5-Land reproducen adecuadamente la variabilidad térmica a nivel regional, reforzando su confiabilidad como referencia para el downscaling en la región del Pacífico colombiano.

Sin embargo, la precipitación horaria presentó mayores desafíos. Los datos observados no mostraron una correlación significativa con los registros del GCM ni con ERA5-Land, alcanzando valores inferiores a $R^2 = 0.005$ y correlaciones de Spearman por debajo de 0.174. Estas limitaciones pueden atribuirse a la ausencia de datos en algunas estaciones meteorológicas, lo que afecta la calidad y consistencia de las observaciones. Este problema resalta la necesidad de aumentar la densidad y mejorar la distribución de estaciones para garantizar una validación más robusta de los modelos. Sin embargo, las técnicas empleadas permitieron identificar patrones locales a una mejor resolución que la que ofrecen los GCM.

El uso del ACP fue una estrategia efectiva para reducir la dimensionalidad de las variables climáticas proporcionadas por el GCM GFDL-ESM4, el cual cuenta con 11 variables climáticas proyectadas (Temperatura, Precipitación, Componente este del viento, Componente norte del viento, Humedad específica, Humedad del suelo, Presión, Radiación de onda corta descendente, ascendente y Radiación de onda larga descendente y ascendente). Esto permitió identificar las componentes más relevantes para el modelado, preservando la información esencial y mejorando la eficiencia computacional. Este enfoque fue clave para captar la variabilidad local influenciada por factores geográficos, como la longitud, latitud y altura, optimizando así la precisión de los

modelos de predicción.

El estudio destacó la eficacia de las técnicas de inteligencia artificial, particularmente las redes ConvLSTM2D, para el downscaling estadístico de temperatura y precipitación a escala local. Estas redes combinan convoluciones espaciales con LSTM, capturando simultáneamente patrones espaciales y temporales complejos en los datos climáticos. La arquitectura también permitió explorar la distribución espacial de las variables climáticas con mayor detalle, identificando patrones diferenciados en zonas costeras y montañosas.

Técnicas tradicionales, como la Regresión Lineal, sirvieron como referencia base, mientras que modelos como Random Forest mostraron robustez en conjuntos de datos pequeños. Aunque el modelo SVR presentó limitaciones en términos de escalabilidad, su evaluación proporcionó información valiosa para comparar enfoques. El ajuste iterativo de hiperparámetros, incluyendo filtros, tamaños de kernel y tasas de aprendizaje, fue fundamental para optimizar el rendimiento de los modelos y evitar el sobreajuste.

La arquitectura de red neuronal ConvLSTM2D demostró ser altamente efectiva en el downscaling estadístico, logrando mejorar la resolución espacial de 1° a 0.1° y la resolución temporal de 3 horas a 1 hora, generando predicciones en dos pasos de tiempo. Para la variable de temperatura, el modelo alcanzó un coeficiente de determinación R^2 de 0.972, con un RECM de 1.001°C y EAM de 0.733°C . Estos resultados representan una mejora significativa en comparación con los valores obtenidos por la regresión lineal, que presentó $R^2 = 0.848$, un RECM de 1.850°C y un EAM de 1.423°C . Esto evidencia la capacidad del modelo ConvLSTM2D para capturar dependencias espaciales y temporales complejas, superando ampliamente a las técnicas más tradicionales.

Para la variable de precipitación, el modelo ConvLSTM2D obtuvo un R^2 de 0.020, con un RECM de 0.264 mm/h y un EAM de 0.069 mm/h . En comparación, la regresión lineal presentó un desempeño ligeramente mejor en términos de correlación $R^2 = 0.038$, pero con errores más altos: un RECM de 0.619 mm/h y un EAM de 0.342 mm/h . Estos resultados reflejan la alta complejidad inherente a la representación horaria de las precipitaciones, lo que destaca la necesidad de explorar ajustes adicionales en las arquitecturas y enfoques utilizados para mejorar la precisión de las predicciones en esta variable.

Los resultados obtenidos destacan que el downscaling estadístico basado en inteligencia artificial para generar predicciones locales a una mejor precisión, superando el desempeño de modelos tradicionales como la RL, RF, RVS. Esta capacidad resulta esencial para abordar los desafíos asociados al cambio climático, proporcionando información detallada que respalda tanto la planificación adaptativa como la toma de decisiones estratégicas a nivel regional. Este avance es particularmente relevante en la costa del Pacífico colombiano, una región caracterizada por una alta variabilidad en la precipitación y la temperatura, factores críticos en la gestión integrada de

recursos hídricos y la mitigación de riesgos climáticos, especialmente frente a eventos extremos que amenazan la sostenibilidad de las comunidades locales.

Los resultados obtenidos se integraron en una plataforma web diseñada para facilitar el acceso a las proyecciones climáticas generadas por medio de la URL <http://18.222.250.4/climatepro>. La elección de React y MapLibre como herramientas principales permitió manejar grandes volúmenes de datos geoespaciales de forma eficiente. MapLibre posibilitó la implementación de funcionalidades avanzadas, como zoom suave, rotación y superposición de capas de los datos climáticos sobre mapas detallados, manteniendo un alto rendimiento y pocas exigencias computacionales.

La arquitectura modular del frontend, desarrollada en React, no solo facilita la publicación clara y comprensible de los datos, sino que también ofrece flexibilidad para futuras expansiones y mejoras. La plataforma actúa como un puente entre la complejidad de los modelos climáticos y la necesidad de información accesible, promoviendo una toma de decisiones informada y fomentando una mayor conciencia sobre los desafíos climáticos en la región.

A pesar de los resultados prometedores, el estudio enfrenta limitaciones importantes. La dependencia de observaciones históricas para validar los modelos plantea desafíos en un clima cambiante, donde las relaciones pasadas entre variables pueden no mantenerse en el futuro. Esto subraya la necesidad de incorporar enfoques dinámicos que capten tendencias emergentes en escenarios de cambio climático.

Además, la baja densidad y distribución desigual de estaciones meteorológicas introduce incertidumbre en la evaluación de los modelos, particularmente en áreas con cobertura limitada. La integración de fuentes adicionales, como observaciones satelitales y técnicas avanzadas de interpolación espacial, podría mejorar la precisión en estas regiones.

Futuros estudios podrían explorar la incorporación de capas profundas con configuraciones más complejas y series temporales extendidas, ajustadas a los recursos computacionales disponibles. También sería valioso evaluar la aplicabilidad de estas técnicas en diferentes regiones climáticas para validar su robustez y generalización.

