Universidad de Cuenca

Facultad de Ingeniería

Carrera de Ingeniería Civil

IMPACTO DEL CAMBIO CLIMÁTICO EN LAS PRINCIPALES FUENTES DE AGUA POTABLE DE LA PROVINCIA DEL AZUAY

Trabajo de titulación previo a la obtención del título de Ingeniero Civil

Autores:

Nicolás Santiago Arias Valdez

Adrián Felipe Peláez Molina

Director:

Diego Esteban Mora Serrano

ORCID: 0000-0002-1139-3204

Tutores:

Daniel Emilio Mendoza Sigüenza

ORCID: 0000-0002-2117-2587

Esteban Patricio Samaniego Alvarado

ORCID: 0000-0002-8728-491X

Cuenca, Ecuador

2025-03-26

Resumen

Este estudio analiza el impacto del cambio climático en las fuentes de agua potable de la provincia del Azuay, Ecuador, con énfasis en la cuenca del río Tomebamba, fuente clave de agua para la ciudad de Cuenca. Se utilizaron datos históricos y satelitales de precipitación, aplicando regresión Lasso y redes neuronales para predecir las tendencias futuras de precipitación y su efecto en el caudal del río. Además, se empleó un enfoque de downscaling para afinar las proyecciones de modelos climáticos globales a una escala local. Los datos de precipitación históricos (1960-1990) fueron validados mediante curvas de doble masa, y la regresión Lasso permitió identificar las variables más relevantes. Las redes neuronales se usaron para modelar y prever la precipitación futura, evaluando la calidad de los modelos con métricas de error como RMSE y el coeficiente de eficiencia de Nash-Sutcliffe. Los resultados muestran un aumento significativo en la variabilidad de la precipitación, con picos más extremos, lo que sugiere mayor frecuencia de eventos extremos como lluvias intensas y sequías. El modelo climático Germany AWI-CM-1-1-MR, refinado con downscaling, se destacó como el más adecuado para la región. Las proyecciones indican una disminución del caudal del río, lo que podría afectar el suministro de agua potable. Se recomienda usar este modelo en futuros estudios y considerar escenarios críticos para una mejor gestión de los recursos hídricos.

Palabras clave del autor: downscaling, modelación, recursos hídricos

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Cuenca ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por la propiedad intelectual y los derechos de autor.

Repositorio Institucional: https://dspace.ucuenca.edu.ec/

Abstract

This study analyzes the impact of climate change on potable water sources in Azuay Province, Ecuador, focusing on the Tomebamba River basin, a key water source for the city of Cuenca. Historical and satellite precipitation data were used, applying Lasso regression and neural networks to predict future precipitation trends and their impact on river flow. A downscaling approach was employed to refine global climate model projections to a local scale. Historical precipitation data (1960-1990) were validated using double mass curves, and Lasso regression helped identify the most relevant variables. Neural networks were used to model and forecast future precipitation, evaluating model quality with error metrics such as RMSE and the Nash-Sutcliffe efficiency coefficient. Results show a significant increase in precipitation variability, with more extreme peaks, suggesting a higher frequency of extreme events such as heavy rains and droughts. The Germany AWI-CM-1-1-MR climate model, refined through downscaling, was identified as the most suitable for the region. Projections indicate a decrease in the Tomebamba River's flow, which could affect the potable water supply. It is recommended to use this model for future studies and consider critical scenarios for better water resource management.

Author keywords: downscaling, modeling, water resources

The content of this work corresponds to the right of expression of the authors and does not compromise the institutional thinking of the University of Cuenca, nor does it release its responsibility before third parties. The authors assume responsibility for the intellectual property and copyrights.

Institutional Repository: https://dspace.ucuenca.edu.ec/

Índice de contenido

1 INTRODUCCIÓN 9
1.1 Antecedentes 10
1.2 Justificación 11
2 Objetivos 11
2.1.1 Objetivo General 11
2.1.2 Objetivos Específicos 11
3 MATERIALES 11
3.1 Área de Estudio 11
3.2 Datos de Observación 12
3.3 Datos Proyectados 14
4 METODOLOGÍA 16
4.1 Regresión Lineal Múltiple 16
4.2 Interacción de Datos 17
4.3 Identificación de variables significativas mediante LASSO 18
4.4 Modelo Climático: Redes Neuronales “Deep Neuronal Network” 18
4.4.1 Modelo Climático: Redes Neuronales (Calibración) 20
4.5 Modelo Hidrológico Autorregresivo Basado en Datos: 22
4.6 Métricas de Error 23
4.6.1 Nash – Sutcliffe y Root Mean Square Error (RMSE) 23
4.7 Determinación del Mejor GCM. 24
5 RESULTADOS 25
5.1 Control de Calidad de Datos 25
5.2 Regresión Lineal Múltiple 25
5.3 Interacción de Datos 26
5.4 Identificación de variables significativas mediante LASSO 27
5.5 Redes Neuronales 28
5.5.1 Redes Neuronales: Calibración 28
5.5.2 Redes Neuronales: Validación 29
5.6 Determinación del mejor GCM 31
5.7 Redes Neuronales: Predicción de la precipitación. 32
5.8 Modelo Hidrológico Autorregresivo Basado en Datos 35
5.8.1 Modelo Hidrológico Autorregresivo Basado en Datos: Calibración 35
5.8.2 Modelo Hidrológico Autorregresivo Basado en Datos: Validación 37
5.8.3 Modelo Hidrológico Autorregresivo Basado en Datos: Aplicación 38

6 CONCLUSIONES Y RECOMENDACIONES 40
6.1 Conclusiones 40
6.2 Observaciones: 41
6.3 Recomendaciones: 41
Anexos 46

Índice de figuras

Figura 1 Cuenca del Tomebamba. Las estaciones pluviométricas están distribuidas a lo largo de la cuenca 13
Figura 2 Puntos de precipitación del modelo Germany AWI-CM-1-1-MR ubicados con respecto a la cuenca del Tomebamba 16
Figura 3 Esquema de una Red Neuronal (Dinamarca, 2018) 19
Figura 4 Red Neuronal previa al entrenamiento. Usa CESM2 21
Figura 5 Red Neuronal después del entrenamiento. Usa CESM2 21
Figura 6 Curvas de Doble Masa. Estaciones Buenos Aires - Surucucho 25
Figura 7 Uso de regresión lineal múltiple para predecir precipitación. 26
Figura 8 Lambda óptimo en el mínimo RMSE del modelo Europeo 28
Figura 9 Conjunto de Evaluación. Italy CMCC-ESM2 30
Figura 10 Conjunto de Evaluación. Se observa que las variables sin interacción presentan una gran dispersión con respecto a los valores observados 31
Figura 11 Conjunto de Evaluación. Se observa que las variables con interacción reducen significativamente la dispersión con respecto a los valores observados 32
Figura 12 Red Neuronal. Precipitaciones a lo largo de los 30 años del conjunto de calibración - validación 32
Figura 13 Precipitación (mm) predicha bajo el escenario Germany AWI-CM-1-1-MR SSP-245 33
Figura 14 Precipitación (mm) predicha bajo el escenario Germany AWI-CM-1-1-MR SSP-585 33
Figura 15 BoxPlot de precipitaciones de los modelos SSP 2 - 4.5 y SSP 5 - 8.5 comparado con las precipitaciones históricas 34
Figura 16 Comparación de la curva de ajuste “a” entre el dado por MATLAB y el ajustado $aQt-1$ 36
Figura 17 Comparación de la curva de ajuste “b” entre el dado por MATLAB y el ajustado $bQt-1$ 36
Figura 18 Modelo Hidrológico Autorregresivo en función del caudal anterior y precipitación efectiva 37
Figura 19 Validación de Modelo Hidrológico DBM 37
Figura 20 Predicción de Caudales en los distintos escenarios SSP 38
Figura 21 BoxPlot de caudales de los modelos SSP 2 - 4.5 y SSP 5 - 8.5 comparado con los caudales históricos 39

Índice de tablas

Tabla 1 Estaciones meteorológicas usadas 13
Tabla 2 Estación hidrológica usada 14
Tabla 3 GCM usados y número de corridas por modelo para la precipitación futura mensual. 16
Tabla 4 Parámetros de configuración de la Red Neuronal 21
Tabla 5 Creación de nuevas variables mediante PolynomialFeatures 27
Tabla 6 Predictores seleccionados por LASSO 28
Tabla 7 Resultados estadísticos conjunto entrenamiento. Modelos sin Interacción 29
Tabla 8 Resultados estadísticos conjunto entrenamiento. Modelos con Interacción 29
Tabla 9 Estadísticos del conjunto de validación. Variables sin Interacción 30
Tabla 10 Estadísticos del conjunto de validación. Variables con Interacción 31

Índice de ecuaciones

Eq 1 Regresión Lineal Múltiple 17
Eq 2 LASSO 18
Eq 3 Red Neuronal 20
Eq 4 Función de Activación 20
Eq 5 Modelo Autorregresivo 22
Eq 6 Ecuación de Young 23
Eq 7 Modelo Hidrológico 23
Eq 8 RMSE 24
Eq 9 Nash - Sutcliffe 24
Eq 10 Curva de Ajuste “a” 35
Eq 11 Curva de Ajuste “b” 35

1 INTRODUCCIÓN

El cambio climático es uno de los mayores desafíos que enfrenta la humanidad en el siglo XXI. Sus efectos sobre el ciclo hidrológico son especialmente preocupantes, ya que pueden alterar la disponibilidad y calidad de los recursos hídricos, impactando negativamente en la seguridad hídrica de las comunidades. En la provincia del Azuay, Ecuador, donde las fuentes de agua potable son cruciales para el desarrollo económico y social, es esencial entender cómo el cambio climático está afectando estas fuentes para poder diseñar estrategias de adaptación adecuadas.

El presente estudio tiene como objetivo determinar y evaluar la influencia directa del cambio climático en las fuentes hídricas y sus principales impactos en el suministro y uso del agua. Para alcanzar este objetivo, se han obtenido datos históricos de precipitación medidos en la cuenca y datos satelitales de precipitación correspondientes a los mismos años. Estos datos fueron verificados mediante curvas de doble masa, y se utilizó la regresión Lasso para identificar qué datos satelitales representaban mejor el comportamiento de la lluvia en la zona. Posteriormente, se realizó una interacción entre los datos satelitales para generar nuevas variables y se aplicó nuevamente la regresión Lasso para obtener los datos más significativos. Finalmente, estos datos se utilizaron para entrenar un modelo de red neuronal con el fin de predecir la precipitación futura.

El cambio climático puede afectar significativamente la disponibilidad y calidad del agua en embalses al alterar el régimen hidrológico de las cuencas (Costabile et al., 2023). Por ello, es fundamental combinar modelos hidrológicos con proyecciones climáticas regionales para evaluar estos impactos de manera precisa. En este estudio se adoptó un enfoque similar, utilizando modelos de aprendizaje automático y técnicas avanzadas de modelado hidrológico, sin embargo, sigue siendo un estudio exploratorio.

El uso de técnicas avanzadas de aprendizaje automático, como las redes neuronales, ha mejorado significativamente la capacidad predictiva de los modelos hidrológicos. Estas técnicas permiten manejar grandes cantidades de datos y extraer patrones complejos que los modelos tradicionales podrían pasar por alto (Kumar et al., 2023). En particular, la regresión Lasso ha demostrado ser eficaz para seleccionar las variables más relevantes en grandes conjuntos de datos climáticos y satelitales, mejorando la precisión de las predicciones.

Además, la evaluación de técnicas de aprendizaje automático para modelar la sequía hidrológica ha demostrado ser crucial para la gestión a largo plazo de los recursos hídricos

en regiones afectadas por el cambio climático (Achite et al., 2022). Estos enfoques proporcionan herramientas valiosas para anticipar y mitigar los efectos adversos del cambio climático en el suministro de agua. En nuestro estudio, hemos aplicado estas técnicas para obtener predicciones más precisas de la precipitación futura, lo que es esencial para la planificación y gestión de los recursos hídricos en la provincia del Azuay.

El realizar un downscaling con el objetivo de proyectar y modelar los datos climáticos a una escala más local y específica es fundamental para identificar los Modelos Climáticos Globales (GCM) que pueden ser aplicados de manera efectiva a nivel local, asegurando que las proyecciones sean válidas para la región cuenca del Tomebamba.

Este trabajo no solo busca entender los impactos directos del cambio climático en la precipitación, sino también evaluar cómo estos cambios pueden afectar la disponibilidad de agua potable en la región. Los resultados obtenidos hasta ahora proporcionan una base sólida para los siguientes pasos del estudio, que incluyen la validación de los modelos climáticos y la calibración de los modelos hidrológicos.

1.1 Antecedentes

El cambio climático representa una grave amenaza global, con impactos especialmente significativos en regiones vulnerables como América del Sur. Los factores socioeconómicos, como la pobreza, la desigualdad y el acceso limitado a servicios esenciales, junto con la distribución desigual de los recursos hídricos, aumentan la vulnerabilidad de esta región a los efectos del cambio climático (CAF, 2014). La falta de investigaciones y estudios específicos que analicen el impacto de este fenómeno en nuestras principales fuentes de agua subraya la urgente necesidad de abordar este tema en nuestra región.

Algunos estudios han examinado los impactos del cambio climático en el sistema hidrológico del Ecuador. Sin embargo, estos estudios han revelado limitaciones, como el enfoque restringido en la región del páramo sin considerar los impactos directos en áreas urbanas, así como el rendimiento deficiente de los modelos hidrológicos utilizados debido a la escasez de datos y estimaciones de incertidumbre incompletas o faltantes sobre los impactos hidrológicos (Núñez Mejía et al., 2023).

Actualmente, nos enfrentamos a desafíos relacionados con la insuficiencia de caudal en las captaciones de la región. Por lo tanto, resulta relevante llevar a cabo un análisis mediante un modelo basado en datos que incorpore fenómenos físicos, una metodología que aún no se ha utilizado en este contexto. El propósito de este análisis sería obtener parámetros

hidrológicos futuros que puedan servir para generar alertas tempranas en caso de ser necesarias (Mendoza et al., 2019).

1.2 Justificación

La provincia del Azuay se caracteriza por su diversidad geográfica y climática, lo que influye en la disponibilidad de sus recursos hídricos. El cambio climático puede alterar significativamente los patrones de precipitación, afectando la disponibilidad de agua en las cuencas hidrográficas locales. Comprender estos cambios es esencial para garantizar un suministro de agua seguro y sostenible. Este estudio se justifica por la necesidad de contar con información precisa y detallada sobre cómo el cambio climático está afectando las fuentes de agua potable en la provincia del Azuay, lo que permitirá diseñar estrategias de adaptación efectivas.

2 Objetivos

2.1.1 Objetivo General

Determinar y evaluar la influencia directa del cambio climático en las fuentes hídricas y sus principales impactos en el suministro y uso del agua.

2.1.2 Objetivos Específicos

- Implementar un modelo hidrológico basado en datos para analizar el comportamiento de la fuente hídrica ante las variaciones climáticas.
- Implementar un modelo climático basado en datos para proyectar los posibles escenarios de cambio climático que podrían afectar las fuentes hídricas.
- Realizar un análisis estadístico de las variables usadas para comprender mejor las tendencias y patrones.

3 MATERIALES

3.1 Área de Estudio

La provincia del Azuay, ubicada en el sur de Ecuador, es una región caracterizada por una topografía montañosa y un clima variado que va desde templado a frío debido a la influencia de la cordillera de los Andes. Esta provincia es especialmente relevante para el estudio del impacto del cambio climático en las fuentes de agua potable debido a su dependencia de los recursos hídricos provenientes de las cuencas hidrográficas locales.

La capital de la provincia, Cuenca, es la tercera ciudad más grande de Ecuador y se abastece principalmente de los ríos Tomebamba, Yanuncay, Tarqui y Machángara, que nacen en el Parque Nacional Cajas y sus alrededores. Estas fuentes de agua son cruciales no solo para el consumo humano, sino también para la agricultura y la industria local (Ferreira et al., 2023).

La altitud en la provincia varía considerablemente, desde los 1,800 metros en los valles hasta más de 4,000 metros en las montañas, lo que contribuye a una variedad de microclimas y ecosistemas (Ruiz-Ordóñez et al., 2023).

El Parque Nacional Cajas, una de las principales áreas protegidas de la provincia, es conocido por sus lagunas glaciares y páramos, que actúan como importantes reservorios de agua. Este parque es una zona de captación clave que contribuye significativamente al suministro de agua potable para Cuenca y otras comunidades cercanas. La biodiversidad y la capacidad de almacenamiento de agua del parque lo hacen especialmente vulnerable a los efectos del cambio climático, como cambios en los patrones de precipitación y temperatura (Wingfield et al., 2021)

La región experimenta una temporada de lluvias bien definida que va de octubre a mayo, y una temporada seca de junio a septiembre. Sin embargo, las tendencias recientes indican una variabilidad climática creciente, con eventos extremos como lluvias intensas y sequías prolongadas volviéndose más comunes (Ferreira et al., 2023). Estas variaciones climáticas tienen el potencial de afectar negativamente la disponibilidad y calidad del agua, poniendo en riesgo la seguridad hídrica de la población.

Además, la expansión urbana y el crecimiento de la población en Cuenca han aumentado la demanda de agua potable, haciendo aún más crucial la gestión sostenible de los recursos hídricos. La combinación de factores climáticos y antropogénicos subraya la necesidad de estudios detallados y modelos predictivos que puedan informar estrategias de adaptación y mitigación (Ruiz-Ordóñez et al., 2023).

El área de estudio en la provincia del Azuay presenta un escenario complejo y dinámico para la investigación del impacto del cambio climático en las fuentes de agua potable. La diversidad geográfica y climática, junto con la importancia de los recursos hídricos para la región, hacen de este un estudio relevante para el desarrollo sostenible y la seguridad hídrica.

3.2 Datos de Observación

La cuenca de estudio para este trabajo es la Cuenca del Tomebamba, ubicada en la provincia del Azuay, Ecuador. Esta cuenca es de gran importancia para la ciudad de Cuenca y sus alrededores debido a su papel crucial en el suministro de agua potable.

Para obtener los datos de precipitación necesarios para este estudio, se seleccionaron cuatro estaciones meteorológicas ubicadas dentro de la cuenca del Tomebamba, dadas por los datos del INAMHI. Estos datos de precipitación corresponden a una escala mensual. Fueron completados utilizando regresiones lineales múltiples y un método de perturbación de

cuantiles aplicado a un relleno de datos tradicional (Mora et al., 2014). Las estaciones seleccionadas son: Buenos Aires, Sayausí, Surucucho y Quinoas. Las estaciones meteorológicas, mostradas en la figura 1, proporcionaron datos históricos de precipitación desde enero de 1964 hasta diciembre de 1993, cumpliendo con el estándar de al menos 30 años de observaciones. La ubicación de estas estaciones se detalla en la tabla 1.

