



Universidad del Desarrollo
Facultad de Ingeniería

"AquaTendencias Maipo: Una radiografía a las zonas de la cuenca "

POR: NICOLAS ANTONIO RIVERA PIÑONES

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

PROFESOR GUÍA:

(Sr.) MAURICIO HERRERA MARIN

Diciembre 2023

LA SERENA

Tabla de contenido

Resumen	2
Abstract	4
Introducción	5
Trabajo Relacionado.....	7
Hipótesis y Objetivos	11
Hipótesis	11
Objetivo General.....	11
Objetivos Específicos.....	11
Limitaciones	12
Datos y Metodología	13
Datos.....	13
Descripción de los Datos	13
Definición de “Dryspell”	14
Como Ayuda el Estudio de los Dryspell	14
Variables Utilizadas.....	15
Localización Geográfica	15
Procesamiento y Preparación de Datos	18
Conjunto de Datos Utilizado.....	18
Metodología.....	20
Análisis Exploratorio de Datos (EDA).....	20
Test no paramétrico de Mann-Kendall.....	20
Índice de Intensidad de Precipitación Diaria Simple (SDII)	23
Índice de Precipitación Estandarizado(SPI)	25
Búsqueda de Correlación de Temperatura y Dryspell.....	27
Modelo de predicción DeepAR.....	28
Resultados.....	30
Análisis exploratorio de datos (EDA)	30
Interpretación de los Resultados del Test de Mann-Kendall	37
Interpretación de los resultados del SDII	40
Interpretación de los Resultados del Índice Estandarizado de Precipitación SPI para la Estación Húmeda. ...	45
Longitud de Dryspell.....	48
Correlación de variable temperatura y “DrySpell”	50
Conclusiones.....	57
Bibliografía	60

Resumen

El presente estudio, enfocado en la cuenca del río Maipo en la Región Metropolitana de Santiago, Chile, aborda la creciente preocupación por la seguridad hídrica en un contexto marcado por la variabilidad y el cambio climático. La investigación se centra en la posibilidad de que los patrones climáticos históricos y los datos meteorológicos como las precipitaciones puedan ser indicativos para la predicción de períodos secos (dry-spell) en la cuenca del río Maipo, reconociendo la alta variabilidad y complejidad inherente a los sistemas meteorológicos. Los dryspell, definidos como períodos prolongados sin precipitaciones significativas menor a 1mm, son de alta importancia en áreas donde la agricultura es altamente dependiente de la lluvia, estos períodos secos son cruciales para entender y anticipar sequías más severas.

Este proyecto se enfoca en la Región del Maipo, que comprende 14.939 km², incluye 23 estaciones fluviométricas, y se inscribe en un contexto desafiante debido a una sequía persistente y una demanda creciente de recursos hídricos que son cada vez más limitados. Este escenario subraya la necesidad de planificación exhaustiva y estrategias adaptativas en la gestión de recursos hídricos, actividades que son vitales para el sector agrícola y entidades privadas que requieren un manejo eficiente y sostenible del agua.

La metodología utilizada en este estudio combina el análisis de datos hidroclimáticos con modelos estadísticos y de aprendizaje profundo para explorar tendencias y patrones en las precipitaciones durante más de 20 años. El uso de modelos estadísticos y de Deep Learning brindan una perspectiva cuantitativa esencial para la toma de decisiones informadas, utilizando por ejemplo herramientas como el Test de Mann-Kendall y el cálculo de índices como el SPI y SDII sobre series de tiempo de precipitaciones.

Este trabajo reconoce la complejidad de un sistema que entrelaza lo hidroclimático, social, ambiental y territorial, proponiendo un enfoque sistémico para abordar estos aspectos y alinear los resultados con las necesidades de adaptación al cambio climático y la optimización de la gestión del agua. Los hallazgos y recomendaciones presentados buscan contribuir a la construcción de un futuro más resiliente y seguro en términos hídricos para la región más poblada de Chile

Abstract

This study, set in the Maipo River Basin within the Metropolitan Region of Santiago, Chile, addresses the growing concerns over water security amidst climate variability and change. It investigates the potential of historical climatic patterns and meteorological data, such as rainfall, to predict dry spells—defined as prolonged periods with less than 1mm of precipitation—critical for regions heavily reliant on rain-fed agriculture.