BIBLIOGRAFÍA

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems [Manual de software informático]. Descargado de <https://arxiv.org/abs/1603.04467>
- Abdollahi, M., Farjad, B., Gupta, A., y Hassan, Q. K. (2022, JUN). Cmp6-d&a: An r-based software with gui for processing climate data available in network common data format. *SOFTWAREX*, 18. doi: 10.1016/j.softx.2022.101044
- Admasu, L. M., Grant, L., y Thiery, W. (2023). Exploring global climate model downscaling based on tile-level output. *Journal of Applied Meteorology and Climatology*, 62(2), 171 - 190. Descargado de <https://journals.ametsoc.org/view/journals/apme/62/2/JAMC-D-21-0265.1.xml> doi: 10.1175/JAMC-D-21-0265.1
- Agrawal, A., Verma, D., y Gupta, S. (2016). Exploratory data analysis on temperature data of indian states from 1800–2013 (analysis of trends in temperature data for prediction modelling). En *2016 2nd international conference on next generation computing technologies (ngct)* (p. 547-552). doi: 10.1109/NGCT.2016.7877475
- Aguirre, M. A., Ibarra, L. I. L., Trochez, F. V. B., Guevara, D. F. G., y Bermudez, O. B. (2017, 5). Percepción del paisaje, agua y ecosistemas en la cuenca del río Dagua, Valle del Cauca, Colombia. *Perspectiva Geográfica*, 22. doi: 10.19053/01233769.5402
- Apache Friends, A. (2024). *Xampp: Apache + mariadb + php + perl*. Descargado de <https://www.apachefriends.org> (Cross-platform web server solution)
- Araya-Osses, D., Casanueva, A., Roman-Figueroa, C., Manuel Uribe, J., y Paneque, M. (2020, MAY). Climate change projections of temperature and precipitation in chile based on statistical downscaling. *CLIMATE DYNAMICS*, 54(9-10), 4309-4330. doi: 10.1007/s00382-020-05231-4
- Arias, P. A., Ortega, G., Villegas, L. D., y Martínez, J. A. (2021). La climatología colombiana en modelos cmp5/cmp6: Sesgos persistentes y mejoras. *Revista Facultad de Ingeniería*, 75-96. Descargado de www.ideam.gov.co/documents/21021/553571/ doi: 10.17533/udea.redin

- Arun, A. M., Chowdary, V. M., Kesarwani, M., y Neeti, N. (2023, 1). Integrated drought monitoring and assessment using multi-sensor and multi-temporal earth observation datasets: a case study of two agriculture-dominated states of india. *Environmental Monitoring and Assessment*, 195. doi: 10.1007/s10661-022-10550-6
- Ashraf, M., Arshad, A., Patel, P. M., Khan, A., Qamar, H., Siti-Sundari, R., ... Babar, J. R. (2021, 12). Quantifying climate-induced drought risk to livelihood and mitigation actions in balochistan. *Natural Hazards*, 109, 2127-2151. doi: 10.1007/s11069-021-04913-4
- Baghanam, A. H., Eslahi, M., Sheikhababaei, A., y Seifi, A. J. (2020, 8). Assessing the impact of climate change over the northwest of iran: an overview of statistical downscaling methods. *Theoretical and Applied Climatology*, 141, 1135-1150. doi: 10.1007/s00704-020-03271-8
- Baker, J. C. A., de Souza, D. C., Kubota, P. Y., Buermann, W., Coelho, C. A. S., Andrews, M. B., ... Spracklen, D., V (2021, APR). An assessment of land-atmosphere interactions over south america using satellites, reanalysis, and two global climate models. *JOURNAL OF HYDROMETEOROLOGY*, 22(4), 905-922. doi: 10.1175/JHM-D-20-0132.1
- Ballarin, A. S., Sone, J. S., Gesualdo, G. C., Schwambach, D., Reis, A., Almagro, A., y Wendland, E. C. (2023, 12). Climbra - climate change dataset for brazil. *Scientific Data*, 10. doi: 10.1038/s41597-023-01956-z
- Balmaceda-Huarte, B.-M. J. O. M. E. B. M. L., Rocío. (2024, FEB). On the use of convolutional neural networks for downscaling daily temperatures over southern south america in a climate change scenario. *Climate Dynamics*, 62(1), 383-397. Descargado de <https://doi.org/10.1007/s00382-023-06912-6> doi: 10.1007/s00382-023-06912-6
- Bao, R., Zhang, W., Ma, T., y Gan, Z. (2023). Design of management system for wr timing system based on django framework. En *2023 ieee international conference on electrical, automation and computer engineering (iceace)* (p. 1734-1738). doi: 10.1109/ICEACE60673.2023.10442435
- Baquero, O., Jiménez Escobar, H., Carvajal Escobar, Y., López García, J., Mina Viveros, B., Ayala Arboleda, I., y Córdoba Bravo, C. (2018). *Estudio del efecto de la variabilidad climática y el cambio de uso del suelo en la soberanía alimentaria de una comunidad del litoral Pacífico vallecaucano* (Informe de investigación). Cali, Colombia: Universidad del Valle.
- Bartkowiak, P., Castelli, M., y Notarnicola, C. (2019, JUN 1). Downscaling land surface temperature from modis dataset with random forest approach over alpine vegetated areas. *REMOTE SENSING*, 11(11). doi: 10.3390/rs11111319

- Biniz, E. A.-R., Mohamed. (2021, 06). Recognition of tfinagh characters using optimized convolutional neural network. *Sensing and Imaging*, 22. doi: 10.1007/s44196-023-00397-1
- Bordera, J., Valladares, F., Turiel, A., Vilar, F. P., Prieto, F., y Hewlett, T. (2022, 1). Advertencias del panel intergubernamental del cambio climático. *Revista de Economía Institucional*, 24, 237-247. doi: 10.18601/01245996.v24n46.12
- Caicedo, E. F., y López, J. A. (2017). *Una aproximación práctica a las redes neuronales artificiales*. Cali, Colombia: Universidad del Valle. Descargado de <https://hdl.handle.net/10893/10330> doi: 10.25100/peu.64
- Canchala, T., Ocampo-Marulanda, C., Alfonso-Morales, W., Carvajal-Escobar, Y., Cerón, W. L., y Caicedo-Bravo, E. (2022). Techniques for monthly rainfall regionalization in southwestern Colombia. *Anais da Academia Brasileira de Ciencias*, 94, e20201000. doi: 10.1590/0001-3765202220201000
- Cao, S., Cao, C., Li, Y., y Zhu, L. (2023, JUN). A statistical downscaling model based on multiway functional principal component analysis for southern australia winter rainfall. *JOURNAL OF APPLIED METEOROLOGY AND CLIMATOLOGY*, 62(6), 677-689. doi: 10.1175/JAMC-D-22-0101.1
- Carlsonweb. (2022, 1). C-21. *Revistann*, 01, 237-247. doi: 10.18601/01245996.v24n46.12
- Chen, C., Chen, Q., Qin, B., Zhao, S., y Duan, Z. (2020, 11). Comparison of different methods for spatial downscaling of gpm imerg v06b satellite precipitation product over a typical arid to semi-arid area. *Frontiers in Earth Science*, 8. doi: 10.3389/feart.2020.536337
- Cheng, J., Kuang, Q., Shen, C., Liu, J., Tan, X., y Liu, W. (2020). Reslap: Generating high-resolution climate prediction through image super-resolution. *IEEE Access*, 8, 39623-39634. doi: 10.1109/ACCESS.2020.2974785
- Chou, C., Park, J., y Chou, E. (2021, 5). Generating high-resolution climate change projections using super-resolution convolutional lstm neural networks. En (p. 293-298). Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICACI52617.2021.9435890
- Christie, T., y cols. (2024). *Django REST framework*. Descargado de <https://www.django-rest-framework.org> (Version 3.14)
- contributors. MapLibreJS, M. (2024). *Maplibre gl js: Open-source libraries for publishing maps*. Descargado de <https://maplibre.org> (Open source mapping platform)

Copernicus Climate Change Service, C. (2024). Era5-land hourly data from 1950 to present. Descargado de <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview>

CVC, . (2020). Grupo sistema de información ambiental guía temática para el usuario sig corporativo red hídrica tabla de contenido.