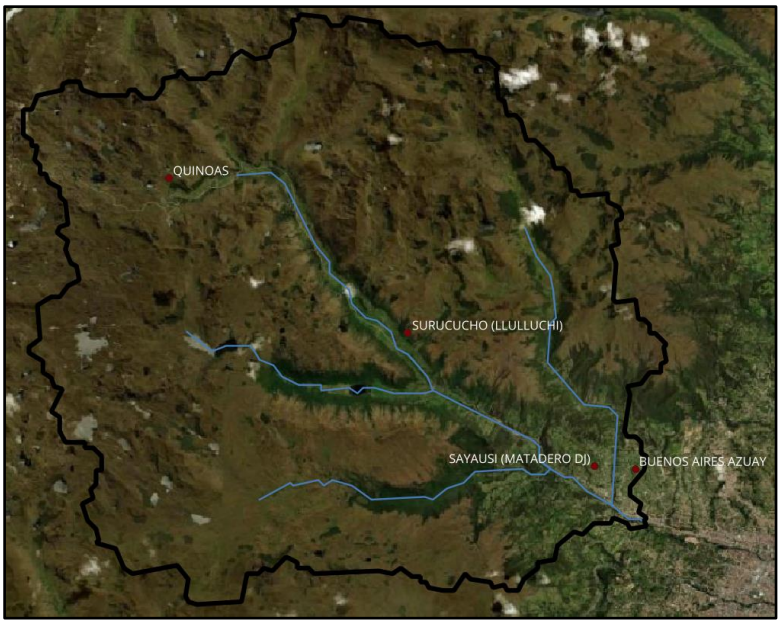


Figura 1
Cuenca del Tomebamba. Las estaciones pluviométricas están distribuidas a lo largo de la cuenca

Tabla 1
Estaciones meteorológicas usadas

Estación	Latitud	Longitud
Buenos Aires	-2.867	-79.064
Quinoas	-2.780	-79.203
Sayausí	-2.866	-79.076
Surucucho	-2.826	-79.132

La precipitación media en la cuenca se determinó utilizando el método de polígonos de Thiessen. Este método permite asignar un peso específico a cada estación meteorológica según su área de influencia, facilitando así la obtención de una estimación más precisa de la precipitación media en toda la cuenca (Thiessen, 1911).

Los datos de precipitación proporcionados por estas estaciones fueron sometidos a un riguroso proceso de verificación para asegurar su exactitud y confiabilidad. Para esto, se aplicó la técnica de curvas de doble masa. Este método es ampliamente utilizado en estudios hidrológicos para detectar inconsistencias y evaluar la homogeneidad de los datos de

precipitación. Las curvas de doble masa comparan la acumulación de precipitación en una estación específica con la acumulación promedio de un grupo de estaciones de referencia (Searcy et al., s/f).

Los datos de caudales usados en este trabajo fueron dados por una estación hidrológica ubicada en la cuenca del Tomebamba. Estos datos de precipitación corresponden a una escala mensual. La ubicación de la estación se presenta en la tabla 2

Tabla 2
Estación hidrológica usada

Estación	Latitud	Longitud
Tomebamba	-2.868	-79.086

3.3 Datos Proyectados

El proyecto Copernicus es el programa de observación de la Tierra de la Unión Europea, diseñado para proporcionar información precisa y accesible sobre el medio ambiente y la seguridad. Gestionado por la Comisión Europea en colaboración con la Agencia Espacial Europea (ESA) y otras agencias, este programa recopila datos a través de una serie de satélites llamados Sentinels (*Sobre Copernicus* | *Copernicus*, s/f).

En este estudio, se ha utilizado el proyecto Copernicus para obtener precipitaciones, debido a que ofrece datos globales casi en tiempo real e históricos mediante observaciones in situ y por satélite. Además, se han empleado los modelos climáticos globales de última generación del Proyecto de Intercomparación de Modelos Acoplados, Fase 6 (CMIP6), que es la última fase disponible actualmente. Para los fines de este estudio, el período histórico se define como los años comprendidos entre 1960 y 1990, utilizando datos de precipitación mensual. La nomenclatura de los modelos climáticos se estructura de la siguiente manera: “SSP x – y”. En esta estructura:

- “x” representa el escenario socioeconómico, que describe diferentes posibles evoluciones de la sociedad.
- “y” indica el nivel de forzamiento radiactivo asociado con el escenario, que refleja la cantidad de emisiones esperadas. En esta escala: 8.5 es el valor asociado al escenario con el mayor forzamiento radiactivo, es decir, el escenario más extremo en términos de emisiones, 1.9 representa el escenario con el menor forzamiento radiactivo, siendo el más optimista en cuanto a las emisiones y su impacto.

Los SSP (Shared Socioeconomic Pathways) describen diferentes posibles evoluciones de la sociedad futura, considerando que no haya cambio climático ni políticas climáticas. Estos escenarios se dividen en varias categorías (O'neill et al., 2016):

SSP1 y SSP5:

Optimistas: Ambos escenarios anticipan un desarrollo humano positivo, con grandes inversiones en educación y salud, crecimiento económico rápido y buenas instituciones. Sin embargo, difieren en el enfoque hacia la energía:

- SSP1: Se enfoca en prácticas sostenibles y una transición hacia una economía más verde.
- SSP5: Asume una economía intensiva en energía y dependiente de combustibles fósiles.

SSP3 y SSP4:

Pesimistas: Estos escenarios presentan tendencias de desarrollo más negativas, con baja inversión en educación y salud, un rápido crecimiento poblacional y creciente desigualdad.

- SSP3: Los países priorizan la seguridad regional, lo que lleva a una alta vulnerabilidad al cambio climático.
- SSP4: Se caracteriza por grandes desigualdades tanto dentro como entre países, resultando en sociedades también muy vulnerables al cambio climático.

SSP2:

Central: Este escenario contempla que las tendencias continúen según los patrones históricos sin grandes desviaciones, representando una trayectoria intermedia.

Los escenarios técnicos examinan cómo podrían evolucionar en el futuro los factores antropogénicos que impactan el cambio climático, tales como las emisiones de gases de efecto invernadero, otros gases reactivos, aerosoles y el uso del suelo. Estos escenarios, que se fundamentan en posibles cambios socioeconómicos, son esenciales para entender y anticipar cómo nuestro clima podría cambiar en el futuro (O'neill et al., 2016). En nuestro caso, se han seleccionado cuatro modelos debido a las limitaciones y objetivos de este trabajo exploratorio. La elección de estos modelos se basa en dos criterios principales:

- Mantener las tendencias actuales: Se ha priorizado la selección de modelos que reflejan las tendencias actuales, especialmente el que asume una economía basada en combustibles fósiles, similar a la ecuatoriana.

- Resolución y disponibilidad: Se eligieron modelos que ofrecen la mejor resolución disponible para la región y que tienen la misma disponibilidad en términos de escenarios SSP.

En consecuencia, los escenarios utilizados son los detallados en la Tabla 3.

Tabla 3
GCM usados y número de corridas por modelo para la precipitación futura mensual.

No.	Modelo GCM	Resolución (Lat. x Lon.)	Histórico	SSP 2 - 4.5	SSP 5 - 8.5
1	Europe EC-Earth3-CC	0.7° x 0.7°	1	1	1
2	Germany AWI-CM-1-1-MR	0.9° x 0.9°	1	1	1
3	Italy CMCC-ESM2	0.9° x 1.25°	1	1	1
4	Usa CESM2	0.9° x 1.25°	1	1	1

En la figura 2 se muestra un ejemplo de la resolución y ubicación de los puntos de precipitación dados por el modelo.

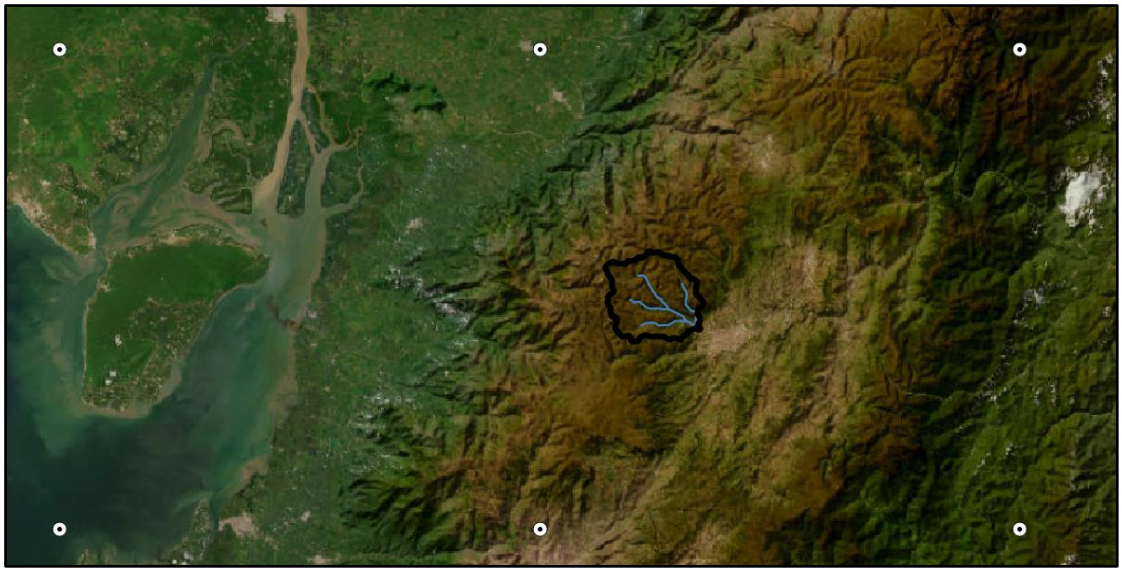


Figura 2
Puntos de precipitación del modelo Germany AWI-CM-1-1-MR ubicados con respecto a la cuenca del Tomebamba

4 METODOLOGÍA

4.1 Regresión Lineal Múltiple

En este estudio, se emplearon datos satelitales de precipitación junto con datos de precipitación media observada para realizar una regresión lineal múltiple. Este análisis tuvo como objetivo obtener coeficientes que permitieran predecir la precipitación en la región de estudio de manera más precisa considerando múltiples variables predictoras. Se optó por

este enfoque con el fin de evaluar cómo se comporta la predicción utilizando este método en comparación con otros enfoques de modelado que se detallan posteriormente.

La regresión lineal simple es una técnica estadística que modela la relación entre una variable dependiente y (precipitación observada) y varias variables independientes $x_1, x_2, \dots, x_n$(datos satelitales de precipitación). La ecuación general de la regresión lineal múltiple es (Farrar & Glauber, 1967):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \qquad \text{Eq 1}$$

Donde:

y es la variable dependiente (precipitación observada),
$x_1, x_2, \dots, x_n$ son las variables independientes (diferentes fuentes de datos satelitales y climáticos),
β_0 es el intercepto,
$\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes de regresión,
ϵ es el término de error

El análisis se realizó utilizando métodos estándar de mínimos cuadrados para ajustar los coeficientes $\beta_0, \beta_1, \dots, \beta_n$. La bondad del ajuste se evaluó mediante el coeficiente de determinación R^2, que indica la proporción de la variabilidad en la precipitación observada explicada por los datos satelitales (Draper & Smith, 2014).

4.2 Interacción de Datos

La idea de que existen diferentes formas de representar los predictores en un modelo, y que algunas de estas representaciones son más efectivas que otras, da lugar al concepto de ingeniería de características. Este proceso consiste en crear representaciones de los datos que aumenten la efectividad del modelo (Kuhn & Johnson, 2019).

La efectividad de un modelo está influenciada por varios factores. Por ejemplo, algunos modelos no pueden tolerar variables que están correlacionadas entre sí (multicolinealidad), muchos modelos no pueden trabajar con muestras que contengan valores faltantes, y algunos se ven gravemente afectados por la presencia de variables irrelevantes en los datos. Dado que en nuestro caso contamos con un número limitado de variables predictoras, recurrimos a la ingeniería de características para mejorar la precisión del modelo y optimizar los resultados.

La ingeniería de características (Feature Engineering) es una técnica que consiste en transformar los datos crudos en características que representen de manera más efectiva el problema en los modelos predictivos, lo cual mejora la precisión del modelo a desarrollar (Kuhn & Johnson, 2019).

Estos términos de interacción son útiles en modelos lineales, donde la relación modelada es una combinación lineal de las características de entrada. Al introducir productos de variables, el modelo puede capturar relaciones no lineales dentro de un marco lineal, permitiendo la creación de nuevas características a partir de la combinación de las existentes. La estructura de estos términos de interacción es la siguiente:

$$[a\ b] = [a\ \ b\ \ a^2\ \ ab\ \ b^2]$$

4.3 Identificación de variables significativas mediante LASSO

Es importante seleccionar cuales de los predictores son mejores, ya que los predictores innecesarios agregan ruido a la estimación (Lee & Singh, 2019):

Los parámetros de un modelo de regresión son calculados minimizando el costo de la función. Cuando la complejidad del modelo aumenta, el uso de este método puede conducir a un modelo con bajo sesgo y alta variación. Significa que, aunque el modelo puede ajustarse bien al conjunto de entrenamiento (sesgo bajo), predice muy mal los nuevos puntos de prueba (varianza alta). Debido a esto, los métodos de regularización son usados.

La regresión LASSO (Tibshiranit, 1996) es una de las técnicas de regularización. Al imponer una penalización L1 al costo total, el crecimiento del valor de los parámetros del modelo se puede controlar, por lo que se obtendría un modelo con menor varianza. LASSO brinda la posibilidad de establecer variables predictoras en cero (Ver eq. 2).

$$Costo\ total = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{s} |\beta_j| \qquad \text{Eq 2}$$

Se ha usado una herramienta gráfica para visualizar y seleccionar el valor óptimo del parámetro de regularización λ. La curva de RMSE versus λ muestra cómo cambia el RMSE del modelo en un conjunto de datos de validación o prueba a medida que se varía el valor de λ (Tibshiranit, 1996).

4.4 Modelo Climático: Redes Neuronales “Deep Neuronal Network”

Debido a su estructura y fundamentos, las redes neuronales artificiales presentan muchas similitudes con el funcionamiento del cerebro. Por ejemplo, son capaces de aprender de la

experiencia, generalizar a partir de casos anteriores para enfrentar nuevos desafíos, y abstraer características esenciales de entradas que pueden contener información irrelevante. (Matich Jorge, 2001). Estas capacidades hacen que las redes neuronales sean muy ventajosas y las convierten en una herramienta valiosa para diversas aplicaciones, incluida la predicción de precipitación. Las redes neuronales pueden aprender a realizar tareas específicas basadas en datos de entrenamiento o experiencias previas. Esto les permite adaptarse a patrones complejos en los datos de precipitación y mejorar su precisión a medida que se expone a más datos.

La arquitectura básica de una neurona artificial y su implementación en redes neuronales se describe detalladamente en el trabajo de (Dinamarca, 2018). Una neurona artificial funciona recibiendo varias entradas, representadas por un vector $x = (x_1, \dots, x_n)$ donde cada entrada x_i corresponde a la señal de salida de otras neuronas. A cada una de estas entradas se le asigna un peso específico $\omega_1, \dots, \omega_n$. Estas entradas ponderadas son combinadas en una suma y luego transformadas a través de una función de activación f no lineal, produciendo una salida y.

Se puede construir una red neuronal artificial mediante un conjunto de neuronas artificiales $f^1, \dots, f^k$ y conectando la salida de cada una de estas entradas a las entradas de otras diferentes, formando una estructura interconectada que permite el procesamiento complejo de información.

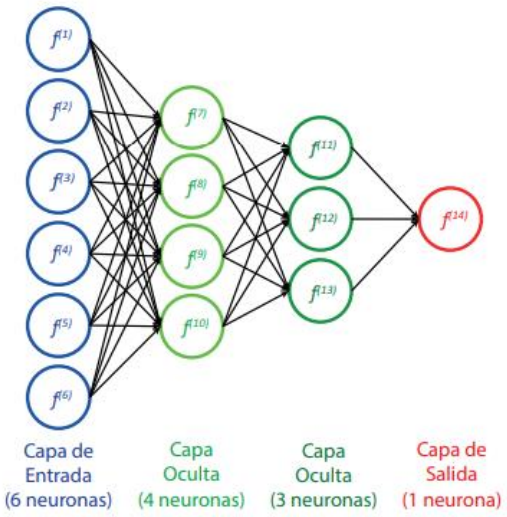


Figura 3
Esquema de una Red Neuronal (Dinamarca, 2018)

El modelo matemático viene dado por la eq 3:

$$y = f\left(\sum_{k=1}^{n} \omega_k x_k + b\right) \qquad \text{Eq 3}$$

Donde $\{x_1, \dots, x_n\}$ son las entradas (señales de salida de n neuronas distintas), $\{\omega_1, \dots, \omega_n\}$ son los respectivos pesos o ponderaciones, b es un sesgo (constante real), f es la función de activación o transferencia usualmente elegida de forma arbitraria, y es la salida producida por la neurona.

La función de activación usada en la red neuronal es la “Leaky Relu”, descrita en la eq 4:

$$f(z) = \begin{cases} z, & z > 0 \\ az, & z \leq 0 \end{cases} \qquad \text{Eq 4}$$

4.4.1 Modelo Climático: Redes Neuronales (Calibración)

El gradiente descendente se destaca como uno de los algoritmos más empleados para optimización, siendo además la metodología predominante para la optimización de DNN. Se ha usado Adam, ya que es eficiente computacionalmente (Goodfellow et al., 2016). En algunos casos, las redes neuronales pueden producir un fenómeno conocido como sobreajuste. El sobreajuste ocurre cuando una red neuronal es extremadamente precisa al predecir los datos con las que fue entrenada, pero tiene un rendimiento deficiente al intentar predecir datos nuevos que no ha visto anteriormente. En este caso, se ha usado el Early Stopping. Esta técnica implica monitorear el rendimiento del modelo en un conjunto de validación durante el entrenamiento y detener el entrenamiento tan pronto como el rendimiento comience a deteriorarse (Prechelt, 1998).