Covering a 14,939 km² Mediterranean region with 23 hydrometric stations, the project unfolds against the backdrop of persistent drought and escalating demands for increasingly scarce water resources. This underscores the imperative for comprehensive planning and adaptive strategies in water resource management, positioning this work as a pivotal tool for agricultural sectors and private entities aiming for efficient and sustainable water use.

Methodologically, the study melds hydroclimatic data analysis with statistical and deep learning models to discern precipitation trends and patterns over two decades. Despite challenges in deep learning modeling, the findings provide a quantitative perspective vital for informed decision-making, employing tools such as the Mann-Kendall Test and indices like SPI and SDII.

Recognizing the complexity of intertwining hydroclimatic, social, environmental, and territorial systems, the work advocates a systemic approach to address these facets, aligning outcomes with climate adaptation needs and water management optimization. The insights and recommendations aim to contribute to a more resilient and secure hydrological future for Chile's most populous region.

Introducción

En el escenario actual, donde la seguridad hídrica se ha convertido en una preocupación primordial para la sustentabilidad de las comunidades, la cuenca del río Maipo emerge como un eje vital en la dinámica socioeconómica y ambiental de la Región Metropolitana de Santiago, Chile. Este estudio surge del intento de predecir episodios de sequía (dry-spell) basándonos en patrones climáticos históricos y datos meteorológicos. Los “dryspell” se refiere a un período prolongado sin precipitaciones significativas. Para este estudio se define un periodo dry-spell como un registro diario con una precipitación menor a 1mm, esta definición es aceptada por la literatura, pero puede tener excepciones dependiendo el contexto geográfico. Los dry spells son especialmente significativos en áreas donde la agricultura depende en gran medida de la lluvia y pueden ser un precursor de sequías más severas si se extienden por períodos más largos, un desafío que se ha intensificado ante la variabilidad y el cambio climático. Este estudio se desarrolla en la cuenca del Maipo, una región mediterránea que se extiende desde las alturas de la cordillera de los Andes hasta su desembocadura en el océano Pacífico, cubriendo un área de 14.939 km² monitoreada por 23 estaciones fluviométricas. La investigación se inscribe en un contexto donde la variabilidad y el cambio climático imponen desafíos significativos a la disponibilidad hídrica, exacerbados por una mega sequía persistente y una demanda creciente de un recurso cada vez más escaso.

La situación crítica de abastecimiento de agua potable en la región, marcada por un margen estrecho entre la oferta y la creciente demanda, pone de manifiesto la necesidad de una planificación exhaustiva y la implementación de estrategias adaptativas para la gestión de recursos hídricos. Este trabajo se propone como una herramienta de instrucción para el sector

agrícola y entidades privadas, con el fin de promover un manejo más eficiente y consciente del recurso hídrico.

La investigación se fundamenta en el análisis de datos hidroclimáticos y la aplicación de modelos estadísticos y de aprendizaje profundo, para explorar tendencias y patrones en las precipitaciones a lo largo de más de 20 años.

A pesar de que los modelos con técnicas de Deep Learning empleados aquí no lograron resultados concluyentes, la experimentación con estos métodos proporciona una base sobre la cual futuras investigaciones pueden edificar. Así mismo, la aplicación de pruebas no paramétricas Test de Mann-Kendall, modelos estadísticos y el cálculo de índices como el SPI y SDII brindan una perspectiva cuantitativa sobre las tendencias de la precipitación, esenciales para la toma de decisiones informadas.

Los problemas de seguridad hídrica en la cuenca del Maipo reflejan la complejidad de un sistema que entrelaza aspectos hidroclimáticos, sociales, ambientales y territoriales. El presente capstone project reconoce esta complejidad y propone un enfoque sistémico para abordarla, alineando los resultados obtenidos con las necesidades de adaptación al cambio climático y la optimización de la gestión del agua tanto para los organismos del Estado como para la empresa privada.