CVC. (2017). CVC, 2017 Balance Oferta – Demanda de agua.

Dash, R. K., Nguyen, T. N., Cengiz, K., y Sharma, A. (2023). Fine-tuned support vector regression model for stock predictions. *Neural Computing and Applications*, 35(32), 23295–23309. Descargado de <https://doi.org/10.1007/s00521-021-05842-w> doi: 10.1007/s00521-021-05842-w

De, M., Adacheski, P., y Virgens Filho, J. (2011, 03). Análise da frequência e da intensidade das chuvas em ponta grossa, estado do paraná, no período entre 1954 e 2001. *Acta Scientiarum Technology*, 33, 57-64. doi: 10.4025/actascitechnol.v33i1.6957

Domala, V., y Kim, T.-W. (2022). A univariate and multivariate machine learning approach for prediction of significant wave height. En *Oceans 2022, hampton roads* (p. 1-7). doi: 10.1109/OCEANS47191.2022.9977028

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., ... Zhao, M. (2020). The gfdl earth system model version 4.1 (gfdl-esm 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, 12(11), e2019MS002015. Descargado de <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002015> (e2019MS002015 2019MS002015) doi: <https://doi.org/10.1029/2019MS002015>

Durbin, J., y Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4), 409–428. Descargado 2025-03-27, de <http://www.jstor.org/stable/2332391>

Ebrahimi, H., y Azadbakht, M. (2019, 3). Downscaling modis land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels. *Computers and Geosciences*, 124, 93-102. doi: 10.1016/j.cageo.2019.01.004

Ecma International. (2024). *ECMAScript 2024 language specification* (Standard n.º ECMA-262). Ecma International. Descargado de <https://www.ecma-international.org/publications-and-standards/standards/ecma-262/> (Edition 14)

- ECMWF. (2019). First ERA5-Land dataset to be released in spring. *ECMWF Newsletter*(159). Descargado de <https://www.ecmwf.int/en/newsletter/159/news/first-era5-land-dataset-be-released-spring> (Consultado el 17 de octubre de 2023)
- Eekhout, J. P., Hunink, J. E., Terink, W., y Vente, J. D. (2018, 11). Why increased extreme precipitation under climate change negatively affects water security. *Hydrology and Earth System Sciences*, 22, 5935-5946. doi: 10.5194/hess-22-5935-2018
- El Qorchi, F., Yacoubi Khebiza, M., Omondi, O. A., Karmaoui, A., Pham, Q. B., y Acharki, S. (2023). Analyzing temporal patterns of temperature, precipitation, and drought incidents: A comprehensive study of environmental trends in the upper draa basin, morocco. *Water*, 15(22). Descargado de <https://www.mdpi.com/2073-4441/15/22/3906> doi: 10.3390/w15223906
- Enriquez, O., Guzmán, A., y Narváez, G. (2014, 01). Análisis del comportamiento de la precipitación en el municipio de Buenaventura (Valle del Cauca, Colombia) en condiciones de desarrollo de los fenómenos El Niño y La Niña. *Cuadernos de Geografía: Revista Colombiana de Geografía*, 23, 165 - 178. Descargado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-215X2014000100013&nrm=iso
- Erlandsen, H. B., Parding, K. M., Benestad, R., Mezghani, A., y Pontoppidan, M. (2020, NOV). A hybrid downscaling approach for future temperature and precipitation change. *JOURNAL OF APPLIED METEOROLOGY AND CLIMATOLOGY*, 59(11), 1793-1807. doi: 10.1175/JAMC-D-20-0013.1
- Estoque, R. C., Ooba, M., Togawa, T., y Hijioka, Y. (2020). Projected land-use changes in the shared socioeconomic pathways: Insights and implications. *Ambio*, 49(12), 1972-1981. Descargado de <https://doi.org/10.1007/s13280-020-01338-4> doi: 10.1007/s13280-020-01338-4
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., y Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937-1958. Descargado de <https://gmd.copernicus.org/articles/9/1937/2016/> doi: 10.5194/gmd-9-1937-201
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., ... Xiang, Y. (2018, 12). Evaluation of svm, elm and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of china. *Agricultural and Forest Meteorology*, 263, 225-241. doi: 10.1016/j.agrformet.2018.08.019

- FAO. (2022, 1). Climate maps for the lao people's democratic republic. part of the land resources information management system (Irimis). *Revistann*, 01, 237-247. doi: 10.18601/01245996.v24n46.12
- Fedesarrollo. (2013). *Ciudades y Cambio Climático en Colombia Proyecto financiado por AFD Fedesarrollo Fundación Ciudad Humana Instituto para la Investigación y Debate sobre la Gobernanza (IRG) Octubre 2013 1*.
- Feijt, M. (2015). *The impact of climate change and climate variability on areal groundwater recharge in Colombia's largest sugar cane region, Valle del Cauca* (Tesis de Master no publicada). Wageningen University, the Netherlands. (Supervisors: dr.ir Martine van der Ploeg, Marta Faneca Sánchez MSc, Micha Werner PhD MSc)
- Feser, F., Rrockel, B., Storch, H., Winterfeldt, J., y Zahn, M. (2011, 9). Regional climate models add value to global model data a review and selected examples. *Bulletin of the American Meteorological Society*, 92, 1181-1192. doi: 10.1175/2011BAMS3061.1
- Foundation, T. A. S. (2024). *The apache http server project*. Descargado de <https://httpd.apache.org> (Version 2.4)
- Galiano, V., Migallon, H., Migallon, V., y Penades, J. (2010, JAN). Pypnetcdf: A high level framework for parallel access to netcdf files. *ADVANCES IN ENGINEERING SOFTWARE*, 41(1), 92-98. doi: 10.1016/j.advengsoft.2009.06.005
- Gaustad, K., Shippert, T., Ermold, B., Beus, S., Daily, J., Borsholm, A., y Fox, K. (2014, OCT). A scientific data processing framework for time series netcdf data. *ENVIRONMENTAL MODELLING & SOFTWARE*, 60, 241-249. doi: 10.1016/j.envsoft.2014.06.005
- Ghaderpour, E., Dadkhah, H., Dabiri, H., Bozzano, F., Scarascia Mugnozza, G., y Mazzanti, P. (2023). Precipitation time series analysis and forecasting for italian regions. *Engineering Proceedings*, 39(1). Descargado de <https://www.mdpi.com/2673-4591/39/1/23> doi: 10.3390/engproc2023039023
- Ghil, M., y Lucarini, V. (2020, Jul). The physics of climate variability and climate change. *Rev. Mod. Phys.*, 92, 035002. Descargado de <https://link.aps.org/doi/10.1103/RevModPhys.92.035002> doi: 10.1103/RevModPhys.92.035002
- Gomez Gonzalez, C. A. (2023). DI4ds—deep learning for empirical downscaling. *Environmental Data Science*, 2, e3. doi: 10.1017/eds.2022.26