La red neuronal, desarrollada en Python, usa las mismas características en todos los modelos. Está formada por 4 capas, de 25-50-50-25 neuronas. Los parámetros de entrada son los predictores seleccionados de cada modelo, para finalmente tener una única salida. Los datos de entrada han sido escalados entre 0 y 1, con el objetivo de mejorar el aprendizaje de la Red Neuronal. Se usó un 80% de datos como entrenamiento del modelo.

Las características del modelo se presentan en la tabla 4.

Tabla 4
Parámetros de configuración de la Red Neuronal

Característica	
Optimizador	AdamW Optimizer
Función de Perdida	Mse
Épocas	5000
Paciencia	2000

Las figuras 3 y 4 muestran las diferencias entre una red neuronal antes del entrenamiento y después del entrenamiento

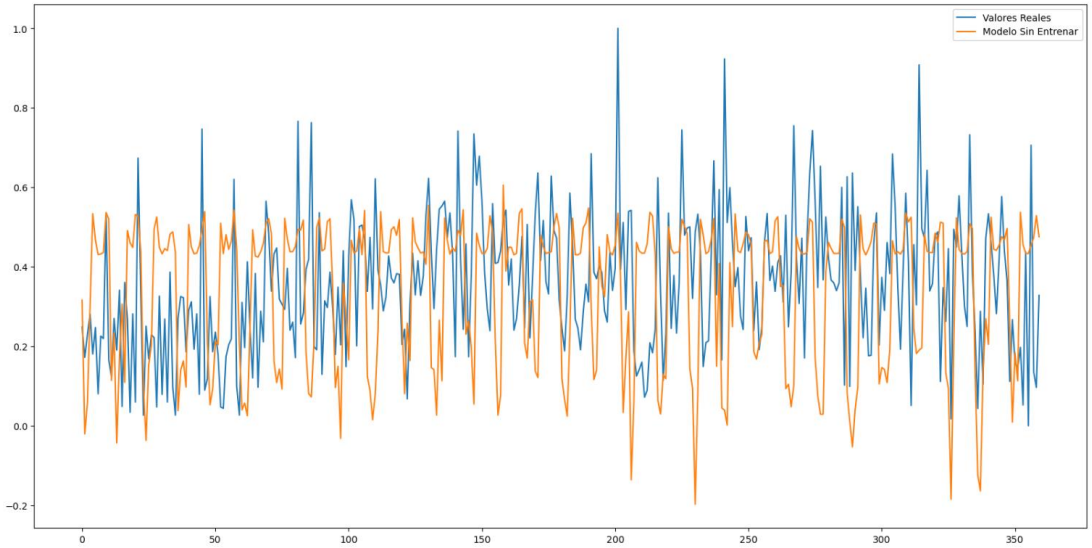


Figura 4
Red Neuronal previa al entrenamiento. Usa CESM2

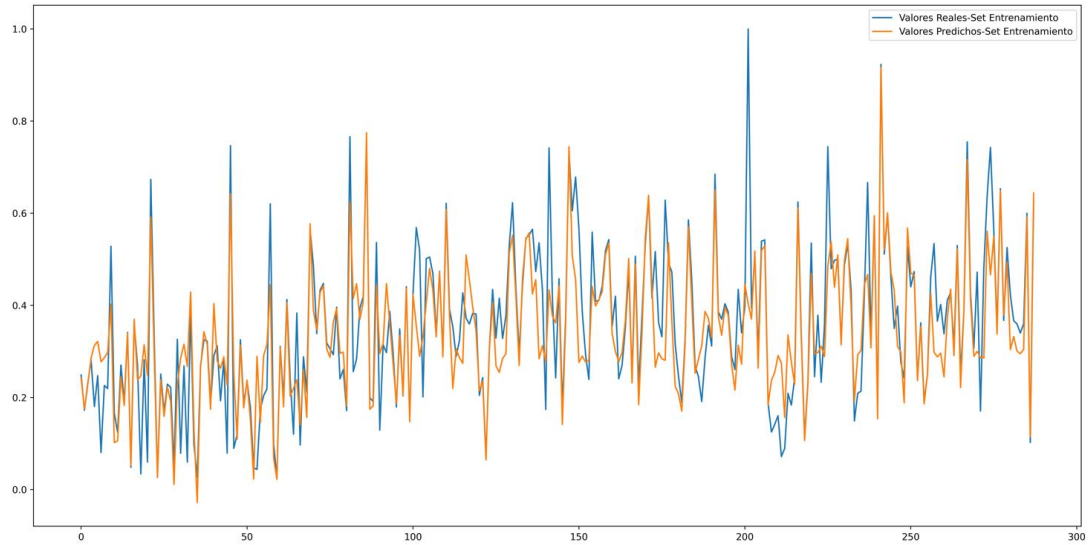


Figura 5
Red Neuronal después del entrenamiento. Usa CESM2

4.5 Modelo Hidrológico Autorregresivo Basado en Datos:

Se ha optado por utilizar un modelo autorregresivo (AR) para el análisis debido a varias ventajas que este enfoque ofrece (Box, 2013). Los modelos AR son relativamente simples en comparación con enfoques más complejos, lo que facilita su interpretación y aplicación. Esta simplicidad no solo hace que el modelo sea más accesible, sino que también reduce el riesgo de sobreajuste al enfocarse en las dependencias temporales entre las observaciones de la serie. Además, los modelos AR tienen la capacidad de capturar patrones de tendencia y estacionalidad en series temporales, lo que es crucial para realizar predicciones precisas en datos que muestran variaciones a lo largo del tiempo

Los modelos AR son relativamente simples en comparación con otros enfoques más complejos. La simplicidad del modelo facilita la interpretación y reduce el riesgo de sobreajuste. Los modelos AR pueden capturar patrones de tendencia y estacionalidad en series temporales.

Los modelos autorregresivos (AR) se emplean frecuentemente para analizar series temporales en hidrología debido a su estructura intuitiva y sencillez. Estos modelos funcionan bajo el principio de que el valor actual en una serie depende de sus valores anteriores, lo que los convierte en una opción simple y directa para modelar fenómenos temporales (Salas et al., 1980).

Se emplea un modelo autorregresivo con variables exógenas para prever los caudales medios mensuales en la cuenca del Tomebamba, utilizando como referencia los registros históricos de caudales de la zona. Este modelo toma en cuenta las mediciones previas para estimar los valores futuros de forma efectiva. El modelo ARX es una función dinámica lineal que se ha utilizado ampliamente para predecir el caudal de los ríos utilizando secuencias observadas de precipitaciones y escorrentías (Wood et al., 1980).

El modelo está representado como se muestra en la eq 5 (Hsu et al., 2009):

$$y_{t+1} = a_i y_{t-i} + b_j r_{t-j} + e_{i+1} \qquad \text{Eq 5}$$

Donde y_t y r_t es el caudal observado y la precipitación observada respectivamente en un tiempo t, a_i y b_i son parámetros.

Los parámetros a_i y b_i son una clase de modelos combinados no lineales (Zhou & Ding, 2020). Estos parámetros se estiman mediante la aplicación de un enfoque para la estimación de parámetros dependientes variables en el tiempo, reflejando aspectos no estacionarios y no lineales del comportamiento del sistema observado (P. Young, 2003).

La ecuación originalmente propuesta por Young, tiene la siguiente forma:

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})} u_e(k-\delta); \qquad u_e(k) = s(k)u_g(k) \qquad \text{Eq 6}$$

Donde k representa el valor de la variable asociada en el instante de muestreo k, $y(k)$ representa el caudal, $u_g(k)$ es la precipitación medida, $u_e(k)$ es la precipitación efectiva, δ es un retraso de tiempo y $s(k)$ es el índice de humedad del suelo.

Dado que se trata de un trabajo exploratorio, se utilizará una versión simplificada de la ecuación de Young. Con base en las consideraciones anteriores, la ecuación que define el modelo hidrológico se presenta en la Eq 7.

$$Q_{t+1} = a_{Q_{t-1}} * Q_{t-1} + b_{Q_{t-1}} * P_{e_t} \qquad \text{Eq 7}$$

Donde Q_t es el caudal predecido, P_{e_t} es la precipitación efectiva, $a_{Q_{t-1}}$ y $b_{Q_{t-1}}$ son parámetros dependientes del estado (SDP). En este modelo autorregresivo no estacionario, el caudal anterior se utiliza como indicador del estado de humedad de la cuenca, reemplazando directamente cualquier medida directa de este estado. Este enfoque corresponde a los mecanismos conceptuales de funcionamiento hidrológico descritos en el modelo basado en datos. Este modelo asume que el caudal previo proporciona una representación adecuada de las condiciones de humedad anteriores, que son cruciales para predecir los flujos futuros en la cuenca. Este proceso es descrito en (P. C. Young & Beven, 1994). El modelado DBM (Modelo Basado en Datos) ha sido desarrollada con el objetivo de producir un enfoque estadístico general y unificado para el modelado de sistemas dinámicos lineales y no lineales (como es este caso), estocásticos, a partir de datos de series temporales (P. Young, 1998).

Este modelo fue desarrollado en MATLAB, a traves de toolbox especializados (*Download Matlab Captain Toolbox – System Identification, Time Series Analysis, Forecasting and True Digital Control*, s/f).

4.6 Métricas de Error

4.6.1 Nash – Sutcliffe y Root Mean Square Error (RMSE)

El error cuadrático medio (RMSE, por sus siglas en inglés) es una métrica comúnmente utilizada para evaluar modelos predictivos (Hodson, 2022). Calcula la diferencia entre los valores que el modelo predice y los valores reales observados. Esta métrica permite

cuantificar el tamaño del error del modelo, expresado en las mismas unidades que la variable que se está analizando. Viene dado por la eq 8:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \qquad \text{Eq 8}$$

Donde y_i son los valores observados, $\widehat{y_i}$ son los valores predichos por el modelo, n es el número de observaciones.

El criterio NSE (Eficiencia de Nash-Sutcliffe) se utiliza para evaluar el rendimiento de los modelos. Este indicador muestra tanto los aspectos positivos como los negativos del modelo y mejora a medida que aumenta la calidad del mismo (Zeybek, 2018). El NSE es una medida estandarizada que evalúa cuán bien el modelo ajusta los datos observados. Un NSE de 1 significa que las predicciones del modelo son perfectas. Un valor de 0 indica que el modelo es tan preciso como simplemente usar el promedio de los datos observados, mientras que un valor negativo sugiere que el modelo predice peor que el promedio de los datos observados. Nash esta dado por la eq 9:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad \text{Eq 9}$$

Donde y_i son los valores observados, $\widehat{y_i}$ son los valores predichos por el modelo, $\bar{y}$ es el promedio de valores observados y n es el número de observaciones.

4.7 Determinación del Mejor GCM.

Inicialmente, seleccionamos cuatro Modelos Climáticos Globales (GCM) adecuados para nuestra región de estudio. Evaluamos estos modelos para verificar su capacidad de predecir los patrones de precipitación en la zona de interés. Recolectamos datos observados de precipitación de estaciones meteorológicas locales, abarcando el periodo de 1960 a 1990, en función de la disponibilidad de estos datos en la región. Estos datos fueron cuidadosamente preprocesados para asegurar que sus resoluciones temporal y espacial coincidieran con las salidas de los modelos GCM.

El procedimiento para identificar el mejor GCM incluyó la selección de predictores relevantes, el entrenamiento y la evaluación de cada modelo. Parte de este proceso involucró comparar cómo los modelos se comportaban con y sin la inclusión de interacciones entre variables, utilizando para ello una red neuronal. Los criterios para evaluar el desempeño de los modelos,

detallados en la sección 4.6.1, incluyeron el análisis de las métricas de Error Cuadrático Medio (RMSE) y la Eficiencia de Nash-Sutcliffe (NSE) tanto en las fases de entrenamiento como en la de validación. Elegimos el modelo que mostró el RMSE más bajo y el NSE más alto en la fase de validación como el más adecuado para predecir la precipitación en nuestra región.

5 RESULTADOS

5.1 Control de Calidad de Datos

Tras seleccionar el período de análisis, se recopiló la información de precipitación de estaciones cercanas a la cuenca del Tomebamba. Esta recopilación de datos se realizó a escala mensual, proporcionando un panorama detallado de las variaciones de precipitación a lo largo del tiempo.

Una vez obtenidos los datos, se procedió a su verificación de la calidad de los datos mediante curvas de doble masa. Los resultados de esta verificación indicaron que los datos de precipitación eran consistentes y adecuados para su uso en este estudio.

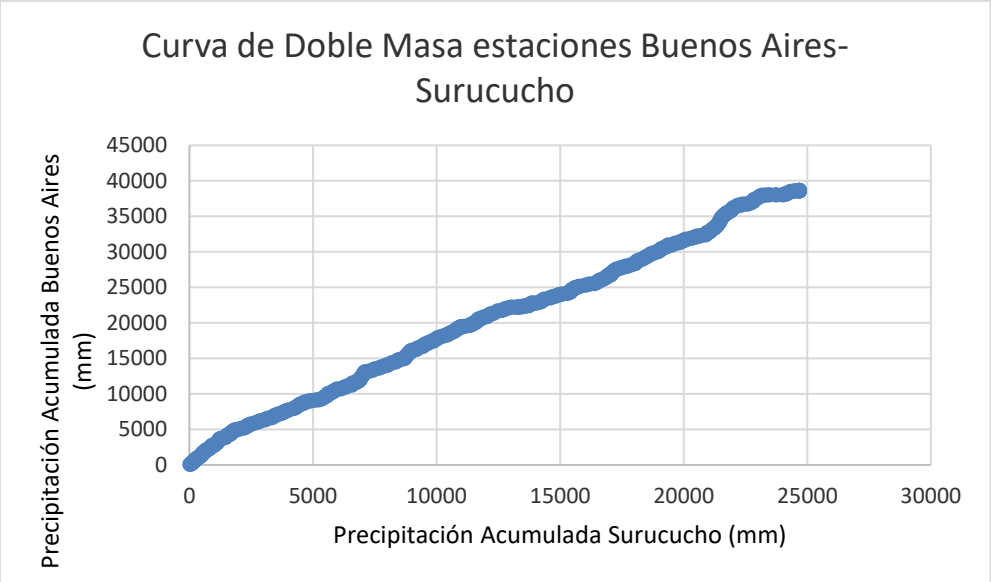


Figura 6
Curvas de Doble Masa. Estaciones Buenos Aires - Surucucho

5.2 Regresión Lineal Múltiple

La regresión lineal múltiple se realizó con la ayuda de Python y la librería statsmodels.api, la cual nos brinda los coeficientes de la regresión. Los coeficientes obtenidos proporcionaron una comprensión más detallada sobre la relación entre los datos satelitales de precipitación y la precipitación media observada.

Sin embargo, uno de los problemas identificados fue la presencia de multicolinealidad, donde las variables independientes estaban altamente correlacionadas entre sí. La multicolinealidad puede inflar las varianzas de los coeficientes de regresión, haciendo que las estimaciones sean menos precisas y dificultando la interpretación de los efectos individuales de las variables predictoras. Debido a estas limitaciones, se determinó que la regresión lineal múltiple no era la metodología más adecuada para predecir la precipitación en este caso específico.

La multicolinealidad puede ser abordada mediante la eliminación de variables altamente correlacionadas, la combinación lineal de variables, o el uso de técnicas avanzadas como la regresión Lasso, que son más robustas frente a la multicolinealidad.

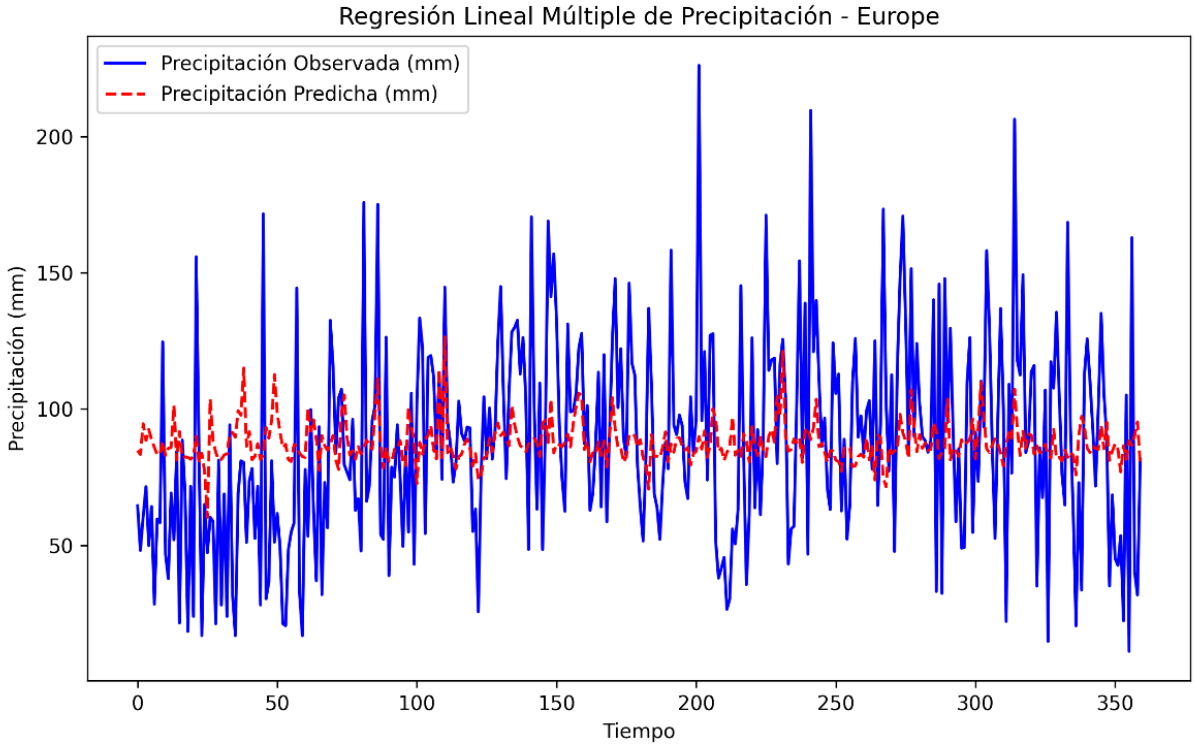


Figura 7
Uso de regresión lineal múltiple para predecir precipitación.

5.3 Interacción de Datos

Para realizar la interacción de datos (Feature Engineering) se utilizó la biblioteca Scikit-learn, específicamente la herramienta PolynomialFeatures. Esta herramienta permite crear términos polinómicos a partir de las características originales, expandiendo así el conjunto de datos y capturando relaciones más complejas entre las variables.