En este trabajo se detalla la metodología adoptada, los hallazgos resultantes y las recomendaciones pertinentes, con el propósito de contribuir a la construcción de un futuro más resiliente y seguro en términos hídricos para la región que alberga la mayor concentración de habitantes de Chile.

Trabajo Relacionado

(1) Impact of the dry-day definition on Mediterranean extreme dry-spell analysis

Rivoire, P., Tramblay, Y., Neppel, L., Hertig, E., & Vicente-Serrano, S. M. (2019). Impact of the dry-day definition on Mediterranean extreme dry-spell analysis. *Natural Hazards and Earth System Sciences*, 19(8), 1629–1638. <https://doi.org/10.5194/nhess-19-1629-2019>

Introducción al Estudio

En el estudio (1) los autores destacan que la región mediterránea, caracterizada por una variabilidad interanual significativa en los patrones de precipitación, se enfrenta a episodios de sequía severa que afectan tanto a la producción agrícola como a los recursos hídricos, especialmente durante la temporada húmeda (invierno). Estudios previos sugieren una tendencia hacia condiciones más cálidas y secas en la región, que se espera que se intensifiquen con el cambio climático futuro.

Metodologías de Análisis de Sequía

Existen diversas metodologías para analizar las sequías, incluyendo índices de sequía y modelización de la frecuencia y duración de periodos secos. Se define meteorológicamente un periodo seco como una secuencia de días consecutivos secos, con precipitación inexistente o por debajo de un umbral específico. Aunque estos periodos no determinan la severidad de la sequía, son útiles para evaluar las diferencias espaciales en el riesgo de sequía y las posibles tendencias asociadas al cambio climático.

Impacto de la Definición de Día Seco

La elección del umbral de precipitación diaria (DRT) es crucial para relacionar el riesgo de periodos secos con impactos en diferentes sectores. Los umbrales fijos no siempre reflejan las

condiciones reales del terreno, dado que la evaporación varía a lo largo del año y entre distintas ubicaciones. La demanda atmosférica de evaporación (AED) puede modular significativamente la precipitación neta disponible para las plantas, afectando los niveles de estrés hídrico.

Objetivo del Estudio Relacionado Presente

Evaluar la influencia de diferentes umbrales de precipitación diaria en la estimación del riesgo de periodos secos extremos estacionales en el Mediterráneo. Se compara un umbral fijo de 1 mm diario, comúnmente utilizado, con un umbral variable definido por el AED.

Tests Estadísticos Utilizados

Para detectar tendencias en las series temporales de las estaciones, se utilizó el test no paramétrico de Mann-Kendall, junto con el método de preblanqueo libre de tendencia y el método de la tasa de falsos descubrimientos para distinguir entre tendencias significativas a nivel local y regional. Para comparar diferentes distribuciones de periodos secos extremos, se empleó el test de Anderson-Darling.

Conclusiones del Estudio

El análisis reveló que, a pesar de las tendencias climáticas en precipitación y evapotranspiración, no se detectan tendencias significativas en el riesgo de periodos secos extremos estacionales en la mayoría de las áreas. Se sugiere que el uso de un umbral fijo subestima el riesgo de periodos secos extremos en comparación con un umbral variable que representa la evapotranspiración durante la temporada de verano extendida. Los resultados subrayan la importancia de considerar la demanda de agua atmosférica al analizar periodos secos, especialmente si el objetivo es relacionarlos con impactos agrícolas.

Uso del índice de precipitación estandarizado para la evaluación de la intensidad de la sequía en la provincia de Holguín

Villazón Gómez, Juan Alejandro, Noris Noris, Pavel, & García Reyes, Roberto Alejandro.