- Gong, X., Wu, X., y Zhou, X. (2023). Deep learning-based security situational awareness and detection technology for power networks in the context of big data. *Applied Mathematics and Nonlinear Sciences*, 8(1), 2939–2956. Descargado de <https://doi.org/10.2478/amns.2023.1.00012> doi: 10.2478/amns.2023.1.00012
- González, N. (2014). Variabilidad espacio-temporal de la susceptibilidad a sequías meteorológicas en la cuenca del río Dagua, Valle del Cauca. *Universidad del Valle*, 77-78. Descargado de <https://hdl.handle.net/10893/15885>
- Gonçalves, A. M., Pereira, F. C., Costa, M., y Leão, C. P. (2023). Meteorological time series: An exploratory statistical and critical analysis. En (p. 197-208). Springer Science and Business Media Deutschland GmbH. doi: 10.1007/978-3-031-09360-9_17
- Goosse, H., Barriat, P.-Y., Loutre, M.-F., y Zunz, V. (2010). *Introduction to climate dynamics and climate modeling*. Centre de recherche sur la Terre et le climat Georges Lemaître - UCLouvain. Descargado de <https://oer.uclouvain.be/jspui/handle/20.500.12279/286> (Consultado el 4 de febrero de 2024)
- Gu, J., Ye, Y., Jiang, Y., Dong, J., Cao, Y., Huang, J., y Guan, H. (2023). A downscaling-calibrating framework for generating gridded daily precipitation estimates with a high spatial resolution. *Journal of Hydrology*, 626, 130371. Descargado de <https://www.sciencedirect.com/science/article/pii/S0022169423013136> doi: <https://doi.org/10.1016/j.jhydrol.2023.130371>
- Guo, H., Bao, A., Chen, T., Zheng, G., Wang, Y., Jiang, L., y Maeyer, P. D. (2021, 4). Assessment of cmip6 in simulating precipitation over arid central asia. *Atmospheric Research*, 252. doi: 10.1016/j.atmosres.2021.105451
- Gutiérrez, S., Carvajal, Y., y Ávila, A. (2013). Analysis of the influence of phenomena el niño-southern oscillation in the water supply of the Dagua river drainage watershed. *Entre Ciencia e Ingeniería*, 26-33.
- Guyon, I. (1997). A scaling law for the validation-set training-set size ratio. *AT&T Bell Laboratories*.
- Hamarashid, H. (2021, 05). Modified long short-term memory and utilizing in building sequential model. *International Journal of Multidisciplinary and Current Research*. doi: 10.14741/ijmcr/v.9.3.2

- Hamed, M. M., Nashwan, M. S., Shahid, S., Wang, X. J., Ismail, T. B., Dewan, A., y Asaduzzaman, M. (2023, 3). Future köppen-geiger climate zones over southeast asia using cmip6 multimodel ensemble. *Atmospheric Research*, 283. doi: 10.1016/j.atmosres.2022.106560
- Hamidi, M., y Roshani, A. (2023). Investigation of climate change effects on iraq dust activity using lstm. *Atmospheric Pollution Research*, 14(10), 101874. Descargado de <https://www.sciencedirect.com/science/article/pii/S1309104223002283> doi: <https://doi.org/10.1016/j.apr.2023.101874>
- Han, J., Chong, A., Lim, J., Ramasamy, S., Wong, N. H., y Biljecki, F. (2024). Microclimate spatio-temporal prediction using deep learning and land use data. *Building and Environment*, 253, 111358. Descargado de <https://www.sciencedirect.com/science/article/pii/S0360132324002002> doi: <https://doi.org/10.1016/j.buildenv.2024.111358>
- Hannah, L. (2015). Chapter 2 - the climate system and climate change. En L. Hannah (Ed.), *Climate change biology (second edition)* (Second ed., p. 13-53). Boston: Academic Press. Descargado de <https://www.sciencedirect.com/science/article/pii/B9780124202184000020> doi: <https://doi.org/10.1016/B978-0-12-420218-4.00002-0>
- Hannah, L. (2022). Chapter 2 - The Climate System and Climate Change. En L. Hannah (Ed.), *Climate change biology (third edition)* (Third ed., p. 13-50). Academic Press. Descargado de <https://www.sciencedirect.com/science/article/pii/B9780081029756000029> doi: <https://doi.org/10.1016/B978-0-08-102975-6.00002-9>
- He, K., Zhang, X., Ren, S., y Sun, J. (2016). Deep residual learning for image recognition. En *2016 ieee conference on computer vision and pattern recognition (cvpr)* (p. 770-778). doi: 10.1109/CVPR.2016.90
- Hernanz, A., Correa, C., Andrés García-Valero, J., Domínguez, M., Rodríguez-Guisado, E., y Rodríguez-Camino, E. (2023). pyclim-sdm: Service for generation of statistically downscaled climate change projections supporting national adaptation strategies. *Climate Services*, 32, 100408. Descargado de <https://www.sciencedirect.com/science/article/pii/S2405880723000705> doi: <https://doi.org/10.1016/j.cliser.2023.100408>
- Hochreiter, S., y Schmidhuber, J. (1997, 12). Long short-term memory. *Neural computation*, 9, 1735-80. doi: 10.1162/neco.1997.9.8.1735
- Hooker, J., Duveiller, G., y Cescatti, A. (2018). Data descriptor: A global dataset of air temperature derived from satellite remote sensing and weather stations. *Scientific Data*, 5. doi: 10.1038/sdata.2018.246