En función del número de variables en cada modelo, se generaron las variables de interacción, que se detallan en la Tabla 5.

El uso de PolynomialFeatures facilita la identificación de patrones no lineales en los datos que las características originales podrían no captar. Esto es particularmente útil en el contexto del análisis climático, donde las relaciones entre las variables pueden ser altamente no lineales. La generación de estos términos polinómicos ayuda a capturar mejor la dinámica subyacente de los datos climáticos.

La implementación de esta técnica en el estudio permitió mejorar la precisión de los modelos predictivos utilizados para analizar el impacto del cambio climático en la cuenca del Tomebamba. Además, este enfoque proporciona una mayor flexibilidad al modelo, permitiendo que se ajuste mejor a las variaciones presentes en los datos de precipitación recopilados.

Tabla 5
Creación de nuevas variables mediante PolynomialFeatures

Modelo	Variables sin Interacción	Variables con Interacción
Europe EC-Earth3-CC	6	78
Germany AWI-CM-1-1-MR	6	78
Italy CMCC-ESM2	4	36
Usa CESM2	4	36

5.4 Identificación de variables significativas mediante LASSO

Para la identificación de las variables significativas, se utilizó un programa en Python con la ayuda de la librería Lasso. Esta herramienta proporciona resultados como el RMSE y el parámetro de regularización λ. Estos resultados son esenciales para construir una curva que permita identificar cuál sería el parámetro de regularización óptimo, ya que un valor adecuado de λ equilibra la complejidad del modelo y su capacidad para generalizar a nuevos datos. Este enfoque no solo optimiza el rendimiento del modelo, sino que también previene el sobreajuste, asegurando que las predicciones sean robustas y fiables.

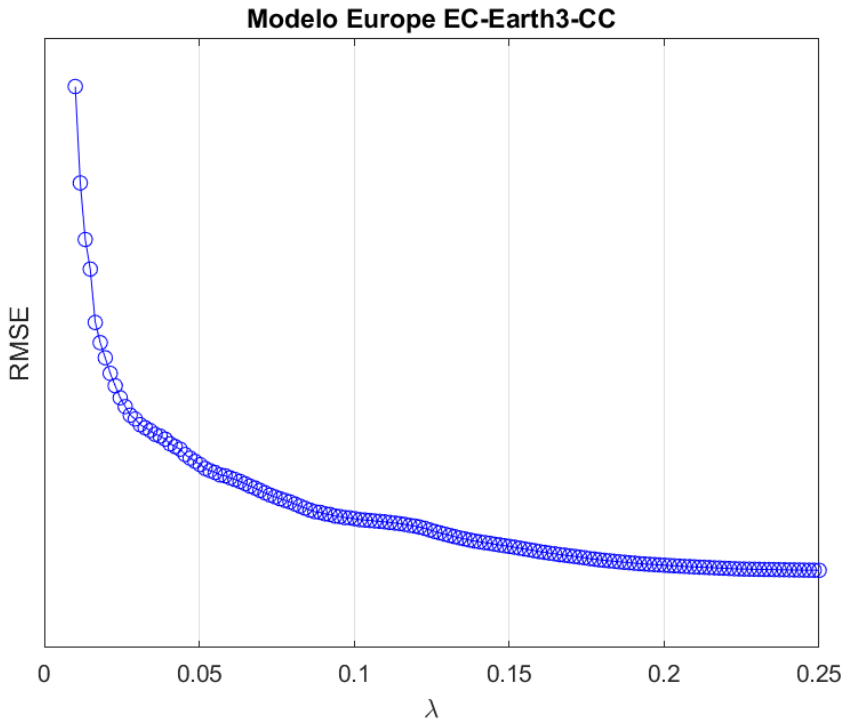


Figura 8
Lambda óptimo en el mínimo RMSE del modelo Europeo

El algoritmo Lasso ayudó a reducir el número de predictores seleccionando solo aquellos que tienen un mayor impacto en el modelo, lo que facilita la interpretación y mejora la precisión de este. Con estos resultados, el programa es capaz de identificar cuáles son las variables más significativas o sensibles. Los predictores seleccionados según Lasso se muestran en la Tabla 6.

Tabla 6
Predictores seleccionados por LASSO

Modelo	Predictores Seleccionados
Europe EC-Earth3-CC	7
Germany AWI-CM-1-1-MR	13
Italy CMCC-ESM2	8
Usa CESM2	6

5.5 Redes Neuronales

5.5.1 Redes Neuronales: Calibración

Los análisis se realizaron utilizando MATLAB y Python, que son plataformas que ofrecen herramientas para el procesamiento y análisis de datos. Estas herramientas nos permitieron manejar eficientemente los cálculos y visualizaciones necesarios para nuestro acercamiento

a la determinación del mejor GCM. Con el 80% de los datos usados para la calibración, los resultados obtenidos se presentan en la tabla 7 y 8.

Tabla 7
Resultados estadísticos conjunto entrenamiento. Modelos sin Interacción

Variables sin Interacción		
Modelo	Entrenamiento	
	RMSE	NSE
Europe EC-Earth3-CC	0.047	0.929
Germany AWI-CM-1-1-MR	0.03	0.971
Italy CMCC-ESM2	0.053	0.91
Usa CESM2	0.095	0.708

Tabla 8
Resultados estadísticos conjunto entrenamiento. Modelos con Interacción

Variables con Interacción		
Modelo	Entrenamiento	
	RMSE	NSE
Europe EC-Earth3-CC	0.093	0.725
Germany AWI-CM-1-1-MR	0.091	0.737
Italy CMCC-ESM2	0.119	0.547
Usa CESM2	0.112	0.599

En el estudio del modelado de series temporales hidrológicas, la efectividad de las redes neuronales artificiales se destaca particularmente al evaluar su desempeño con estadísticos como el coeficiente de eficiencia de Nash-Sutcliffe (NSE) y la raíz del error cuadrático medio (RMSE). Estos indicadores han mostrado resultados sobresalientes en el conjunto de entrenamiento, subrayando la capacidad superior de las redes neuronales para capturar la complejidad y la variabilidad de los datos hidrológicos en comparación con métodos tradicionales o más simplistas.

La combinación de un alto NSE y un bajo RMSE en el conjunto de entrenamiento fortalece la argumentación a favor de las redes neuronales, resaltando su utilidad como una herramienta poderosa.

5.5.2 Redes Neuronales: Validación

Los resultados del conjunto de validación son cruciales para determinar cuán bien los modelos, como las redes neuronales en este contexto, generalizan más allá de los datos de entrenamiento a situaciones no vistas. Este proceso de validación nos ayuda a verificar que

el modelo no solo es eficaz con datos ya conocidos, sino que también mantiene su precisión al enfrentarse a nuevos datos, como los escenarios que se usan para modelar.

La efectividad del modelo en el conjunto de validación no solo confirma su competencia bajo condiciones controladas, sino que también fortalece nuestra confianza en su utilidad práctica al momento de predecir. La figura 9 muestra el conjunto de validación en uno de los modelos.

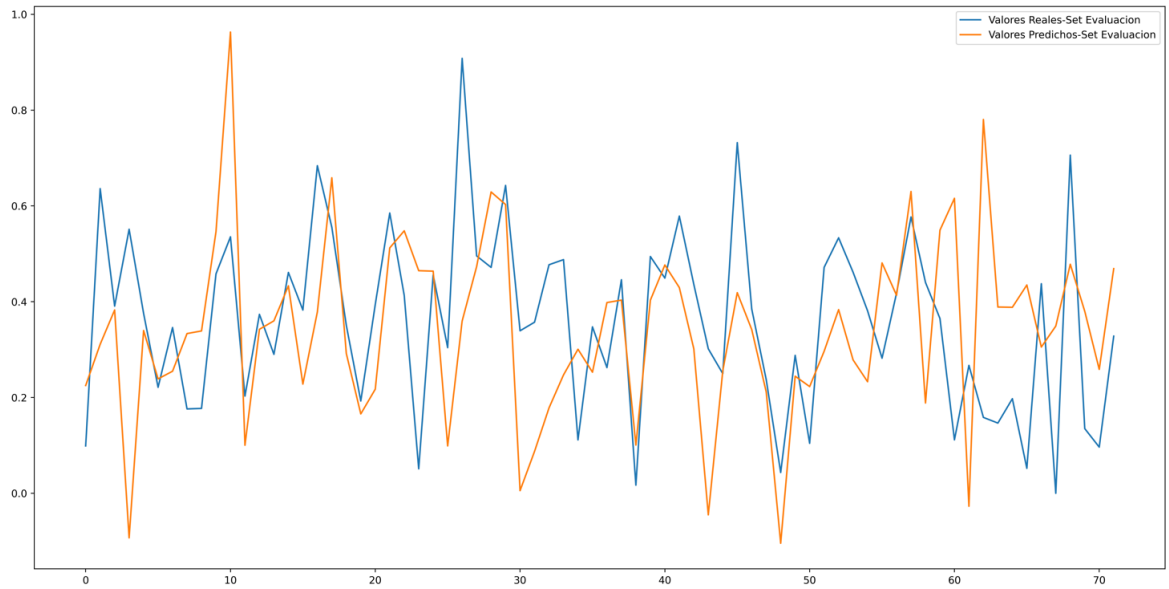


Figura 9
Conjunto de Evaluación. Italy CMCC-ESM2

Los estadísticos del conjunto de validación de todos los modelos se presentan en la tabla 9 y 10.

Tabla 9
Estadísticos del conjunto de validación. Variables sin Interacción

Variables sin Interacción		
Modelo	Validación	
	RMSE	NSE
Europe EC-Earth3-CC	0.311	-1.283
Germany AWI-CM-1-1-MR	0.257	-0.562
Italy CMCC-ESM2	0.246	-0.427
Usa CESM2	0.257	-0.563

Tabla 10
Estadísticos del conjunto de validación. Variables con Interacción

Variables con Interacción		
Modelo	Validación	
	RMSE	NSE
Europe EC-Earth3-CC	0.281	-0.871
Germany AWI-CM-1-1-MR	0.159	0.401
Italy CMCC-ESM2	0.22	-0.142
Usa CESM2	0.246	-0.433

5.6 Determinación del mejor GCM

Considerando ambos escenarios, con y sin interacción, el modelo Germany AWI-CM-1-1-MR que incluye interacciones se destaca por su robustez durante la fase de validación. Es el único que presenta un NSE positivo junto con un RMSE relativamente bajo cuando incluye interacciones. Esto indica que la inclusión de interacciones mejora notablemente su capacidad predictiva, convirtiéndolo en la mejor opción para predecir la precipitación según los datos analizados.

Este análisis nos demuestra que, dadas las características de los datos y los métodos aplicados en este estudio, el modelo Germany AWI-CM-1-1-MR con interacciones es superior. Esto resalta la importancia de considerar las interacciones en el modelado climático cuando estas son relevantes, ya que pueden influir significativamente en la precisión de las predicciones.

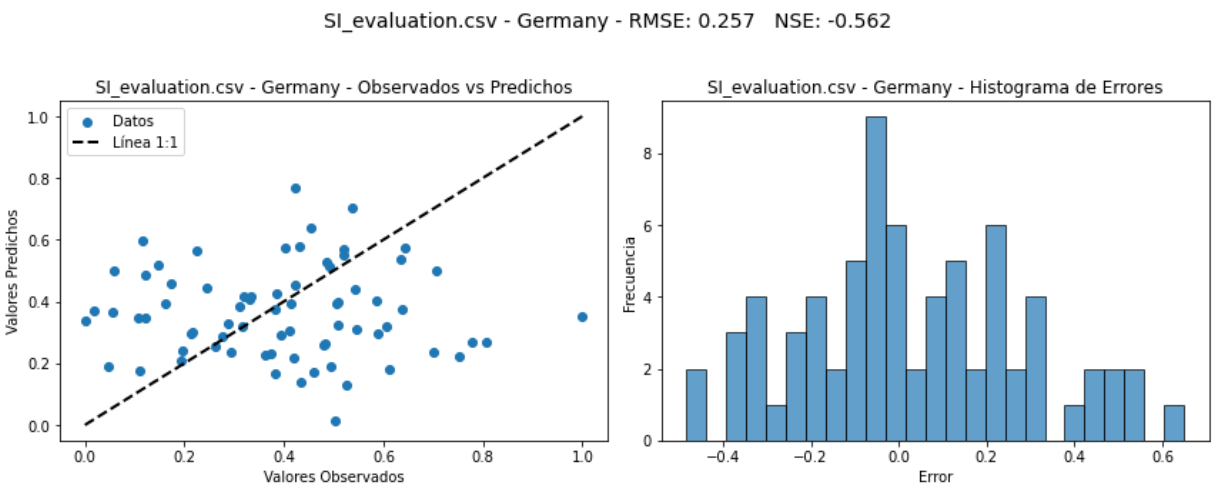


Figura 10
Conjunto de Evaluación. Se observa que las variables sin interacción presentan una gran dispersión con respecto a los valores observados

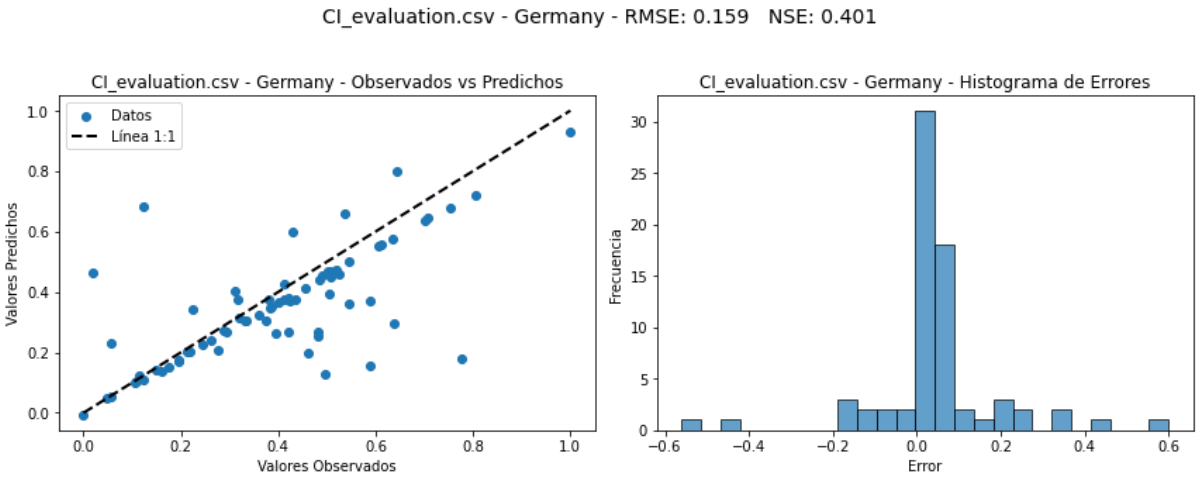


Figura 11
Conjunto de Evaluación. Se observa que las variables con interacción reducen significativamente la dispersión con respecto a los valores observados

En base a los resultados, el conjunto completo evaluado en la red neuronal del modelo Germany es el siguiente:

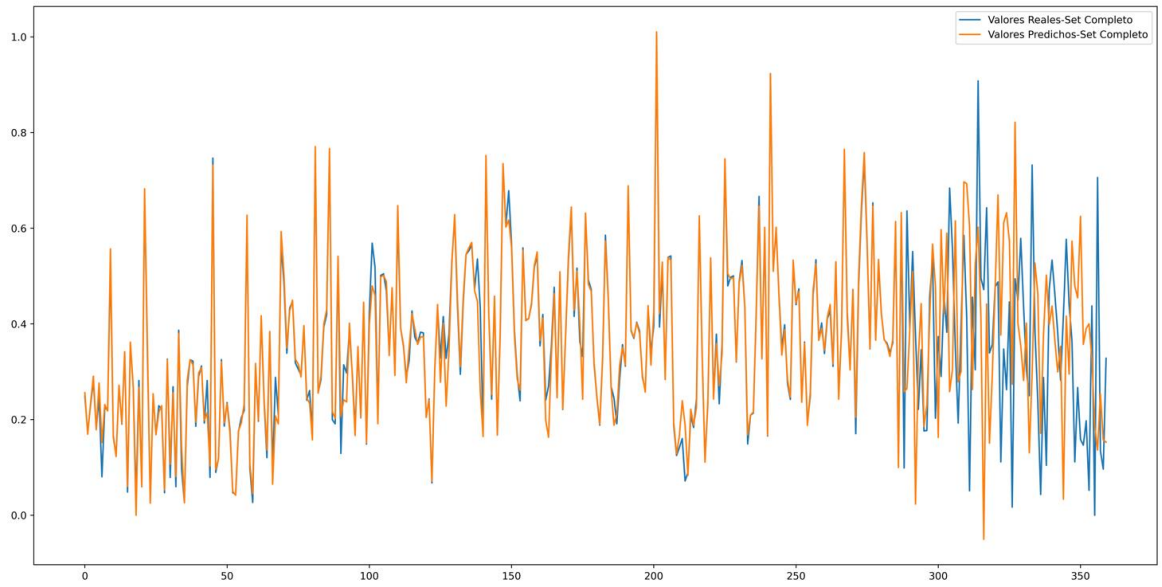


Figura 12
Red Neuronal. Precipitaciones a lo largo de los 30 años del conjunto de calibración - validación

5.7 Redes Neuronales: Predicción de la precipitación.

Una vez seleccionado el modelo con el que se va a trabajar, se recopilaron los datos de precipitación futuros correspondientes al período comprendido entre los años 2015 y 2100. Con estos datos, se realizó el feature engineering, seleccionando exactamente las mismas interacciones que se utilizaron en el conjunto de validación. Este enfoque asegura la consistencia y la comparabilidad de los resultados.

Los datos futuros de precipitación fueron ingresados a la red neuronal previamente entrenada. Esto permitió obtener la precipitación media mensual futura para el período mencionado. La red neuronal, al estar ya entrenada con datos históricos y validados, proporciona una proyección precisa y confiable de las tendencias de precipitación futura. Los resultados obtenidos de estas proyecciones se presentan en las Figuras 13 y 14.