(2021). Uso del índice de precipitación estandarizado para la evaluación de la intensidad de la sequía en la provincia de Holguín. *Idesia (Arica)*, 39(4), 103-

109. <https://dx.doi.org/10.4067/S0718-34292021000400103>

Monitoreo y Evaluación de Sequías

La evaluación de la sequía es un componente crítico de la gestión de recursos hídricos, especialmente en regiones propensas a variaciones extremas en las precipitaciones. El índice de precipitación estandarizado (SPI) se destaca como una herramienta esencial para la detección y monitorización de sequías agrícolas, meteorológicas e hidrológicas, debido a su simplicidad y efectividad. Este índice requiere registros de precipitaciones durante un período de al menos 30 años para proporcionar una evaluación precisa (Campos, 2014; Esquivel et al., 2017).

Aplicación del SPI en la Provincia de Holguín

En un estudio llevado a cabo en la provincia de Holguín, Cuba, se utilizó el SPI para caracterizar eventos de sequía sobre diferentes escalas de tiempo, desde 1 hasta 24 meses. Esta adaptabilidad temporal permite una comprensión detallada de las sequías de corta y larga duración, lo cual es vital para la gestión agrícola y la planificación de recursos hídricos. Los datos recopilados entre 1975 y 2017 en cuatro estaciones meteorológicas distintas proporcionaron una perspectiva profunda sobre la frecuencia e intensidad de los eventos de sequía en la región.

Resultados Clave del SPI en la Evaluación de Sequías

Los resultados de la investigación revelaron que las sequías en las zonas llanas ocurrieron con mayor frecuencia en comparación con las regiones montañosas. Se identificaron períodos significativos de sequía que coincidieron en las cuatro estaciones entre 1999 y 2006, con la severidad de los eventos de sequía impactando de manera diversa las reservas de agua y los ecosistemas locales. El estudio enfatiza la importancia de considerar las diferencias topográficas y de suelo en la evaluación de sequías y en la planificación de la gestión del agua.

Implicaciones para la Investigación Futura

Los hallazgos subrayan la necesidad de integrar enfoques como el SPI en la monitorización y gestión de sequías a largo plazo. Estas metodologías podrían ser particularmente relevantes para áreas como la cuenca del río Maipo, donde las variaciones en las precipitaciones y los eventos extremos de sequía podrían tener impactos significativos en los recursos hídricos y la agricultura.

Hipótesis y Objetivos

Hipótesis

Los patrones climáticos históricos y los datos meteorológicos pueden ser utilizados para la predicción de períodos secos (dry-spell) en la cuenca del río Maipo, reconociendo la alta variabilidad y complejidad inherente a los sistemas meteorológicos.

Objetivo General

Evaluar la relación entre patrones climáticos históricos, datos meteorológicos y la ocurrencia de días secos en la cuenca del río Maipo, investigando la capacidad predictiva de modelos estadísticos y de Deep Learning para pronosticar eventos de sequía mediante el uso de series de tiempo de precipitaciones, con el fin de contribuir al conocimiento y manejo eficiente del recurso hídrico.

Objetivos Específicos

- Documentar y analizar los intentos de aplicar modelos estadísticos y de aprendizaje profundo para la predicción de días secos, incluyendo las barreras encontradas y posibles soluciones.
- Evaluar la variabilidad y la incertidumbre en los modelos predictivos del clima y su impacto en la planificación de recursos hídricos en el sector privado y agrícola.

Limitaciones

Complejidad de los Modelos Meteorológicos

La precisión de los modelos de predicción de períodos secos (dry spells) está significativamente influenciada por la complejidad inherente a los sistemas meteorológicos. Los patrones climáticos son el resultado de interacciones dinámicas y multifactoriales que pueden ser difíciles de modelar con precisión, especialmente con las limitaciones computacionales y metodológicas a nivel de tesis de grado.

Disponibilidad y Calidad de Datos

La disponibilidad de datos meteorológicos históricos de alta calidad y alta resolución es esencial para entrenar modelos predictivos robustos. Sin embargo, puede haber lagunas en los datos, errores de medición y variaciones en la calidad de los datos a lo largo del tiempo y entre diferentes ubicaciones geográficas.

Recursos y Expertise

El desarrollo de modelos predictivos avanzados generalmente requiere la colaboración de un equipo multidisciplinario con expertos en meteorología, climatología, así como recursos computacionales significativos, los cuales pueden estar más allá del alcance de una persona capstone project.