- Hoyer, S., y Hamman, J. J. (2017). xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5(1). Descargado de <https://doi.org/10.5334/jors.148> doi: 10.5334/jors.148
- Hoyos, I., Baquero-Bernal, A., Jacob, D., y Rodríguez, B. A. (2013, 4). Variability of extreme events in the colombian pacific and caribbean catchment basins. *Climate Dynamics*, 40, 1985-2003. doi: 10.1007/s00382-012-1487-9
- Hrisko, J., Ramamurthy, P., Yu, Y., Yu, P., y Melecio-Vázquez, D. (2020, 2). Urban air temperature model using goes-16 l1 and a diurnal regressive neural network algorithm. *Remote Sensing of Environment*, 237. doi: 10.1016/j.rse.2019.111495
- IDEAM. (2005). *Atlas climatológico de Colombia*. Ideam (Instituto de Hidrología, Meteorología y Estudios Ambientales).
- IDEAM. (Consultado en el años 2023). *Análisis de las Precipitaciones Horarias en Colombia* (Inf. Téc.). Instituto de Hidrología, Meteorología y Estudios Ambientales. Descargado de <http://www.ideam.gov.co/web/tiempo-y-clima/clima> (Subdirección de Meteorología, Grupo de Climatología, Agroclimatología y Meteorología Marina)
- IDEAM - UNAL. (2018). *Variabilidad Climática y Cambio Climático en Colombia*. Bogotá, D.C.: Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) y Universidad Nacional de Colombia (UNAL).
- IDEAM-AQUARIUS, . (2024). Aquarius web portal. Descargado de <http://aquariuswebportal.ideam.gov.co/Data/Map/Parameter/NoParameter/Interval/Latest>
- Ioffe, S., y Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. En *International conference on machine learning* (pp. 448–456). Descargado de <https://arxiv.org/abs/1502.03167> doi: <https://doi.org/10.48550/arXiv.1502.03167>
- IPCC. (2023). *Climate change 2022: Synthesis report. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change* (Inf. Téc.). Geneva, Switzerland: Intergovernmental Panel on Climate Change. Descargado de <https://www.ipcc.ch/report/ar6/syr/>
- Jalota, S., Vashisht, B., Sharma, S., y Kaur, S. (2018). Chapter 2 - climate change projections. En S. Jalota, B. Vashisht, S. Sharma, y S. Kaur (Eds.), *Understanding climate*

change impacts on crop productivity and water balance (p. 55-86). Academic Press. Descargado de <https://www.sciencedirect.com/science/article/pii/B9780128095201000021> doi: <https://doi.org/10.1016/B978-0-12-809520-1.00002-1>

JAXA. ([2024]). *Alos palsar l1.0 raw data*. Descargado de <https://asf.alaska.edu/datasets/daac/alos-palsar/> (Accessed through ASF DAAC [2024])

Jing, Y., Lin, L., Li, X., Li, T., y Shen, H. (2022, 10). An attention mechanism based convolutional network for satellite precipitation downscaling over china. *Journal of Hydrology*, 613. doi: 10.1016/j.jhydrol.2022.128388

Kamworapan, S., Thao, P. T. B., Gheewala, S. H., Pimonsree, S., y Prueksakorn, K. (2021, 11). Evaluation of cmip6 gcms for simulations of temperature over thailand and nearby areas in the early 21st century. *Heliyon*, 7. doi: 10.1016/j.heliyon.2021.e08263

Karaman, H., y Akyürek, Z. (2023). Evaluation of near-surface air temperature reanalysis datasets and downscaling with machine learning based random forest method for complex terrain of turkey. *Advances in Space Research*. doi: 10.1016/j.asr.2023.02.006

Khazaei, M., Hamzeh, S., Samani, N. N., Muhuri, A., Goita, K., y Weng, Q. (2023). A web-based system for satellite-based high-resolution global soil moisture maps. *Computers - Geosciences*, 170, 105250. Descargado de <https://www.sciencedirect.com/science/article/pii/S0098300422001996> doi: <https://doi.org/10.1016/j.cageo.2022.105250>

Kheir, A. M., Elnashar, A., Mosad, A., y Govind, A. (2023). An improved deep learning procedure for statistical downscaling of climate data. *Heliyon*, 9(7), e18200. Descargado de <https://www.sciencedirect.com/science/article/pii/S2405844023054087> doi: <https://doi.org/10.1016/j.heliyon.2023.e18200>

Kim, J., Choi, J., Choi, C., y Park, S. (2013, 5). Impacts of changes in climate and land use/land cover under ipcc rcp scenarios on streamflow in the hoeya river basin, korea. *Science of the Total Environment*, 452-453, 181-195. doi: 10.1016/j.scitotenv.2013.02.005

Kim, J., Lee, M., Han, H., Kim, D., Bae, Y., y Kim, H. S. (2022). Case study: Development of the cnn model considering teleconnection for spatial downscaling of precipitation in a climate change scenario. *Sustainability*, 14(8). Descargado de <https://www.mdpi.com/2071-1050/14/8/4719> doi: 10.3390/su14084719

Li, Y., y Yan, X. (2017). Statistical downscaling of monthly mean temperature for kazakhstan in central asia. *Climate Research*, 72(2), 101-110. Descargado de <https://www.int-res.com/abstracts/cr/v72/n2/p101-110/> doi: 10.3354/cr01456

- Lima, C. H., Kwon, H. H., y Kim, Y. T. (2021, 6). A bayesian kriging model applied for spatial downscaling of daily rainfall from gcms. *Journal of Hydrology*, 597. doi: 10.1016/j.jhydrol.2021.126095
- Liu, D. L., y Zuo, H. (2012, 11). Statistical downscaling of daily climate variables for climate change impact assessment over new south wales, australia. *Climatic Change*, 115, 629-666. doi: 10.1007/s10584-012-0464-y
- Liu, L., Dong, Z., Gong, H., Wang, L., Chen, W., y Wu, R. (2022, 12). Climatology and trends of wintertime diurnal temperature range over east asia in cmip6 models: Evaluation and attribution. *Atmospheric Research*, 280. doi: 10.1016/j.atmosres.2022.106438
- Liu, Y., Jing, W., Wang, Q., y Xia, X. (2020). Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms. *Advances in Water Resources*, 141, 103601. Descargado de <https://www.sciencedirect.com/science/article/pii/S0309170819311972> doi: <https://doi.org/10.1016/j.advwatres.2020.103601>
- Lyra Fialho Breda, J. P., Dias de Paiva, R. C., Chou, S. C., y Collischonn, W. (2022, DEC 30). Assessing extreme precipitation from a regional climate model in different spatial-temporal scales: A hydrological perspective in South America. *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, 42(16), 8904-8927. doi: 10.1002/joc.7782
- Mehta, D., Yadav, S., Ladavia, C., y Caloiero, T. (2023). Drought projection using gcm & statistical downscaling technique: A case study of sirohi district. *Results in Engineering*, 20, 101605. Descargado de <https://www.sciencedirect.com/science/article/pii/S2590123023007326> doi: <https://doi.org/10.1016/j.rineng.2023.101605>
- Mhaweji, M., y Abunnasr, Y. (2022, 11). Daily ten-st-gee: An open access and fully automated 10-m lst downscaling system. *Computers and Geosciences*, 168. doi: 10.1016/j.cageo.2022.105220
- Michna, P., y Woods, M. (2013, DEC). Rnetcdf - a package for reading and writing netcdf datasets. *R JOURNAL*, 5(2), 29-36.
- MinCencias. (2022). *Convocatoria de la asignación para la CTel del SGR para la conformación de un listado de propuestas de proyecto elegibles para la vinculación de jóvenes investigadores e innovadores en las regiones para atención de demandas definidas por los CODECTI*. Ministerio de Ciencia, Tecnología e Innovación de Colombia. Descargado de <https://www.minciencias.gov.co/> (Ministerio de Ciencia, Tecnología e Innovación de Colombia)