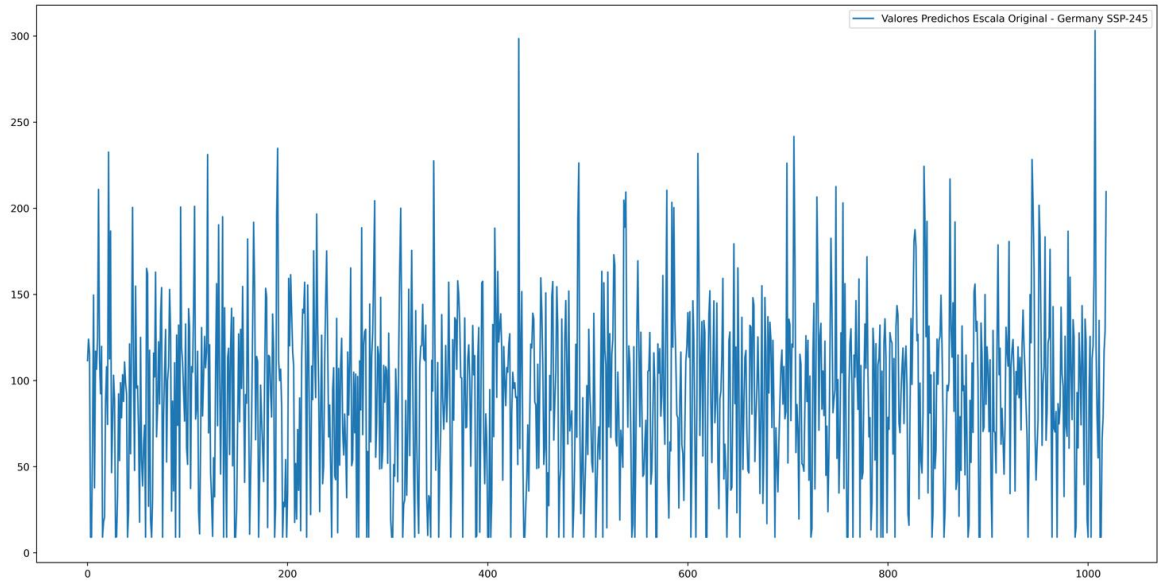


Figura 13
Precipitación (mm) predicha bajo el escenario Germany AWI-CM-1-1-MR SSP-245

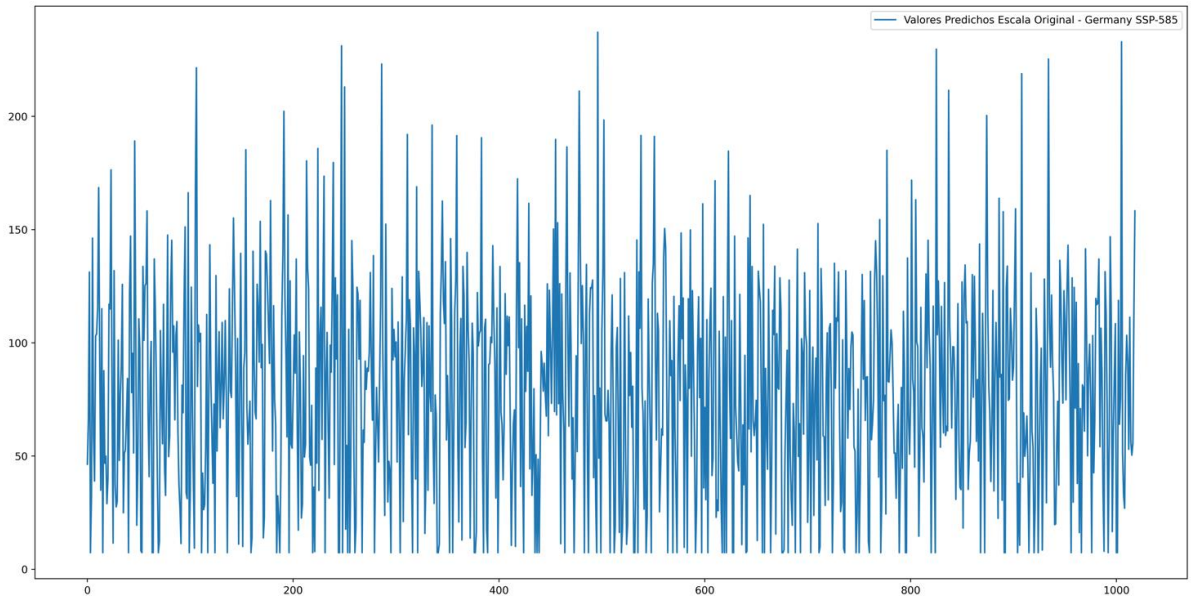


Figura 14
Precipitación (mm) predicha bajo el escenario Germany AWI-CM-1-1-MR SSP-585

La gráfica de precipitación muestra un aumento en los picos altos, lo que indica que en el futuro habrá periodos con mayor cantidad de lluvia. Sin embargo, también se observa un

incremento en los picos bajos de precipitación, lo que sugiere que los periodos de sequía serán más frecuentes. Esto implica que, aunque habrá episodios de lluvias intensas, estos serán menos comunes que en la actualidad, resultando en un aumento en la frecuencia y duración de las épocas de sequía. Para un mejor entendimiento, se hace uso de un BoxPlot de precipitaciones

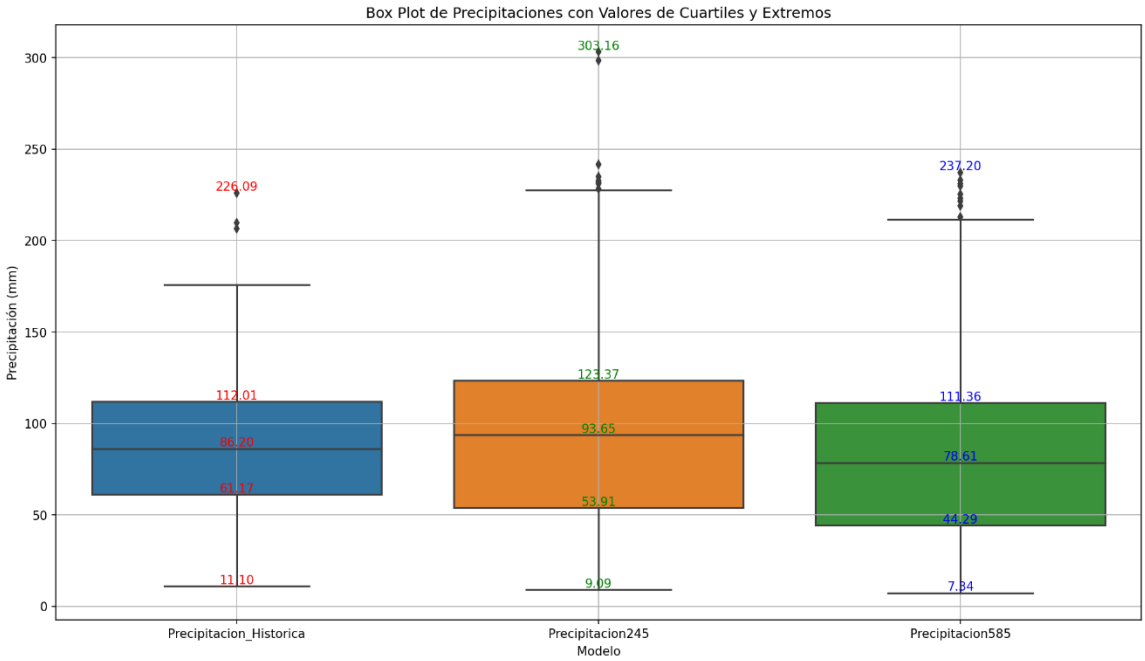


Figura 15
BoxPlot de precipitaciones de los modelos SSP 2 - 4.5 y SSP 5 - 8.5 comparado con las precipitaciones históricas

En base al BoxPlot, se puede extraer varios resultados:

- Mediana: La precipitación mediana es mayor en el escenario SSP 2 – 4.5 (93.65 mm) comparado con la histórica (86.20 mm) y SSP 5 – 8.5 (78.61 mm).
- Distribución de Datos: La variabilidad de los datos es mayor en los escenarios de modelado comparado con los datos históricos, lo que indica una mayor dispersión en los datos futuros.
- Valores Atípicos: Los escenarios modelados presentan valores atípicos más altos (303.16 mm para SSP 245 y 237.20 mm para SSP 585) en comparación con el valor atípico histórico de 226.09 mm.
- Rango de Precipitaciones: El rango total de precipitaciones (diferencia entre los bigotes) es mayor en los escenarios de modelado comparado con los datos históricos, lo que sugiere que los modelos predicen una mayor variabilidad en la precipitación

futura. Los valores atípicos más altos en estos escenarios también indican que eventos extremos de precipitación podrían ser más frecuentes o severos en el futuro.

- La mediana de la precipitación en los escenarios futuros SSP 2 – 4.5 y SSP 5 – 8.5 muestra un ligero aumento (SSP 2 – 4.5) y una ligera disminución (SSP 5 – 8.5) en comparación con la precipitación histórica. Esto indica que los diferentes escenarios de emisión de gases de efecto invernadero pueden tener impactos variados en los patrones de precipitación.

5.8 Modelo Hidrológico Autorregresivo Basado en Datos

5.8.1 Modelo Hidrológico Autorregresivo Basado en Datos: Calibración

Se llevó a cabo un análisis independiente de estado utilizando series temporales de caudales para examinar condiciones previas. Este estudio culminó en la formulación de ecuaciones que adoptan un modelo logístico. Este proceso se encuentra detallado en (P. C. Young et al., 2001). Estas curvas de ajuste, desarrolladas con MATLAB y su paquete especializado, se presentan en las figuras 16 y 17. Estas gráficas se comparan con las Ecuaciones 10 y 11 derivadas anteriormente, demostrando una alta coincidencia con los ajustes realizados por el software.

El análisis de las curvas muestra que, a medida que la precipitación efectiva se incrementa hasta un punto de saturación, se produce un cambio significativo en la curva 'b'. Asimismo, la curva 'a' ilustra cómo el aumento del caudal altera su relación con el caudal precedente, subrayando que la influencia del caudal anterior sobre el actual es más pronunciada.

$$a_{Q_{t-1}} = 0.1654 * e^{\left(-0.001221 * X_1^{1.843}\right)} + 0.2946 \qquad \text{Eq 10}$$

$$b_{Q_{t-1}} = \frac{0.32}{0.8717 + e^{-0.00681 * X_2^{1.494}}} \qquad \text{Eq 11}$$

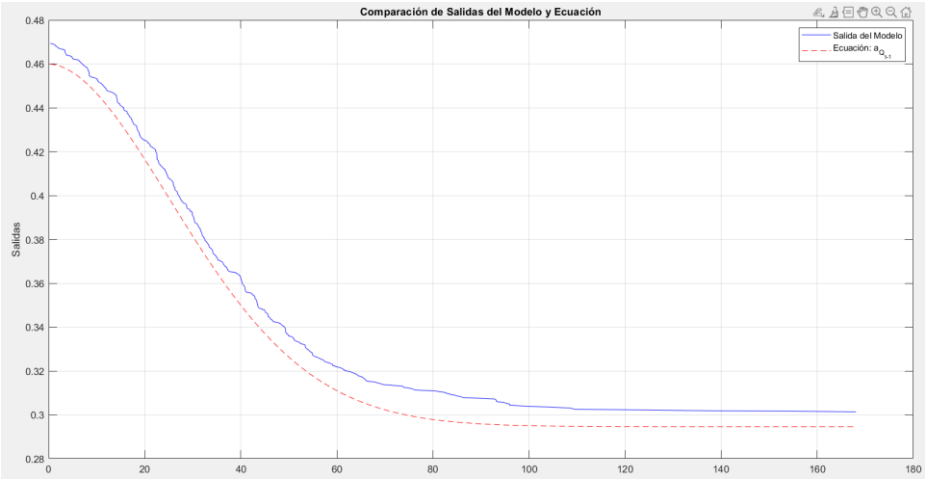


Figura 16
Comparación de la curva de ajuste "a" entre el dado por MATLAB y el ajustado $a_{Q_{t-1}}$

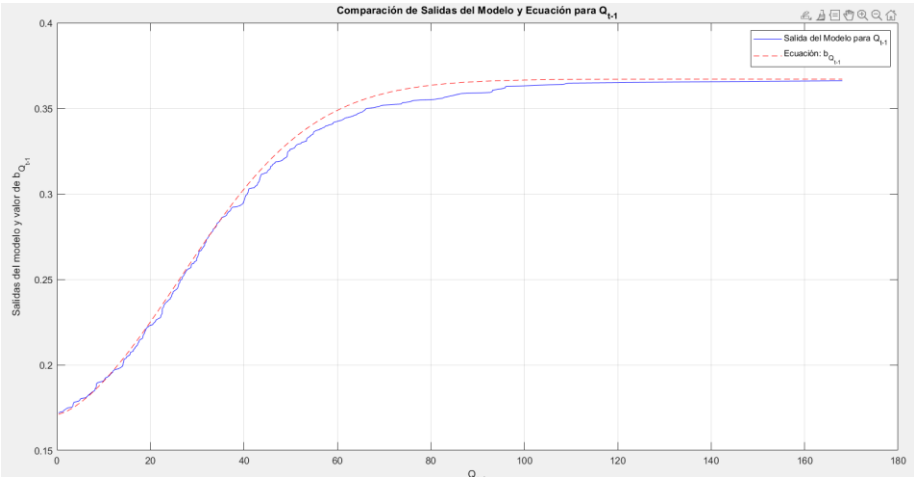


Figura 17
Comparación de la curva de ajuste "b" entre el dado por MATLAB y el ajustado $b_{Q_{t-1}}$

Una vez que se han determinado todos los parámetros necesarios para el modelo, los resultados de la predicción de caudales se visualizan en la figura 18. Esta visualización sirve para interpretar cómo el modelo aplica los parámetros ajustados para prever el comportamiento del caudal bajo diversas condiciones de precipitación.

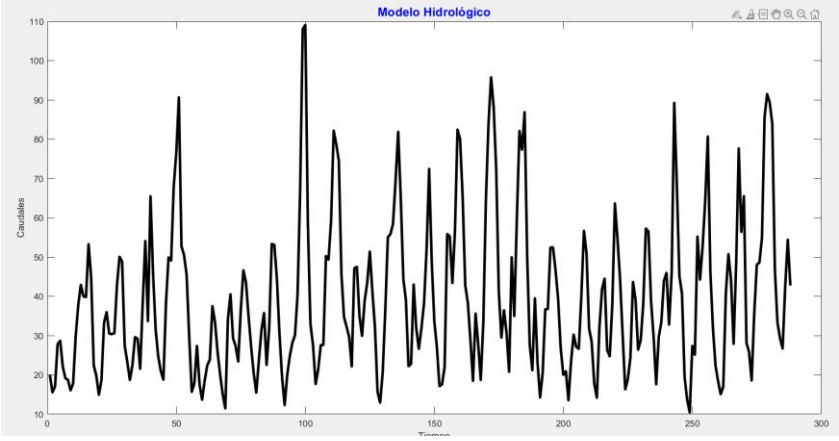


Figura 18
Modelo Hidrológico Autorregresivo en función del caudal anterior y precipitación efectiva

5.8.2 Modelo Hidrológico Autorregresivo Basado en Datos: Validación

Para validar la precisión del modelo hidrológico, se utilizó el coeficiente de eficiencia de Nash-Sutcliffe, que es una medida estándar para evaluar la exactitud de las simulaciones de caudales. En este caso, el modelo alcanzó un coeficiente de Nash-Sutcliffe de 0.546, indicando una capacidad moderada del modelo para simular los caudales comparados con los datos observados en la cuenca del Tomebamba. Este resultado se visualiza en la figura 19, que muestra una comparación gráfica entre los caudales simulados y los observados, proporcionando una representación visual de cómo el modelo reproduce la variabilidad y magnitud de los caudales reales.

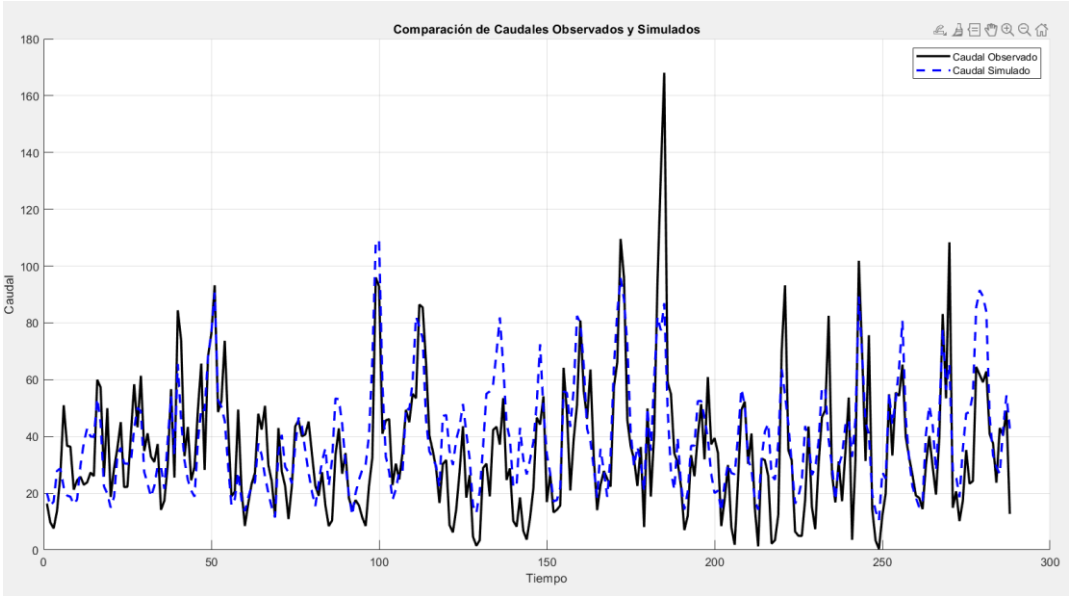


Figura 19
Validación de Modelo Hidrológico DBM

5.8.3 Modelo Hidrológico Autorregresivo Basado en Datos: Aplicación

Se presentan los resultados obtenidos de las simulaciones de caudales futuros, basados en precipitaciones proyectadas bajo dos escenarios climáticos distintos: SSP-245 y SSP-585. El análisis se realizó utilizando el modelo climático global Germany AWI-CM-1-1-MR. El escenario SSP-245 representa una situación de cambio climático menos intensa comparado con el SSP-585, que asume condiciones más extremas. Estos se muestran en la figura 20