Datos y Metodología

Datos

Descripción de los Datos

Los datos para este estudio fueron obtenidos del Center for Climate and Resilience Research (CR2), a través de su sitio web cr2.cl. La base de datos específica utilizada es parte del proyecto CAMELS-CL, detallado en el artículo de Alvarez-Garreton et al., "The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset", publicado en Hydrology and Earth System Sciences (<https://doi.org/10.5194/hess-22-5817-2018>).

Se utilizaron variadas librerías que ayudan en la investigación tales como:

- Pandas
- Numpy
- Scikitlearn
- Matplotlib.pyplot
- Seaborn
- Ipywidgets

Definición de “Dryspell”

Se refiere a un período prolongado sin precipitaciones significativas. Para este estudio se define un periodo dry-spell como un registro diario con una precipitación menor a 1mm, esta definición es aceptada por la literatura pero puede tener excepciones dependiendo el contexto geográfico.

Como Ayuda el Estudio de los Dryspell

Planificación y Gestión de Riesgos

Comprender la frecuencia, duración e intensidad de los dry spells permite a los planificadores y gestores de recursos hídricos desarrollar estrategias de manejo y mitigación más efectivas, preparándose mejor para estos eventos.

Mejora en la Agricultura

Los agricultores pueden beneficiarse de la información sobre los períodos secos para planificar la siembra, el riego y la cosecha de manera más efectiva, lo que puede llevar a una mayor eficiencia en el uso del agua y mejor rendimiento de los cultivos.

Investigación y Desarrollo

Aportan información valiosa para la investigación científica, ayudando a entender mejor los patrones climáticos y sus efectos. Esto puede conducir al desarrollo de nuevas tecnologías y prácticas para enfrentar los desafíos del cambio climático.

Concienciación y Educación

La sensibilización sobre los impactos y la gestión de los dry spells es crucial para fomentar el uso sostenible del agua y la adaptación al cambio climático a nivel comunitario y nacional.

Variables Utilizadas

Aunque la base de datos de CR2 contiene una amplia gama de variables meteorológicas, para este estudio se seleccionaron específicamente:

- Registros de precipitaciones diarias, que constituyen la variable principal de interés.
- Temperaturas diarias (máxima y mínima) como co- variables.

Estas variables fueron elegidas por su relevancia directa en la investigación de días secos y su impacto en la cuenca del río Maipo.

Periodo de Tiempo

La investigación se centra en el periodo entre enero de 1998 y diciembre de 2019. Este intervalo fue seleccionado por la consistencia y la completa disponibilidad de los registros diarios.

Localización Geográfica

El estudio se enfocó en la cuenca del río Maipo, una zona clave en el contexto hidrológico de la Región Metropolitana, Chile. Se incluyeron datos de 23 estaciones meteorológicas situadas en esta cuenca, siendo las siguientes:

Tabla 1

Se describe el código de la estación y su nombre respectivo para un mejor entendimiento.

CODIGO DE LA ESTACIÓN	NOMBRE DE LA ESTACIÓN
5701001	Río Maipo en las Hualtatas
5701002	Río Maipo en las Melosas
5702001	Rio Volcán en Queltehues
5703006	Estero Glaciar Echaurren
5704002	Río Maipo en San Alfonso

5705001	Río Colorado antes Junta Rio Olivares
5706001	Rio Olivares
5707002	Río Colorado antes junta Río Maipo
5710001	Río Maipo en el Manzano
5715001	Rio Paine en Longitudinal
5716001	Río Angostura en Valdivia de Paine
5720001	Rio Molina antes Junta San Francisco
5721001	Estero Yerba Loca ante Junta San Francisco
5721016	Río San Francisco antes Junta Estero Yerba Loca
5722001	Estero Arrayán en la Montosa
5722002	Río Mapocho en los Almendros
5730005	Zanjón de la Aguada
5730008	Quebrada Ramón en recinto Emos
5734001	Estero Polpaico en Chicauma
5737002	Río Mapocho Rinconada de Maipú
5741001	Estero Puangue en Boquerón
5746001	Estero Puangue en Ruta 78
5748001	Río Maipo en Cabimbao

Procesamiento y Preparación de Datos

Conjunto de Datos Utilizado

Se seleccionó el archivo CSV (archivo separado por coma) **precip_cr2met_mm_day.csv** y **tmax_cr2met_C_mon** de la base de datos CAMELS-CL, proporcionada por el Centro de Investigación del Clima y la Resiliencia (CR2). Este conjunto de datos se prefirió debido a su integridad y la consistencia de los registros diarios de precipitación, fundamentales para el análisis de series temporales.