- Ministerio de Ambiente y Desarrollo Sostenible. (2017). *Política nacional de cambio climático* (Inf. Téc.). Bogotá, D.C., Colombia: Ministerio de Ambiente y Desarrollo Sostenible. Descargado de <https://www.minambiente.gov.co/wp-content/uploads/2022/01/9.-Politica-Nacional-de-Cambio-Climatico.pdf> (Ministro: Luis Gilberto Murillo)
- Misra, S., Sarkar, S., y Mitra, P. (2018, 11). Statistical downscaling of precipitation using long short-term memory recurrent neural networks. *Theoretical and Applied Climatology*, 134, 1179-1196. doi: 10.1007/s00704-017-2307-2
- Muñoz, W. (2018). Integración Multiescala entre modelos climáticos globales y regionales mediante técnicas de inteligencia artificial. *Universidad del Valle*. Descargado de <https://bibliotecadigital.univalle.edu.co/handle/10893/11050>
- Nishant, N., Di Virgilio, G., Ji, F., Tam, E., Beyer, K., y Riley, M. L. (2022). Evaluation of present-day cmip6 model simulations of extreme precipitation and temperature over the australian continent. *Atmosphere*, 13(9). Descargado de <https://www.mdpi.com/2073-4433/13/9/1478> doi: 10.3390/atmos13091478
- NOAA. (2024). National oceanic and atmospheric administration. *US Department of Commerce*. Descargado de <https://www.noaa.gov/>
- Nourani, V., Paknezhad, N. J., Mohammadisepasi, S., y Zhang, Y. (2024). Regionalization of grace data in shorelines by ensemble of artificial intelligence methods. *Journal of Hydrology*, 636, 131268. Descargado de <https://www.sciencedirect.com/science/article/pii/S0022169424006632> doi: <https://doi.org/10.1016/j.jhydrol.2024.131268>
- Olén, N. B., y Lehsten, V. (2022, 2). High-resolution global population projections dataset developed with cmip6 rcp and ssp scenarios for year 2010–2100. *Data in Brief*, 40. doi: 10.1016/j.dib.2022.107804
- Ortega, G., Arias, P. A., Villegas, J. C., Marquet, P. A., y Nobre, P. (2021). Present-day and future climate over central and south america according to cmip5/cmip6 models. *International Journal of Climatology*, 41(15), 6713-6735. Descargado de <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.7221> doi: <https://doi.org/10.1002/joc.7221>
- Osborn, T. J., Jones, P. D., Lister, D. H., Morice, C. P., Simpson, I. R., Winn, J. P., ... Harris, I. C. (2021, 1). Land surface air temperature variations across the globe updated to 2019: The crutem5 data set. *Journal of Geophysical Research: Atmospheres*, 126. doi: 10.1029/2019JD032352

- Pahlavan, H. A., Zahraie, B., Nasser, M., y Mahdipour Varnousfaderani, A. (2018). Improvement of multiple linear regression method for statistical downscaling of monthly precipitation. *International Journal of Environmental Science and Technology*, 15(9), 1897–1912. Descargado de <https://doi.org/10.1007/s13762-017-1511-z> doi: 10.1007/s13762-017-1511-z
- Peng, S., Wang, C., Li, Z., Mihara, K., Kuramochi, K., Toma, Y., y Hatano, R. (2023, 12). Climate change multi-model projections in cmip6 scenarios in central hokkaido, japan. *Scientific Reports*, 13. doi: 10.1038/s41598-022-27357-7
- Potgieter, A. B., Schepen, A., Brider, J., y Hammer, G. L. (2022, 9). Lead time and skill of australian wheat yield forecasts based on enso-analogue or gcm-derived seasonal climate forecasts – a comparative analysis. *Agricultural and Forest Meteorology*, 324. doi: 10.1016/j.agrformet.2022.109116
- Pour, S. H., Shahid, S., Chung, E. S., y Wang, X. J. (2018, 11). Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of bangladesh. *Atmospheric Research*, 213, 149-162. doi: 10.1016/j.atmosres.2018.06.006
- Preisendorfer, R., y Mobley, C. (1988). *Principal component analysis in meteorology and oceanography*. Elsevier. Descargado de https://books.google.com.co/books?id=Mt_scAAACAAJ
- Rajanikanth, J., y Kanth, T. V. R. (2017). An explorative data analysis on bangalore city weather with hybrid data mining techniques using r. En *2017 international conference on current trends in computer, electrical, electronics and communication (ctceec)* (p. 1121-1125). doi: 10.1109/CTCEEC.2017.8455008
- Rampal, N., Gibson, P. B., Sood, A., Stuart, S., Fauchereau, N. C., Brandolino, C., ... Meyers, T. (2022, 12). High-resolution downscaling with interpretable deep learning: Rainfall extremes over new zealand. *Weather and Climate Extremes*, 38. doi: 10.1016/j.wace.2022.100525
- Randall, D., Wood, R., Bony, S., Colman, R., Fichet, T., Fyfe, J., ... Taylor, K. (2007). Climate models and their evaluation. En S. Solomon y cols. (Eds.), *Climate change 2007: The physical science basis* (pp. 589–662). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Descargado de <https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-chapter8-1.pdf> (Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change)

- Reboita, M. S., Ferreira, G. W. d. S., Ribeiro, J. G. M., y Ali, S. (2024, JUN 1). Assessment of precipitation and near-surface temperature simulation by cmip6 models in south america. *ENVIRONMENTAL RESEARCH-CLIMATE*, 3(2). doi: 10.1088/2752-5295/ad3fdb
- Reboita, M. S., Ferreira, G. W. d. S., Ribeiro, J. G. M., da Rocha, R. P., y Rao, V. B. (2023). South american monsoon lifecycle projected by statistical downscaling with cmip6-gcms. *Atmosphere*, 14(9). Descargado de <https://www.mdpi.com/2073-4433/14/9/1380> doi: 10.3390/atmos14091380
- Riebesell, J. (2024). *Tikz collection*. Descargado 2025-03-09, de <https://github.com/janosh/tikz>
- Rivera, J. A., y Arnould, G. (2020, 9). Evaluation of the ability of cmip6 models to simulate precipitation over southwestern south america: Climatic features and long-term trends (1901–2014). *Atmospheric Research*, 241. doi: 10.1016/j.atmosres.2020.104953
- Rohat, G. (2018). Projecting drivers of human vulnerability under the shared socioeconomic pathways. *International Journal of Environmental Research and Public Health*, 15(3). Descargado de <https://www.mdpi.com/1660-4601/15/3/554> doi: 10.3390/ijerph15030554
- Rosopa, P., Schaffer, M., y Schroeder, A. (2013, 09). Managing heteroscedasticity in general linear models. *Psychological methods*, 18, 335-51. doi: 10.1037/a0032553
- Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., y Perera, B. J. (2018, 11). Statistical downscaling of precipitation using machine learning techniques. *Atmospheric Research*, 212, 240-258. doi: 10.1016/j.atmosres.2018.05.022
- Salman, A. G., Heryadi, Y., Abdurahman, E., y Suparta, W. (2018). Single layer & multi-layer long short-term memory (lstm) model with intermediate variables for weather forecasting. *Procedia Computer Science*, 135, 89-98. Descargado de <https://www.sciencedirect.com/science/article/pii/S187705091831439X> (The 3rd International Conference on Computer Science and Computational Intelligence (ICSCSI 2018) : Empowering Smart Technology in Digital Era for a Better Life) doi: <https://doi.org/10.1016/j.procs.2018.08.153>
- Sanfilippo, K., Timm, O. E., Frazier, A. G., y Giambelluca, T. W. (2024, FEB). Effects of systematic predictor selection for statistical downscaling of rainfall in hawai'i. *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, 44(2), 571-591. doi: 10.1002/joc.8345
- Saraf, V. R., y Regulwar, D. G. (2016). Assessment of climate change for precipitation and temperature using statistical downscaling methods in upper godavari river basin, india. *Journal of Water Resource and Protection*, 08, 31-45. doi: 10.4236/jwarp.2016.81004