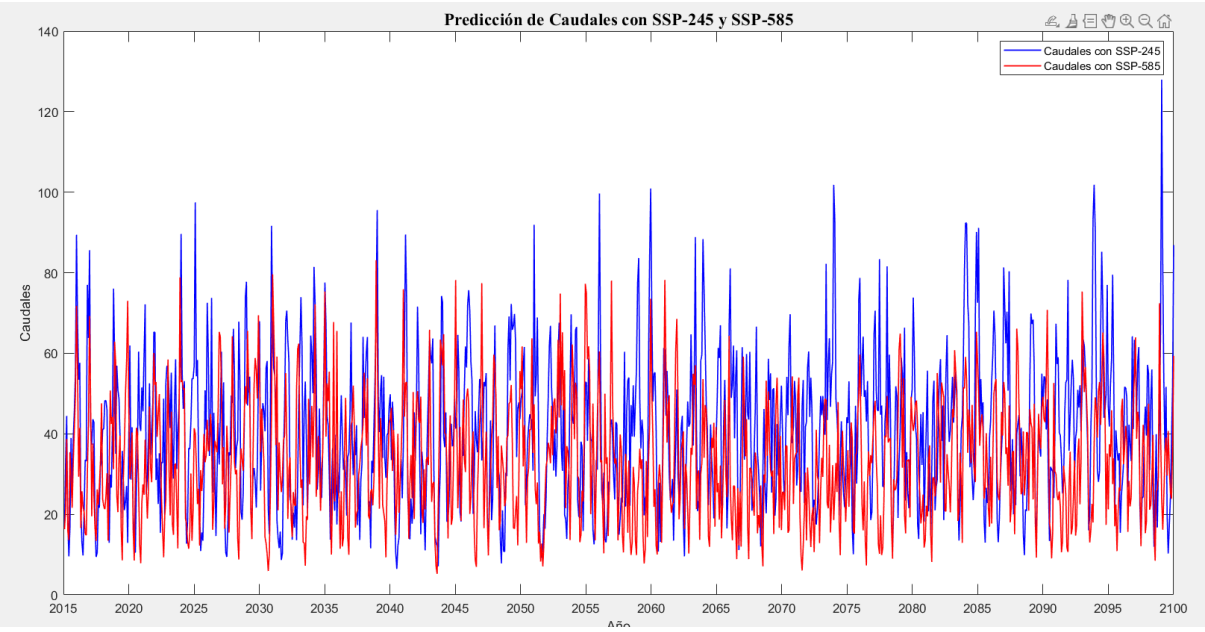


Figura 20
Predicción de Caudales en los distintos escenarios SSP

Las gráficas presentadas muestran las variaciones en los caudales pronosticados para ambos escenarios. Las diferencias entre ellos son:

- SSP-245: Este escenario muestra variaciones de caudal más contenidas a lo largo del tiempo. Esto indica precipitaciones menos extremas, resultando en caudales que fluctúan dentro de límites manejables. Este patrón sugiere un impacto menos severo del cambio climático en la región.
- SSP-585: Las predicciones bajo este escenario revelan una volatilidad mucho mayor, con picos significativos que reflejan eventos de precipitación intensa. Este comportamiento apunta a la posibilidad de enfrentar eventos climáticos extremos como inundaciones, lo que podría requerir estrategias de gestión del agua más estrictas y preparación para emergencias.

La variabilidad en los caudales observada bajo el escenario SSP-585 necesitará de infraestructuras robustas y estrategias adaptativas para mitigar posibles desastres naturales. En contraste, el escenario SSP-245 permite una planificación más flexible y medidas de

adaptación menos costosas, pero cabe acotar que, en general, la sociedad siempre ha estado por encima de los escenarios construidos, por lo que es factible tomar el peor escenario como guía para el futuro. Para un mejor análisis de resultados, se hace un análisis mediante un BoxPlot de todos los escenarios.

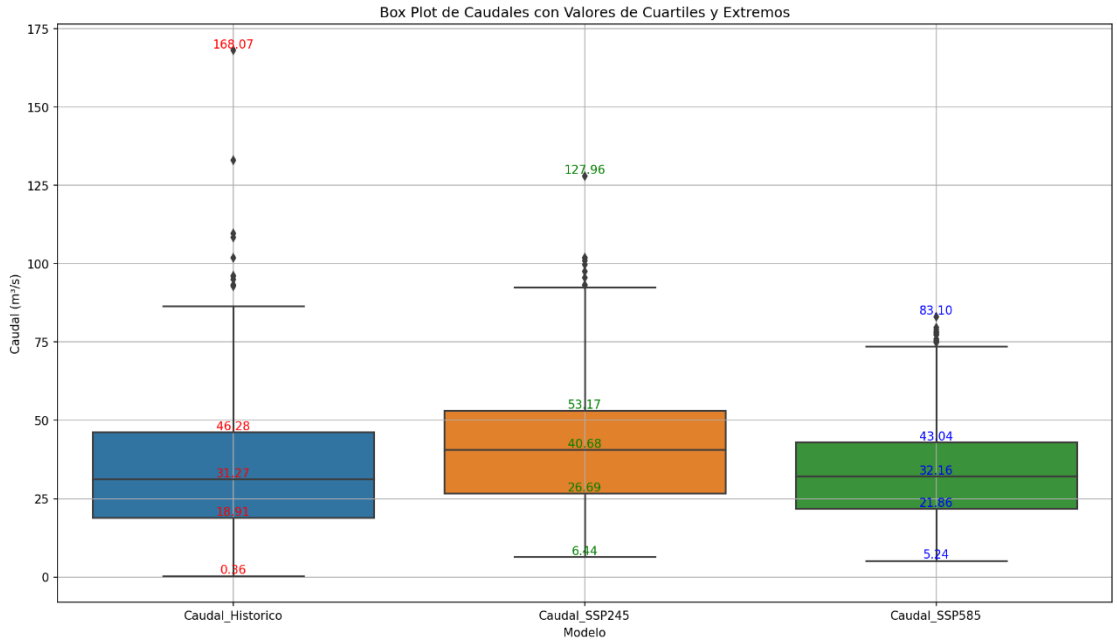


Figura 21
BoxPlot de caudales de los modelos SSP 2 - 4.5 y SSP 5 - 8.5 comparado con los caudales históricos

En base al BoxPlot, se obtienen los siguientes resultados:

- La distribución de caudales en el escenario SSP 245 tiene un rango intercuartil más amplio que en los datos históricos, lo que indica una mayor dispersión y posiblemente una mayor frecuencia de eventos extremos.
- En contraste, el escenario SSP 585 muestra una distribución de caudales más compacta en comparación con el escenario histórico, lo que podría sugerir una estabilización o una reducción en la variabilidad del caudal.
- La presencia de valores atípicos más altos en los escenarios futuros sugiere que los eventos extremos de caudal podrían ser más comunes o más intensos, lo cual es una consideración importante para el diseño de infraestructuras de control de inundaciones y almacenamiento de agua.
- La mediana del caudal en los escenarios futuros SSP 245 y SSP 585 muestra un ligero aumento (SSP 245) y una ligera disminución (SSP 585) en comparación con el caudal histórico, lo que puede indicar diferentes respuestas hidrológicas a los escenarios de cambio climático.
- La mayor variabilidad en el escenario SSP 245 sugiere una mayor incertidumbre en las predicciones de caudal futuro bajo este escenario en comparación con el histórico y el SSP 585.
- La presencia de valores atípicos más altos en los escenarios modelados podría ser un indicador de eventos de caudal más extremos en el futuro.

6 CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones

Tras concluir el estudio exploratorio, se evaluó el impacto del cambio climático en la Cuenca del Tomebamba. Los resultados revelaron un incremento significativo tanto en los picos altos como en los bajos de precipitación en la zona. Este patrón de variabilidad extrema en las precipitaciones resultó en una disminución en el caudal del río Tomebamba, lo que sugiere un escenario preocupante para el suministro de agua en el futuro.

Estos cambios en el régimen de precipitación y por ende el impacto en el caudal del río tienen implicaciones directas para la gestión de los recursos hídricos en la Cuenca del Tomebamba. La disminución en el caudal proyecta un futuro donde el suministro de agua potable podría no ser suficiente para satisfacer la demanda de la población y las actividades agrícolas e industriales de la zona.

Una vez completado el modelo climático, se concluyó que las tendencias de las precipitaciones se mantuvieron consistentemente durante las etapas de entrenamiento y validación, así como en el período de predicción. Esto sugiere que el modelo ha sido entrenado de manera adecuada.

Además, se observó que la ingeniería de características (feature engineering) contribuyó significativamente al enriquecimiento de las variables. Los mejores resultados se obtuvieron de un modelo que incorporó interacciones entre las variables, lo que mejoró notablemente su capacidad predictiva. Específicamente, se observó una disminución considerable en las métricas de error, como el índice de Nash y el RMSE, al utilizar variables con interacciones.

De igual manera, el uso de la regresión Lasso fue fundamental para seleccionar las variables más relevantes y sensibles para este estudio, evitando así el problema de sobreajuste (overfitting) al realizar las predicciones. El sobreajuste ocurre cuando un modelo se ajusta demasiado a los datos específicos de entrenamiento, lo que resulta en un rendimiento deficiente al aplicarse a nuevos datos.

El uso de una red neuronal como herramienta de análisis en este estudio demostró ser significativamente más eficaz que la regresión lineal. Esto se debe a la capacidad de las redes neuronales para capturar características específicas y complejas de cada variable, incluso cuando estas presentan relaciones no lineales. Como resultado, las predicciones obtenidas con la red neuronal fueron más precisas y confiables.

El empleo del modelo hidrológico en las fases de validación, calibración y predicción proporcionó una comprensión más profunda de los efectos potenciales del cambio climático sobre los recursos hídricos. Durante el periodo inicial del estudio, los caudales se mantuvieron relativamente estables, sin mostrar cambios significativos. Sin embargo, a medida que avanzaba el tiempo, se evidenció una tendencia decreciente en los caudales, caracterizada por un incremento en los episodios de caudales bajos. Estos hallazgos sugieren que, bajo el escenario climático SSP 558, podríamos enfrentar una reducción en la disponibilidad de agua en el futuro.

6.2 Observaciones:

Es importante considerar que, al realizar un estudio utilizando la metodología de descarga de datos de Copernicus, se debe verificar la resolución de dichos datos. Necesario que estos estén en la misma escala de tiempo que los datos locales para asegurar la precisión en el análisis.

También es necesario contar con una serie de datos que abarque un largo período de tiempo. Esto es crucial para evitar la necesidad de rellenar datos faltantes, lo que podría comprometer la calidad de los datos locales y, en consecuencia, la precisión del estudio.

Asimismo, es importante considerar que los parámetros utilizados en la calibración de este método fueron desarrollados específicamente para las condiciones particulares de la cuenca y el estado de los datos disponibles en ese momento. Es posible que estos parámetros necesiten ser actualizados periódicamente, requiriendo nuevas calibraciones para mantener la precisión del modelo.

6.3 Recomendaciones:

Al término de este estudio, se seleccionó el modelo Germany AWI-CM-1-1-MR como el más favorable. Por lo tanto, se recomienda utilizar este modelo para estudios posteriores que compartan las mismas características y condiciones descritas en este documento.

A pesar de que en el período analizado se consideraron los escenarios críticos presentados entre los años 1960 y 1990, se sugiere que las instituciones pertinentes inviertan en un estudio más detallado, tomando en cuenta escenarios aún más desfavorables dentro de la cuenca. Esto permitirá obtener resultados mucho más cercanos a la realidad en términos de predicción.

Es necesario fortalecer la robustez del modelo, especialmente en el proceso de reescalado de la precipitación. En este estudio, las distribuciones de probabilidad de la lluvia no se consideraron debido a que el enfoque del trabajo es exploratorio. Por lo tanto, el análisis

detallado de estas distribuciones queda fuera del alcance de este proyecto, aunque es un área importante para futuras investigaciones que podrían mejorar la precisión del modelo. Aunque este estudio es exploratorio, ya se observa un impacto en los caudales futuros, lo que puede representar una alerta importante a tener en cuenta.

Referencias

Achite, M., Jehanzaib, M., Elshaboury, N., & Kim, T. W. (2022). Evaluation of Machine Learning Techniques for Hydrological Drought Modeling: A Case Study of the Wadi Ouahrane Basin in Algeria. *Water 2022, Vol. 14, Page 431*, *14*(3), 431. https://doi.org/10.3390/W14030431

Matich Jorge. (2001). *Cátedra: Informática Aplicada a la Ingeniería de Procesos-Orientación I Redes Neuronales: Conceptos Básicos y Aplicaciones*.

Box, G. (2013). Box and Jenkins: Time Series Analysis, Forecasting and Control. *A Very British Affair*, 161–215. https://doi.org/10.1057/9781137291264_6

CAF. (2014). *Vulnerability Index to climate change in the Latin American and Caribbean Region*.

Costabile, P., Morales-Hernández, M., Bellos, V., Savino, M., Todaro, V., Maranzoni, A., & D'oria, M. (2023). Combining Hydrological Modeling and Regional Climate Projections to Assess the Climate Change Impact on the Water Resources of Dam Reservoirs. *Water 2023, Vol. 15, Page 4243*, *15*(24), 4243. https://doi.org/10.3390/W15244243

Dinamarca, A. (2018). *Aprendizaje y Análisis de Redes Neuronales Artificiales Profundas*.

Download Matlab Captain Toolbox – System Identification, Time Series Analysis, Forecasting and True Digital Control. (s/f). Recuperado el 19 de junio de 2024, de https://wp.lancs.ac.uk/captaintoolbox/

Draper, N. R., & Smith, H. (2014). Applied regression analysis. *Applied Regression Analysis*, 1–716. https://doi.org/10.1002/9781118625590

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics*, *49*(1), 92. https://doi.org/10.2307/1937887

Ferreira, C., Manuel Abreu Freire Diogo, A., Vinicio Carrera-Villacrés, D., Rodríguez-Espinosa, F., & Toulkeridis, T. (2023). Potential Solutions for the Water Shortage Using Towers of Fog Collectors in a High Andean Community in Central Ecuador. *Sustainability 2023, Vol. 15, Page 9237*, *15*(12), 9237. https://doi.org/10.3390/SU15129237

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://books.google.com.ec/books?hl=es&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=Deep+Learning%22+por+Ian+Goodfellow,+Yoshua+Bengio+y+Aaron+Courville&ots=MON6bvtzUT&sig=3drpVtC_lfi-jRUGD4B1dDAP8ao&redir_esc=y#v=onepage&q&f=false

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geosci. Model Dev*, *15*, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Hsu, K. L., Moradkhani, H., & Sorooshian, S. (2009). A sequential Bayesian approach for hydrologic model selection and prediction. *Water Resources Research*, *45*(1). https://doi.org/10.1029/2008WR006824

Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. *Feature Engineering and Selection: A Practical Approach for Predictive Models*, 1–297. https://doi.org/10.1201/9781315108230/FEATURE-ENGINEERING-SELECTION-MAX-KUHN-KJELL-JOHNSON

Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., & Caloiero, T. (2023). Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow Prediction Models. *Water 2023, Vol. 15, Page 2572*, *15*(14), 2572. https://doi.org/10.3390/W15142572

Lee, T., & Singh, V. P. (2019). *Statistical Downscaling for Hydrological and Environmental Applications*.

Mendoza, D. E., Samaniego, E. P., Mora, D. E., Espinoza, M. J., & Campozano, L. V. (2019). Finding teleconnections from decomposed rainfall signals using dynamic harmonic regressions: a Tropical Andean case study. *Climate Dynamics*, *52*(7–8). https://doi.org/10.1007/s00382-018-4400-3

Mora, D. E., Wyseure, G., & Willems, P. (2014). *GAP FILLING BASED ON A QUANTILE PERTURBATION FACTOR TECHNIQUE*.

Núñez Mejía, S. X., Mendoza Paz, S., Tabari, H., & Willems, P. (2023). Climate change impacts on hydrometeorological and river hydrological extremes in Quito, Ecuador. *Journal of Hydrology: Regional Studies*, *49*. https://doi.org/10.1016/j.ejrh.2023.101522

O'neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev*, *9*, 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Prechelt, L. (1998). *Early Stopping - But When?* 55–69. https://doi.org/10.1007/3-540-49430-8_3

Ruiz-Ordóñez, D. M., Solano-Correa, Y. T., Maysels, R., & Figueroa-Casas, A. (2023). Land-Use Dynamics and Water Quality in Andean Basins. *Sustainability 2023, Vol. 15, Page 15965*, *15*(22), 15965. https://doi.org/10.3390/SU152215965

Salas, J. D. (Jose D.), Delleur, J., Yevjevich, V., & Lane, W. (1980). Applied Modeling of Hydrologic Time Series. *Water Resources Publications*, 484.

Searcy, J. K., Hardison, C. H., & Langbein, W. B. (s/f). *Double-Mass Curves With a section Fitting Curves to Cyclic Data Manual of Hydrology: Part 1. General Surface-Water Techniques*.

Sobre Copernicus | Copernicus. (s/f). Recuperado el 20 de julio de 2024, de https://www.copernicus.eu/es/sobre-copernicus

Thiessen, A. (1911). *PRECIPITATION AVERAGES FOR LARGE AREAS in: Monthly Weather Review Volume 39 Issue 7 (1911)*. https://doi.org/https://doi.org/10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2

Tibshiranit, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Statist. Soc. B*, *58*(1), 267–288. https://academic.oup.com/jrsssb/article/58/1/267/7027929

Wingfield, S., Martínez-Moscoso, A., Quiroga, D., & Ochoa-Herrera, V. (2021). Challenges to Water Management in Ecuador: Legal Authorization, Quality Parameters, and Socio-Political Responses. *Water 2021, Vol. 13, Page 1017*, *13*(8), 1017. https://doi.org/10.3390/W13081017

Wood, E. F., Szöllösi-Nagy, A., & International Institute for Applied Systems Analysis. (1980). *Real-Time Forecasting/Control of Water Resource Systems; Selected Papers from an IIASA Workshop, October 18-21,1976*. 330.