Filtrado Geográfico y Temporal

Los datos se filtraron para incluir solo las estaciones ubicadas en la cuenca del río Maipo (identificadas por el prefijo "57") y cubrir el período desde enero de 1998 hasta diciembre de 2019. Esta selección se basó en que el prefijo caracteriza esa zona en este conjunto de datos y también mantiene una disponibilidad de datos consistentes durante el período de estudio.

División de Datos

Criterios de División

La división de los datos en conjuntos de entrenamiento, validación y prueba se basó en criterios temporales para preservar la naturaleza secuencial de la serie.

Importancia de la División Temporal

Esta división aseguró que el modelo aprendiera de patrones pasados y fuera evaluado en condiciones futuras, reflejando un escenario realista de pronóstico.

Manejo de Series Temporales

Descomposición Estacional

La aplicación de técnicas de descomposición estacional permitió desglosar los datos de precipitación en componentes de tendencia, estacionalidad y ruido. Esto ayudó a entender cómo las variaciones estacionales afectan los patrones de precipitación en la cuenca del río Maipo.

Consideraciones Climáticas

Dado el clima mediterráneo de la región, con marcadas diferencias entre estaciones húmedas y secas, esta descomposición fue crucial para ajustar los modelos y entender los patrones subyacentes.

Si bien los datos estaban limpios y no requirieron una limpieza adicional, es importante mencionar que el uso de datos completos y precisos es crucial en el análisis de series temporales. La elección del conjunto de datos precip_cr2met_mm_day y tmax_cr2met_C_mon fue fundamental para asegurar la integridad y confiabilidad del análisis posterior.

Metodología

Análisis Exploratorio de Datos (EDA)

Inspección Inicial

Se realizó una revisión preliminar para identificar la calidad de los datos, verificar la presencia de valores atípicos utilizando `isnull().sum()` y comprender la estructura general del dataset utilizando `.info()`

Análisis Descriptivo

Se llevaron a cabo estadísticas descriptivas para obtener una comprensión general de los niveles de precipitación, incluyendo medidas como la media, mediana, desviación estándar y rangos intercuartílicos.

Visualización de Tendencias y Patrones

Se emplearon gráficos de series temporales para visualizar las tendencias y patrones estacionales. Esto fue particularmente importante para identificar ciclos de sequía y períodos de alta precipitación.

Test no paramétrico de Mann-Kendall

Propósito y Aplicación

El test de Mann-Kendall es un método no paramétrico utilizado para identificar tendencias monótonas ya sea creciente y decreciente en series temporales. Es especialmente útil en el análisis de datos climáticos, donde se busca determinar si existe una tendencia significativa en variables como la temperatura o la precipitación a lo largo del tiempo. Este test no asume ninguna distribución específica para los datos y es menos sensible a las desviaciones de la normalidad y la presencia de valores atípicos.

Cómo funciona

Existen dos hipótesis

Hipótesis nula (H0): No hay tendencia (los datos son aleatorios, es decir, independientes y no idénticamente distribuidos).

Hipótesis alternativa (H1): Existe una tendencia (los datos no son aleatorios).

Procedimiento del Test

Cálculo de la Estadística S

Se calcula la estadística S, que es la suma de los signos de las diferencias entre pares de datos consecutivos en la serie temporal.

Estimación de la Varianza

Se calcula la varianza de S para determinar la dispersión de los datos.

Normalización

La estadística S se normaliza para obtener una distribución aproximadamente normal, lo que facilita la interpretación de los resultados.