- Sathe, A. M., Upadhye, N. S., y Agnieszka. (2023, 6). Forecasting of symmetric stable autoregressive models by time series approach supported by artificial neural networks. *Journal of Computational and Applied Mathematics*, 425. doi: 10.1016/j.cam.2022.115051
- Schoof, J., y Pryor, S. (2001, JUN 15). Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, 21(7), 773-790. doi: 10.1002/joc.655
- Sebbar, B.-e., Khabba, S., Merlin, O., Simonneaux, V., Hachimi, C. E., Kharrou, M. H., y Chehbouni, A. (2023). Machine-learning-based downscaling of hourly era5-land air temperature over mountainous regions. *Atmosphere*, 14(4). Descargado de <https://www.mdpi.com/2073-4433/14/4/610> doi: 10.3390/atmos14040610
- Shangguan, Y., Min, X., y Shi, Z. (2023). Inter-comparison and integration of different soil moisture downscaling methods over the qinghai-tibet plateau. *Journal of Hydrology*, 617, 129014. Descargado de <https://www.sciencedirect.com/science/article/pii/S0022169422015840> doi: <https://doi.org/10.1016/j.jhydrol.2022.129014>
- Shrestha, N. K., Akhtar, T., Ghimire, U., Rudra, R. P., Goel, P. K., Shukla, R., y Daggupati, P. (2021). Can-glws: Canadian great lakes weather service for the soil and water assessment tool (swat) modelling. *Journal of Great Lakes Research*, 47(1), 242-251. Descargado de <https://www.sciencedirect.com/science/article/pii/S038013302030246X> doi: <https://doi.org/10.1016/j.jglr.2020.10.009>
- Sun, L., Lan, Y., y Jiang, R. (2023). Using cnn framework to improve multi-gcm ensemble predictions of monthly precipitation at local areas: An application over china and comparison with other methods. *Journal of Hydrology*, 623, 129866. Descargado de <https://www.sciencedirect.com/science/article/pii/S0022169423008089> doi: <https://doi.org/10.1016/j.jhydrol.2023.129866>
- Tadić, L., Bonacci, O., y Brleković, T. (2019, 10). An example of principal component analysis application on climate change assessment. *Theoretical and Applied Climatology*, 138. doi: 10.1007/s00704-019-02887-9
- Talhar, J. A., Ramgirkar, A. G., y Jena, P. K. (2022, 4). Time series analysis and forecasting of precipitation (rainfall) data. *ECS Transactions*, 107, 1271-1307. doi: 10.1149/10701.1271ecst
- Team, R. C., y contributors. (2024). React documentation [Manual de software informático]. Descargado de <https://react.dev/learn> (Version 18.2)

- Teegavarapu, R. S., V, y Goly, A. (2018, APR). Optimal selection of predictor variables in statistical downscaling models of precipitation. *WATER RESOURCES MANAGEMENT*, 32(6), 1969-1992. doi: 10.1007/s11269-017-1887-z
- Tian, B., y Dong, X. (2020, 4). The double-itsz bias in cmip3, cmip5, and cmip6 models based on annual mean precipitation. *Geophysical Research Letters*, 47. doi: 10.1029/2020GL087232
- Toigo, A. D., Lee, C., Newman, C. E., y Richardson, M. I. (2012, 9). The impact of resolution on the dynamics of the martian global atmosphere: Varying resolution studies with the marswrf gcm. *Icarus*, 221, 276-288. doi: 10.1016/j.icarus.2012.07.020
- T R, M., Thakur, A., Sinha, D., Mishra, K., Kumar, V. V., y Guluwadi, S. (2024, 01). Transformative breast cancer diagnosis using cnns with optimized reducelronplateau and early stopping enhancements. *International Journal of Computational Intelligence Systems*, 17. doi: 10.1007/s44196-023-00397-1
- Tran Anh, D., Pham, S., Dang, T., y Hoang, L. (2019, 04). Downscaling rainfall using deep learning long short-term memory and feedforward neural network. *International Journal of Climatology*, 39, 4170-4188. doi: 10.1002/joc.6066
- Trojer, H. (2017, Dec.). Meteorología y climatología de la vertiente del pacífico colombiano. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, 41(Suplemento), 467-490. Descargado de <https://raccefyn.co/index.php/raccefyn/article/view/583> doi: 10.18257/raccefyn.583
- Turco, M., Llasat, M. C., Herrera, S., y Gutiérrez, J. M. (2017). Bias correction and downscaling of future rcm precipitation projections using a mos-analog technique. *Journal of Geophysical Research*, 122, 2631-2648. doi: 10.1002/2016JD025724
- Unidata. (2024, 01). Netcdf version 4.9.1. boulder, CO: UCAR/Unidata Program Center. , 0, 0. Descargado de <http://doi.org/10.5065/D6RN35XM> doi: <http://doi.org/10.5065/D6RN35XM>
- Urrea, V., Ochoa, A., y Mesa, O. (2019). Seasonality of Rainfall in Colombia. *Water Resources Research*, 55(5), 4149-4162. Descargado de <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023316> doi: <https://doi.org/10.1029/2018WR023316>
- V, P., P, V., K, R. K., P, S., Khan, O., y Krishna, C. N. (2022). A django web application to promote local service providers. En *2022 6th international conference on computing methodologies and communication (iccmc)* (p. 1517-1521). doi: 10.1109/ICCMC53470.2022.9754099