Young, P. (1998). Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling & Software*, *13*(2), 105–122. https://doi.org/10.1016/S1364-8152(98)00011-5

Young, P. (2003). Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. *Hydrological Processes*, *17*(11), 2195–2217. https://doi.org/10.1002/hyp.1328

Young, P. C., & Beven, K. J. (1994). Data-based mechanistic modelling and the rainfall-flow non-linearity. *Environmetrics*, *5*(3), 335–363. https://doi.org/10.1002/ENV.3170050311

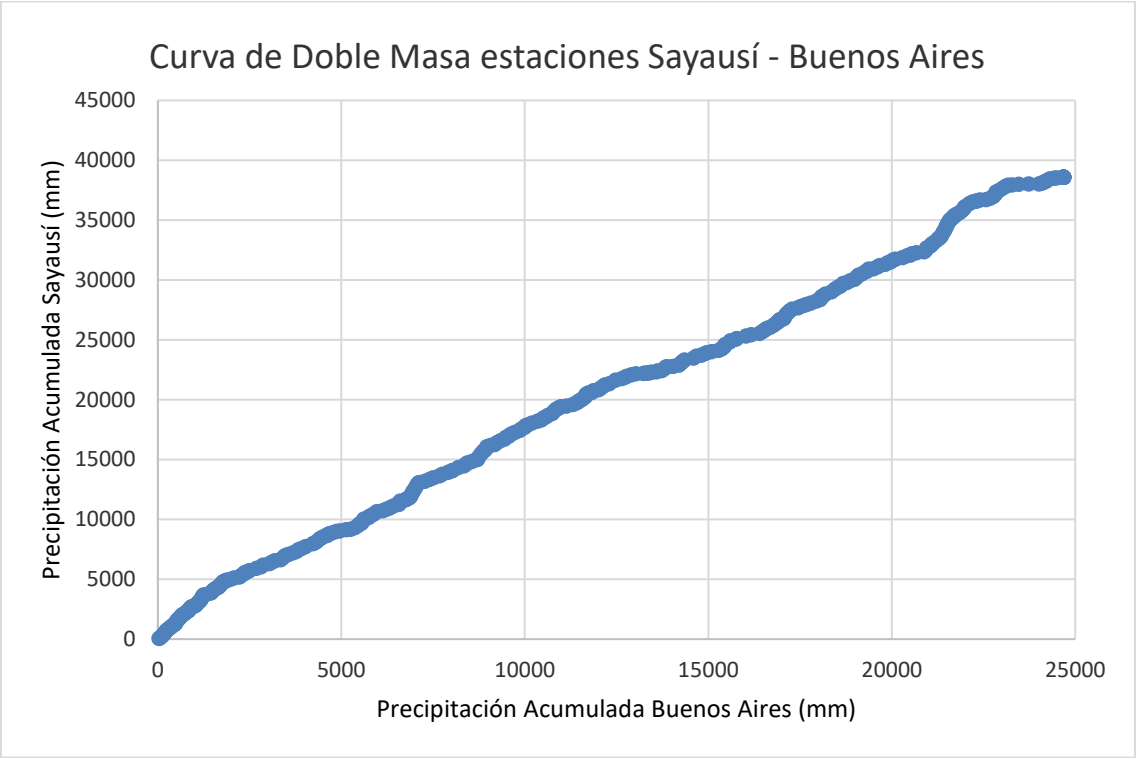
Young, P. C., Mckenna, P., & Bruun, J. (2001). Identification of non-linear stochastic systems by state dependent parameter estimation. *International Journal of Control*, *74*(18), 1837–1857. https://doi.org/10.1080/00207170110089824

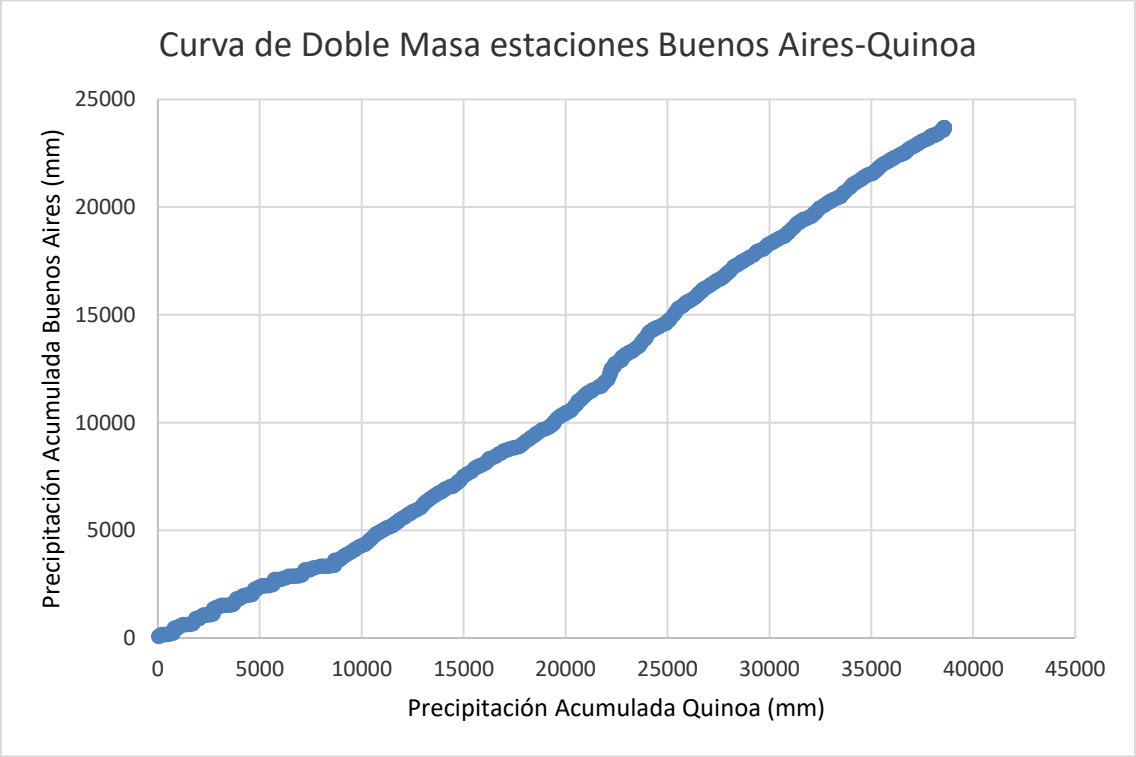
Zeybek, M. (2018). Nash-Sutcliffe Efficiency Approach For Quality Improvement. *Journal of Applied Mathematics and Computation (JAMC)*, *2*(2), 496–503. https://doi.org/10.26855/jamc.2018.11.001

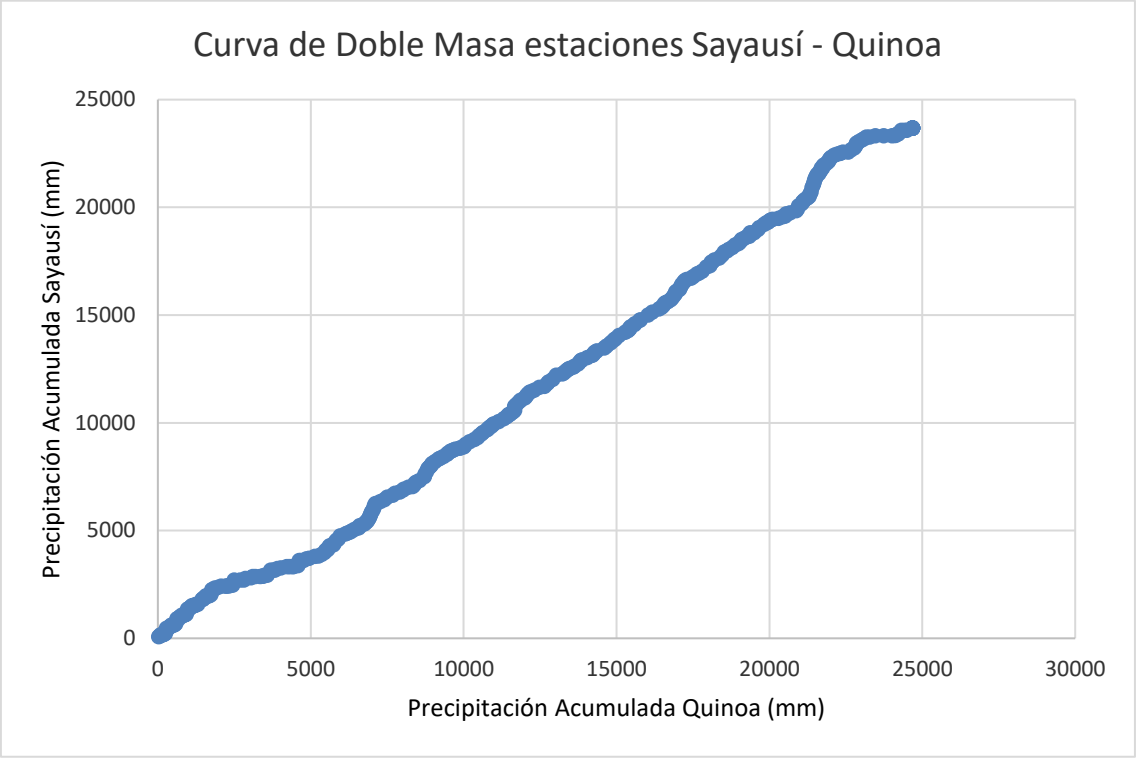
Zhou, Y., & Ding, F. (2020). Modeling Nonlinear Processes Using the Radial Basis Function-Based State-Dependent Autoregressive Models. *IEEE Signal Processing Letters*, *27*, 1600–1604. https://doi.org/10.1109/LSP.2020.3021925

Anexos

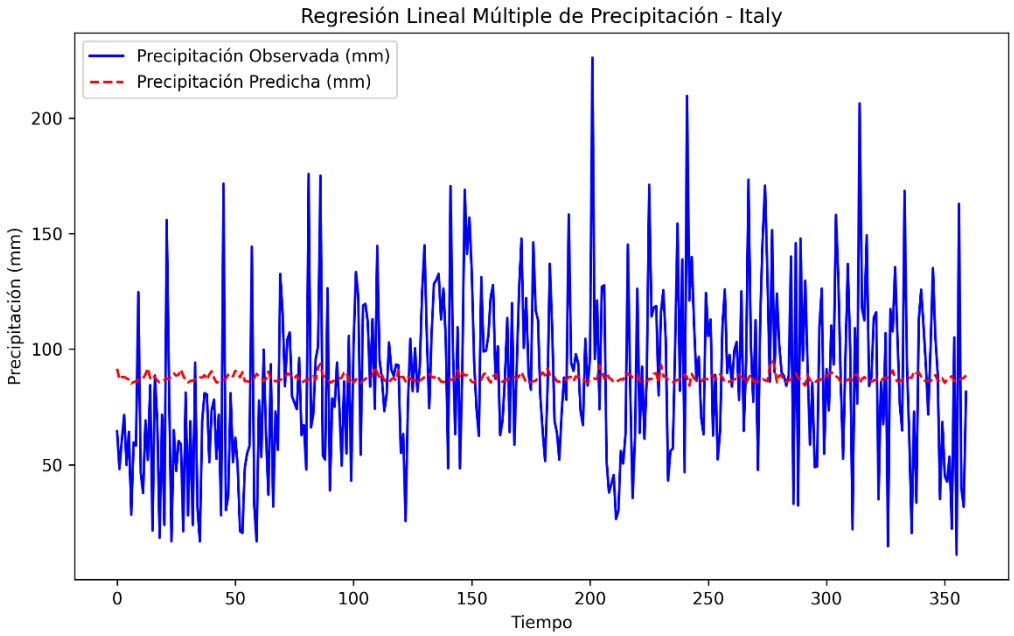
ANEXO A – CURVAS DE DOBLE MASA

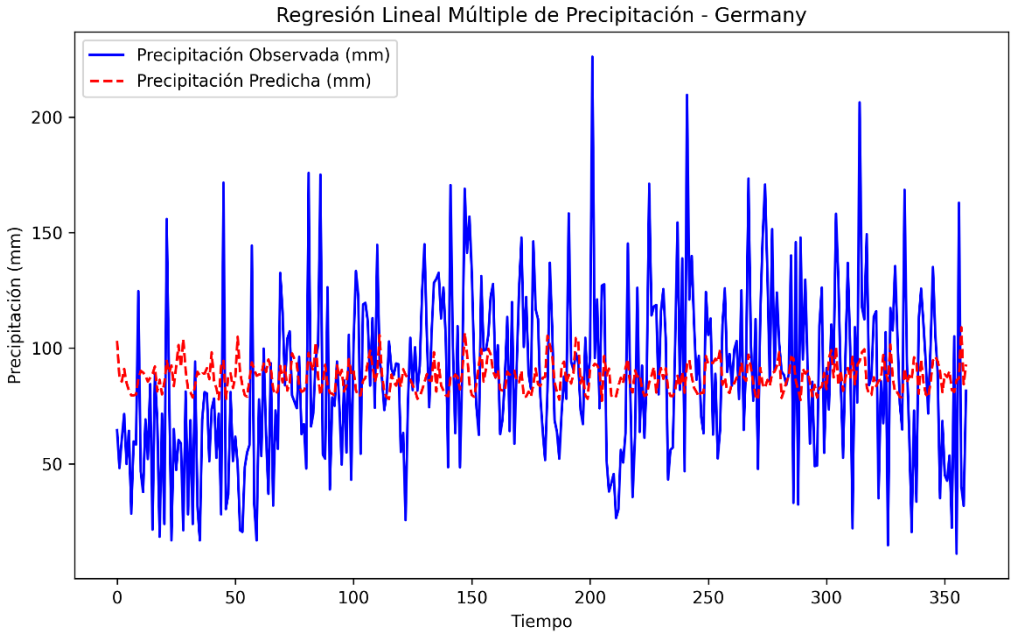


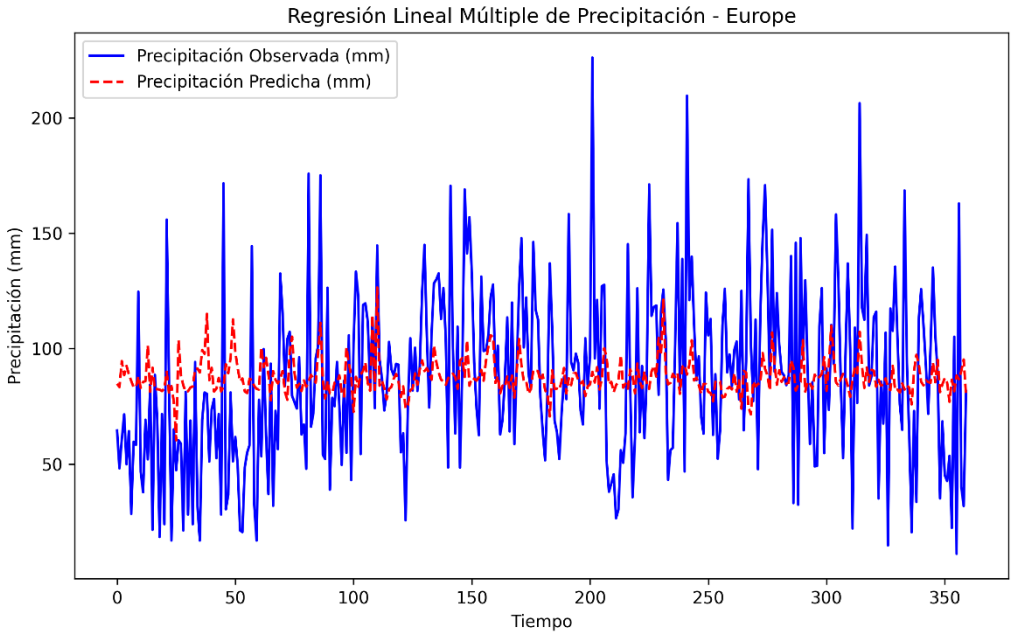


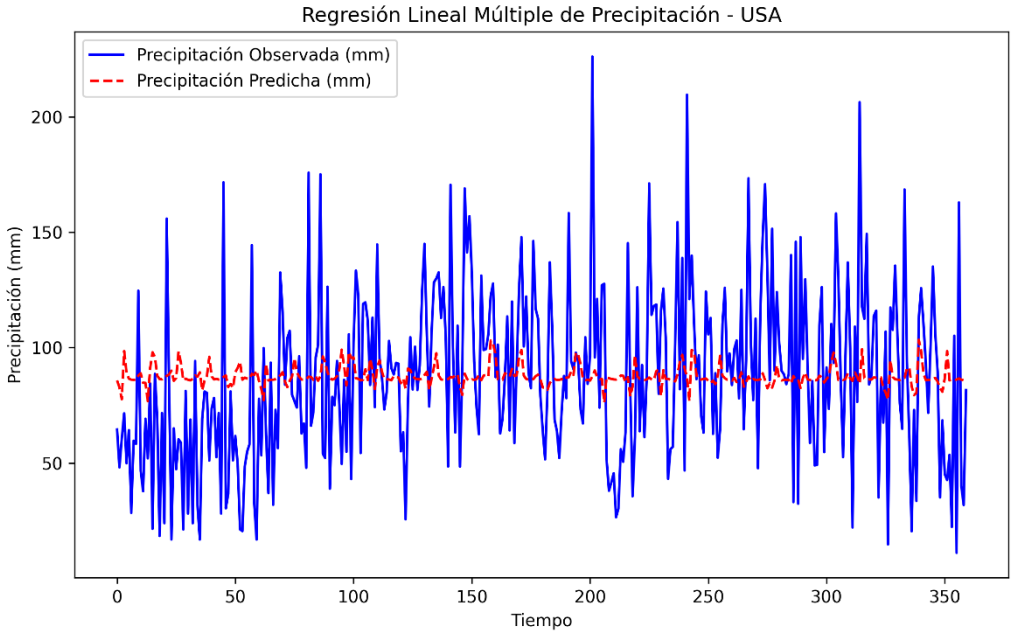

Curva de Doble Masa estaciones Sayausí - Quinoa
Precipitación Acumulada Sayausí (mm)
25000
20000
15000
10000
5000
0
0
5000
10000
15000
20000
25000
30000
Precipitación Acumulada Quinoa (mm)