Determinación de Tendencias

Si la estadística normalizada es significativamente diferente de cero, se concluye que existe una tendencia en los datos (ya sea creciente o decreciente). Una estadística positiva indica una tendencia ascendente, mientras que una negativa indica una tendencia descendente.

Consideraciones en su Aplicación

Se requiere una serie temporal de longitud suficiente para detectar tendencias. El test es robusto frente a valores atípicos, pero estos pueden influir en la detección de tendencias.

Es por esto que la aplicación de este test no paramétrico fue esencial en el proceso, ya que de sus resultados se puede ir corroborando, la diferencia de tendencias entre estaciones y descubrir información nueva para contrastar con los demás test y modelos.

Índice de Intensidad de Precipitación Diaria Simple (SDII)

El Índice de Intensidad de Precipitación Diaria Simple (SDII, por sus siglas en inglés) es un índice climático utilizado para caracterizar la intensidad de las precipitaciones. Este índice se calcula como el total anual de precipitación dividido por el número de días con precipitación en ese año, y se expresa en milímetros por día de precipitación (mm/día).

$$SDII = \frac{\text{Total de precipitación anual en días de lluvia}}{\text{Número total de días de lluvia en el año}}$$

El SDII es útil para entender las tendencias en la intensidad de las precipitaciones, lo que puede ser crucial para estudios relacionados con el cambio climático y la gestión de recursos hídricos.

Aplicaciones

Cambio Climático

El SDII ayuda a entender cómo están cambiando las características de las precipitaciones con el tiempo, lo cual es crucial en el estudio del cambio climático.

Gestión del Agua

Para la planificación de recursos hídricos, es importante saber no solo cuántos días llueve, sino también cuánta agua se espera en esos días.

Agricultura

La agricultura depende enormemente de la precipitación. Un cambio en la intensidad de las lluvias puede tener un gran impacto en la planificación y gestión agrícola

El SDII proporciona una visión complementaria a los análisis de tendencia como el test de Mann-Kendall, ofreciendo una perspectiva diferente sobre cómo están cambiando las precipitaciones en la cuenca del río Maipo.

Índice de Precipitación Estandarizado(SPI)

Es un indicador ampliamente utilizado para monitorear las sequías. Fue desarrollado principalmente para determinar la sequía

Definición

El SPI es un índice basado en la probabilidad, que se calcula para diferentes escalas de tiempo (como 1 mes, 3 meses, 6 meses, etc.).

Cálculo

Se calcula ajustando los datos de precipitación a una distribución de probabilidad (gamma), que luego se transforma en una distribución normal. De esta manera, el índice cuantifica la desviación de la precipitación para un período específico con respecto a la precipitación histórica.

Interpretación

Tabla 2

Valor del índice de precipitación estandarizado	Categorías
≥ 2.00	Extremadamente húmedo
1.50 a 1.99	Muy húmedo
1.00 a 1.49	Moderadamente húmedo
0.99 a -0.99	Normal o aproximadamente normal
-1.00 a -1.49	Moderadamente seco
-1.50 a -1.99	Muy seco
≤ -2.00	Extremadamente seco

En esta investigación se usó un límite inferior para el SPI ya que existían datos que tenían extremos muy secos o extremos muy húmedos. Por lo que se estableció un límite inferior para el SPI (-3) para todos aquellos valores que resulten ser -inf.

Incluir el SPI demuestra un enfoque integral en el estudio de la sequía y la variabilidad climática, utilizando una herramienta reconocida y ampliamente utilizada en la climatología y la hidrología.

Búsqueda de Correlación de Temperatura y Dryspell

Definición de la Correlación de Pearson

La correlación de Pearson es un método estadístico utilizado para medir la fuerza y la dirección de la asociación lineal entre dos variables cuantitativas. El coeficiente de correlación de Pearson, denotado comúnmente como r , varía entre -1 y +1. Un valor de +1 indica una correlación positiva perfecta, donde un aumento en una variable se asocia con un aumento proporcional en la otra. Un valor de -1 indica una correlación negativa perfecta, donde un aumento en una variable se asocia con una disminución proporcional en la otra. Un valor de 0 indica que no hay correlación lineal entre las variables. Los valores de r cercanos a +1 o -1 sugieren una fuerte correlación, mientras que los valores cercanos a 0 sugieren una correlación débil.