- Van der Meer, M., de Roda Husman, S., y Lhermitte, S. (2023, JUN). Deep learning regional climate model emulators: A comparison of two downscaling training frameworks. *JOURNAL OF ADVANCES IN MODELING EARTH SYSTEMS*, 15(6). doi: 10.1029/2022MS003593
- Vincent, W. S. (2020). *Django for beginners: Build websites with python and django*. WelcomeToCode.
- Wang, B., Zheng, L., Liu, D. L., Ji, F., Clark, A., y Yu, Q. (2018). Using multi-model ensembles of cmip5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in australia. *International Journal of Climatology*, 38(13), 4891-4902. Descargado de <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5705> doi: <https://doi.org/10.1002/joc.5705>
- Wang, L., Li, Y., Li, M., Li, L., Liu, F., Liu, D. L., y Pulatov, B. (2022, 12). Projection of precipitation extremes in china's mainland based on the statistical downscaled data from 27 gcms in cmip6. *Atmospheric Research*, 280. doi: 10.1016/j.atmosres.2022.106462
- Wang, W., Hong, T., Xu, X., Chen, J., Liu, Z., y Xu, N. (2019). Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm. *Applied Energy*, 248, 217-230. Descargado de <https://www.sciencedirect.com/science/article/pii/S0306261919307494> doi: <https://doi.org/10.1016/j.apenergy.2019.04.085>
- Wang, Z., Wu, D., Chen, X., y Qiao, R. (2013). Enso indices and analyses. *Advances in Atmospheric Sciences*, 30(5), 1491-1506. Descargado de <https://doi.org/10.1007/s00376-012-2238-x> doi: 10.1007/s00376-012-2238-x
- Williams, D. N., Balaji, V., Cinquini, L., Denvil, S., Duffy, D., Evans, B., ... Trenham, C. (2016). The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*, 61, 213-224. doi: 10.1016/j.future.2015.07.024
- Xie, X., Li, A., Tian, J., Wu, C., y Jin, H. (2023, 1). A fine spatial resolution estimation scheme for large-scale gross primary productivity (gpp) in mountain ecosystems by integrating an eco-hydrological model with the combination of linear and non-linear downscaling processes. *Journal of Hydrology*, 616. doi: 10.1016/j.jhydrol.2022.128833
- Xu, T., Guan, K., Peng, B., Wei, S., y Zhao, L. (2021). Machine learning-based modeling of spatio-temporally varying responses of rainfed corn yield to climate, soil, and management in the u.s. corn belt. *Frontiers in Artificial Intelligence*, 4. Descargado de <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2021.647999> doi: 10.3389/frai.2021.647999

- Xu, W., Zhang, Z., Long, Z., y Qin, Q. (2021). Downscaling smap soil moisture products with convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 4051-4062. doi: 10.1109/JSTARS.2021.3069774
- Yan, X., Chen, H., Tian, B., Sheng, S., Wang, J., y Kim, J.-S. (2021). A downscaling–merging scheme for improving daily spatial precipitation estimates based on random forest and cokriging. *Remote Sensing*, 13(11). Descargado de <https://www.mdpi.com/2072-4292/13/11/2040> doi: 10.3390/rs13112040
- Yang, H., Yan, J., Liu, Y., y Song, Z. (2022). Statistical downscaling of numerical weather prediction based on convolutional neural networks. *Global Energy Interconnection*, 5(2), 217-225. Descargado de <https://www.sciencedirect.com/science/article/pii/S209651172200038X> doi: <https://doi.org/10.1016/j.gloi.2022.04.018>
- Yang, Q., Zheng, J., y Liu, Z. (2016). Which one is the best: The interpolation methods comparison on average annual temperature in xinjiang from 1951 to 2013. En *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (p. 4466-4469). doi: 10.1109/IGARSS.2016.7730164
- Yang, X., Li, Y. P., Huang, G. H., Li, Y. F., Liu, Y. R., y Zhou, X. (2022, 5). Development of a multi-gcms bayesian copula method for assessing multivariate drought risk under climate change: A case study of the aral sea basin. *Catena*, 212. doi: 10.1016/j.catena.2022.106048
- Yazdian, H., Salmani-Dehaghi, N., y Alijanian, M. (2023, NOV). A spatially promoted svm model for grace downscaling: Using ground and satellite-based datasets. *JOURNAL OF HYDROLOGY*, 626(A). doi: 10.1016/j.jhydrol.2023.130214
- Yin, S., Liang, G., Wang, C., y Zhou, Z. (2022, 12). Asynchronous seasonal patterns of soil microorganisms and plants across biomes: A global synthesis. *Soil Biology and Biochemistry*, 175. doi: 10.1016/j.soilbio.2022.108859
- Yuan, L., Zhang, X. C. J., Busteed, P., y Flanagan, D. C. (2022, 7). Simulating the potential effects of elevated co2 concentration and temperature coupled with storm intensification on crop yield, surface runoff, and soil loss based on 25 gcms ensemble: A site-specific case study in oklahoma. *Catena*, 214. doi: 10.1016/j.catena.2022.106251
- Yudin, A. (2020, 01). Building versatile mobile apps with python and rest, restful web services with django and react. *Apress; 1st edition*. doi: 10.1007/978-1-4842-6333-4
- Zelle, J. (2016). *Python programming: An introduction to computer science* (3.^a ed.). Franklin, Beedle & Associates.

- Zhang, C., Yang, Z., Zhao, H., Sun, Z., Di, L., Bindlish, R., ... Yueh, S. H. (2022). Crop-casma: A web geoprocessing and map service based architecture and implementation for serving soil moisture and crop vegetation condition data over u.s. cropland. *International Journal of Applied Earth Observation and Geoinformation*, 112, 102902. Descargado de <https://www.sciencedirect.com/science/article/pii/S1569843222001042> doi: <https://doi.org/10.1016/j.jag.2022.102902>
- Zhang, H., Immerzeel, W. W., Zhang, F., de Kok, R. J., Gorrie, S. J., y Ye, M. (2021, 5). Creating 1-km long-term (1980–2014) daily average air temperatures over the tibetan plateau by integrating eight types of reanalysis and land data assimilation products downscaled with modis-estimated temperature lapse rates based on machine learning. *International Journal of Applied Earth Observation and Geoinformation*, 97. doi: 10.1016/j.jag.2021.102295
- Zhang, X., Huang, T., Gulakhmadov, A., Song, Y., Gu, X., Zeng, J., ... Niyogi, D. (2022, 8). Deep learning-based 500 m spatio-temporally continuous air temperature generation by fusing multi-source data. *Remote Sensing*, 14. doi: 10.3390/rs14153536
- Zhang, Y., Feng, M., Zhang, W., Wang, H., y Wang, P. (2021, JUL). A gaussian process regression-based sea surface temperature interpolation algorithm. *JOURNAL OF OCEANOLOGY AND LIMNOLOGY*, 39(4), 1211-1221. doi: 10.1007/s00343-020-0062-1
- Zhao, N., y Wu, X. (2023). Improving daily precipitation estimations in a high mountainous watershed by developing a new downscaling method with spatially varying coefficients. *Journal of Hydrology*, 626, 130367. Descargado de <https://www.sciencedirect.com/science/article/pii/S0022169423013094> doi: <https://doi.org/10.1016/j.jhydrol.2023.130367>
- Zhou, H., Zhou, W., Liu, Y., Huang, J., Yuan, Y., y Liu, Y. (2023, 2). Climatological spatial scales of meteorological droughts in china and associated climate variability. *Journal of Hydrology*, 617. doi: 10.1016/j.jhydrol.2022.129056