ANEXO B – REGRESIONES LINEALES MÚLTIPLES

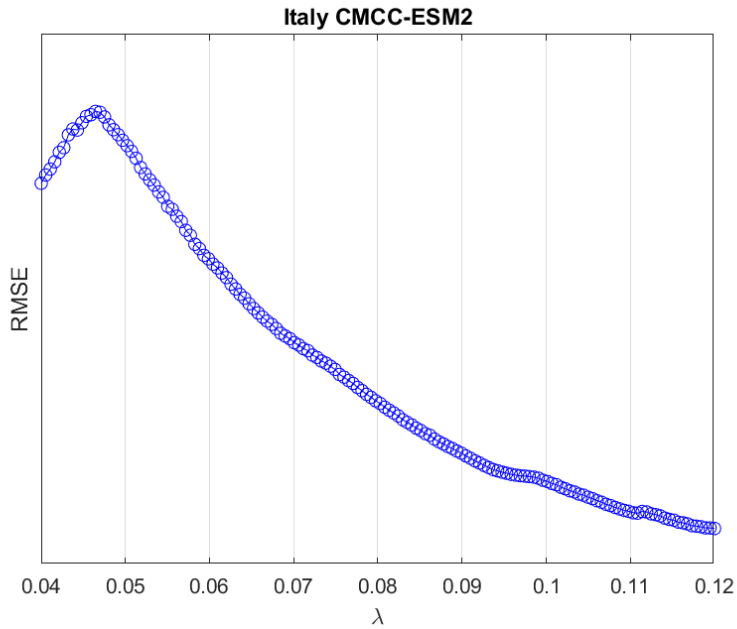


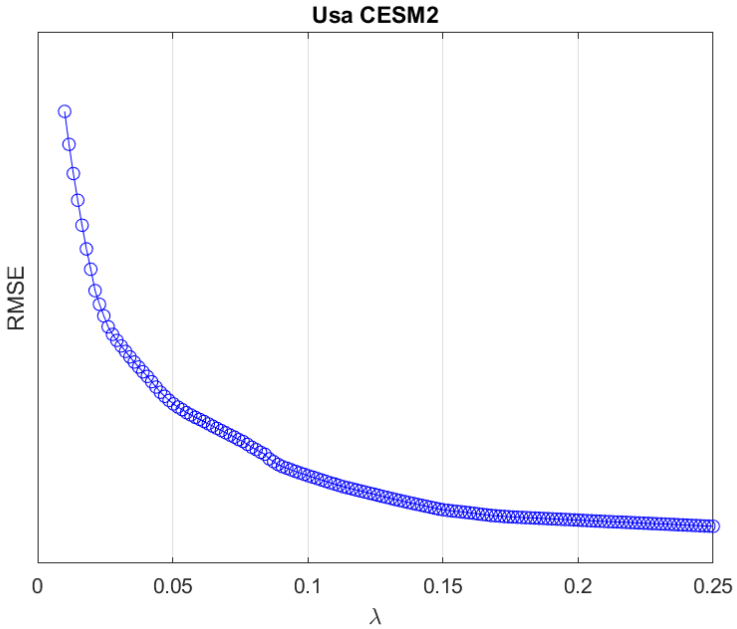


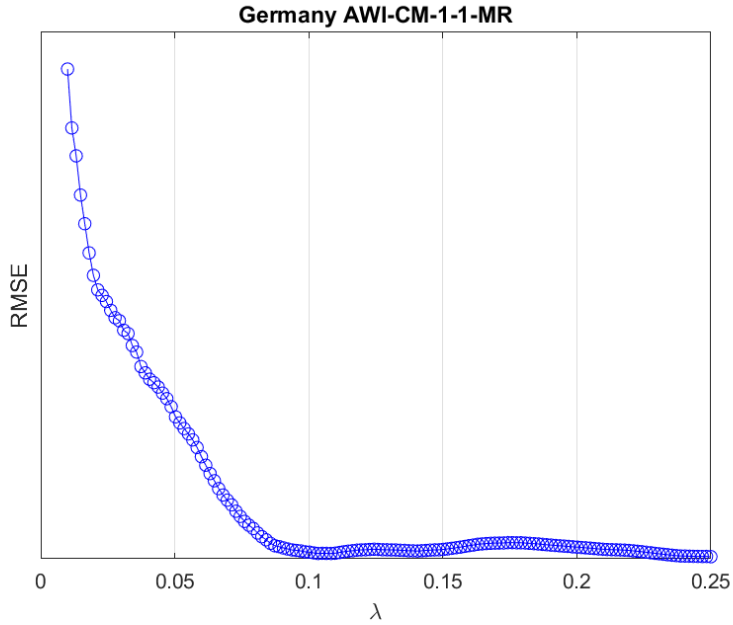

Regresión Lineal Múltiple de Precipitación - Europe
Precipitación Observada (mm)
Precipitación Predicha (mm)
Precipitación (mm)
200
150
100
50
0
50
100
150
200
250
300
350
Tiempo


Regresión Lineal Múltiple de Precipitación - USA
Precipitación Observada (mm)
Precipitación Predicha (mm)
Precipitación (mm)
200
150
100
50
0
50
100
150
200
250
300
350
Tiempo

ANEXO C – CURVAS RSME VS LAMBDA

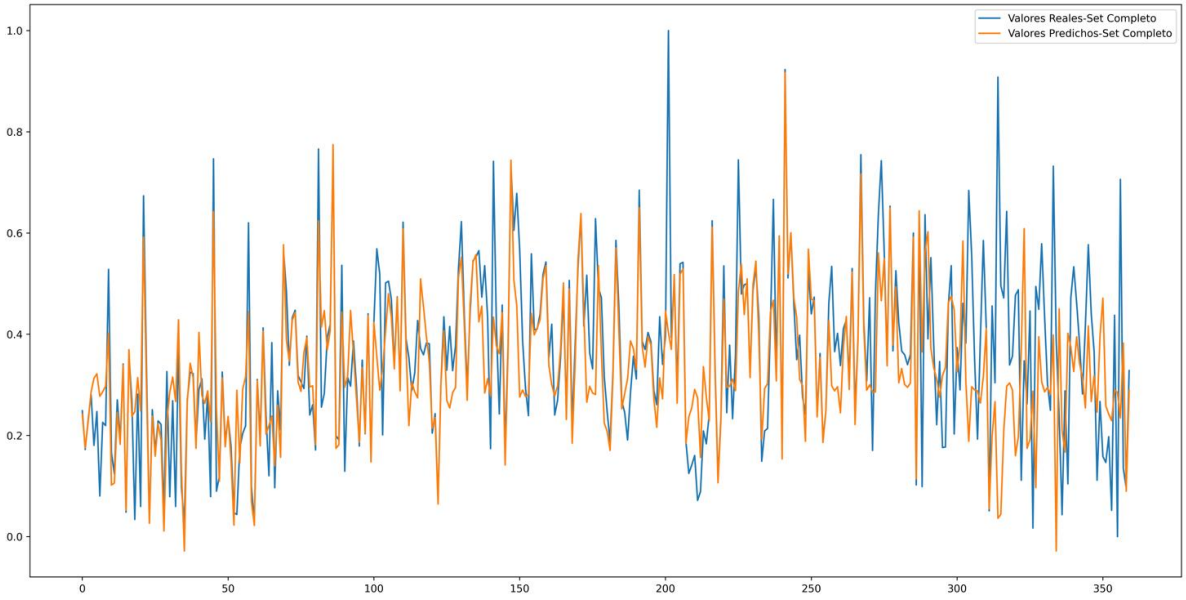




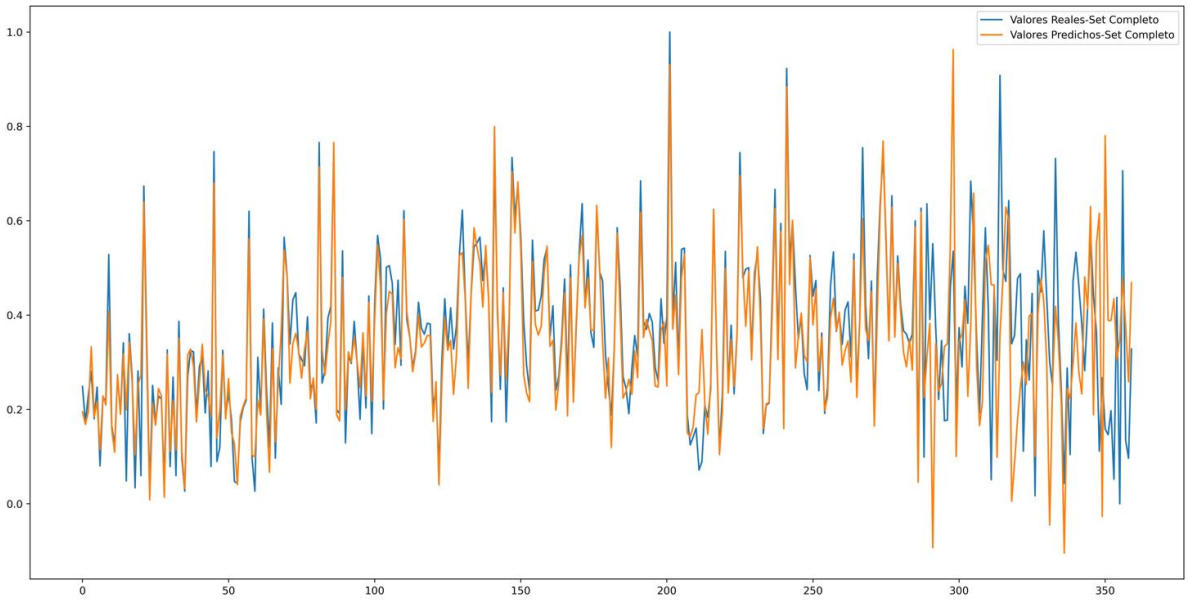

Germany AWI-CM-1-1-MR
RMSE
0
0.05
0.1
0.15
0.2
0.25
λ

ANEXO D – REDES NEURONALES DE MODELOS GCM SIN INTERACCIÓN

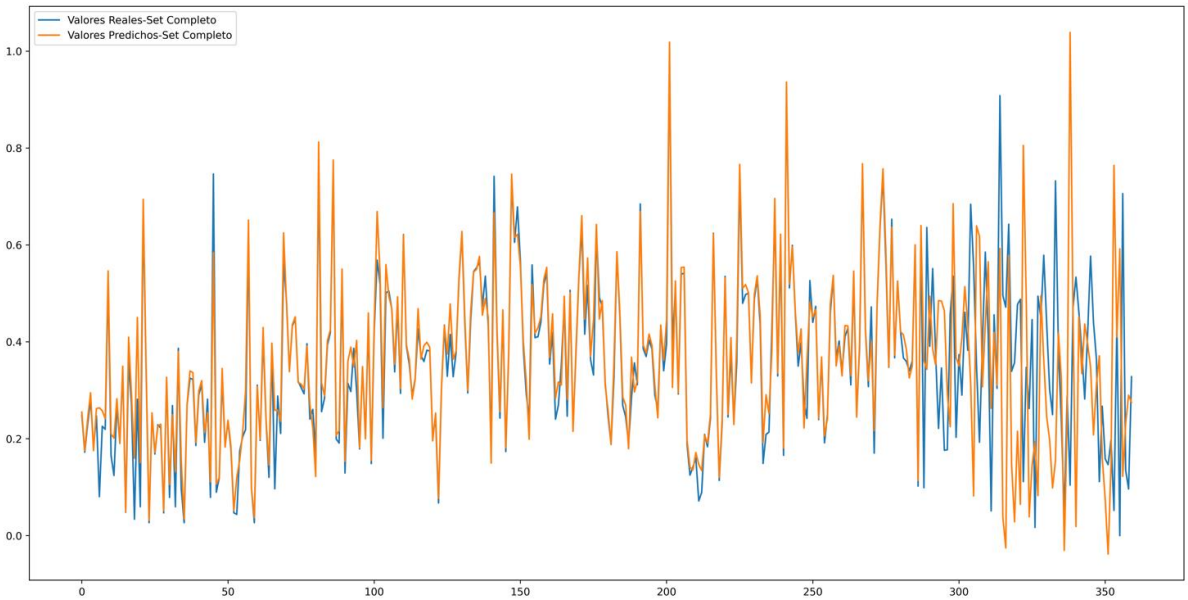
USA:



ITALY:

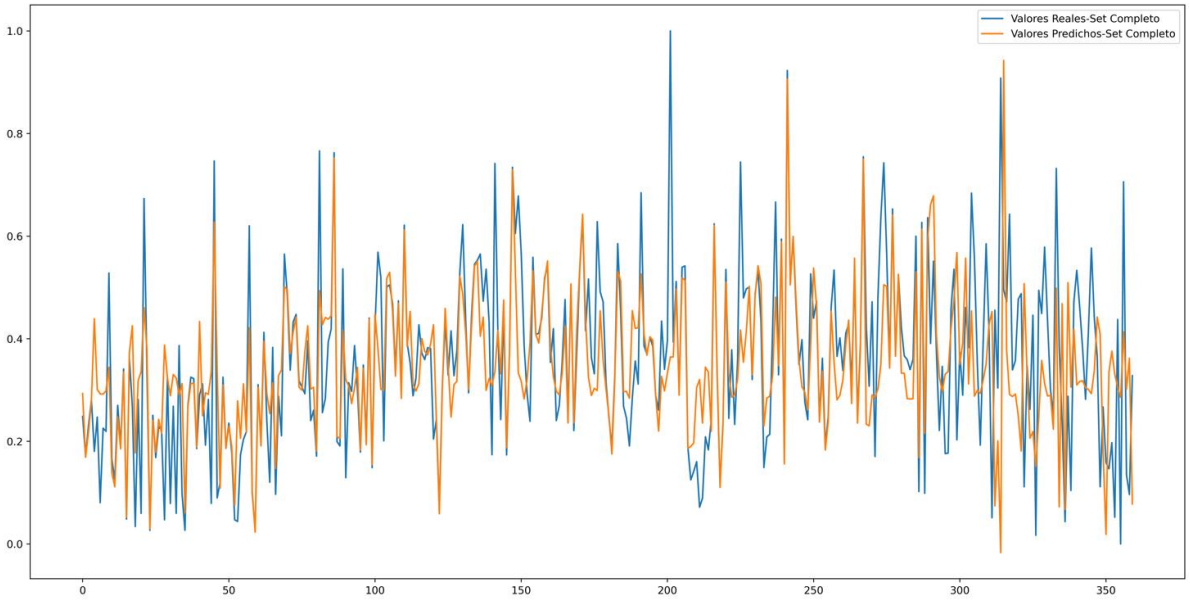


EUROPE:

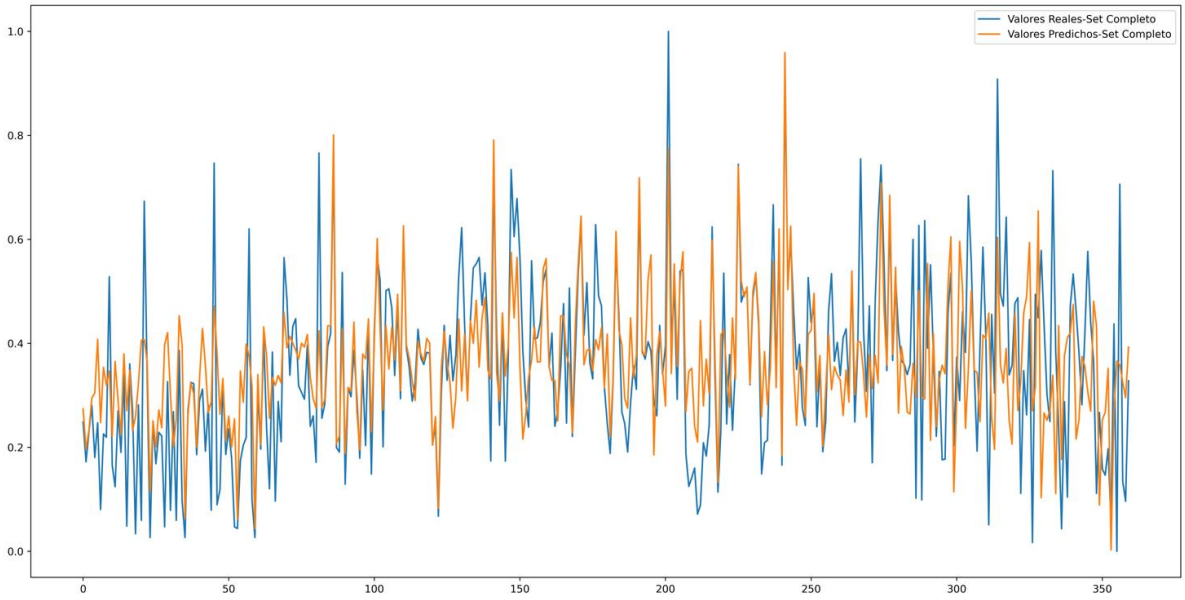

Valores Reales-Set Completo
Valores Predichos-Set Completo
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
350

ANEXO E – REDES NEURONALES DE MODELOS GCM CON INTERACCIÓN

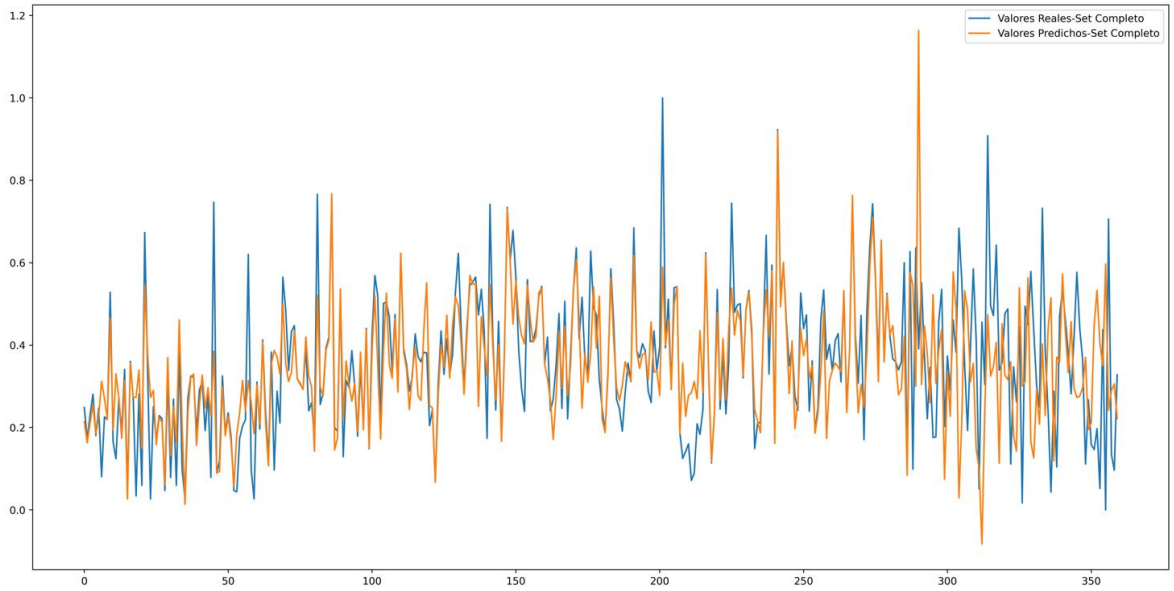
USA:



ITALY:



EUROPE:


Valores Reales-Set Completo
Valores Predichos-Set Completo
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
350