Modelo de predicción DeepAR

Se utilizó como modelo de predicción para evaluar la posibilidad de predecir los “dryspell” el modelo DeepAR el cuál es un modelo de pronóstico probabilístico basado en redes neuronales recurrentes (RNNs), desarrollado por Amazon para su biblioteca GluonTS. Está diseñado para manejar series de tiempo con patrones complejos y se utiliza ampliamente en tareas de predicción en varios campos.

Se detallan características importantes del modelo

Características Clave:

- **Modelado Probabilístico:** A diferencia de los modelos tradicionales que predicen un único valor futuro, DeepAR genera una distribución de probabilidad, ofreciendo una medida de incertidumbre en las predicciones.
- **Redes Neuronales Recurrentes (RNNs):** Utiliza RNNs, en particular Long Short-Term Memory (LSTM) o Gated Recurrent Units (GRU), que son eficaces para capturar dependencias temporales en series de tiempo.
- **Capacidad de manejar múltiples series de tiempo:** Puede entrenarse con un conjunto de series de tiempo relacionadas, lo que permite al modelo aprender patrones complejos y estacionales a partir de datos agregados.
- **Entradas Exógenas:** Permite la inclusión de variables exógenas (externas) para mejorar la precisión de las predicciones.

Desafíos Enfrentados

- **Complejidad Climática:** La predicción de periodos secos implica considerar una amplia gama de factores climáticos interconectados, lo que hace que la modelización sea extremadamente compleja.
- **Datos Históricos Limitados:** Los modelos como DeepAR requieren una gran cantidad de datos históricos para entrenamiento. La disponibilidad limitada o la calidad de los datos pueden afectar la precisión del modelo.
- **Influencia de Variables Externas:** Factores como el cambio climático, fenómenos meteorológicos extremos y variabilidad estacional pueden influir significativamente en la precisión del modelo.
- **Determinación de Hiperparámetros:** La elección de hiperparámetros óptimos para DeepAR puede ser un proceso de prueba y error que consume tiempo.

Resultados

Análisis exploratorio de datos (EDA)

A continuación se detallan los resultados exploratorios para el archivo:

precip_cr2met mm day.csv este conjunto de datos cuenta con 23 columnas con código estación, y una única columna con fecha correspondiente a cada registro diario.

Se cuenta con 8035 registros hasta la fecha final de diciembre de 2019, cero valores nulos.

(Véase código 1.1)

Código 1.1

Descripción general del dataset.

```
In [8]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8035 entries, 0 to 8034
Data columns (total 24 columns):
#   Column      Non-Null Count  Dtype
---  -
0   date        8035 non-null   datetime64[ns]
1   5701001     8035 non-null   float64
2   5701002     8035 non-null   float64
3   5702001     8035 non-null   float64
4   5703006     8035 non-null   float64
5   5704002     8035 non-null   float64
6   5705001     8035 non-null   float64
7   5706001     8035 non-null   float64
8   5707002     8035 non-null   float64
9   5710001     8035 non-null   float64
10  5715001     8035 non-null   float64
11  5716001     8035 non-null   float64
12  5720001     8035 non-null   float64
13  5721001     8035 non-null   float64
14  5721016     8035 non-null   float64
15  5722001     8035 non-null   float64
16  5722002     8035 non-null   float64
17  5730005     8035 non-null   float64
18  5730008     8035 non-null   float64
19  5734001     8035 non-null   float64
20  5737002     8035 non-null   float64
21  5741001     8035 non-null   float64
22  5746001     8035 non-null   float64
23  5748001     8035 non-null   float64
dtypes: datetime64[ns](1), float64(23)
memory usage: 1.5 MB
```

Además se suma un análisis descriptivo utilizando `.describe` de la librería pandas el cuál es útil para entender las características generales de las precipitaciones en las diferentes estaciones en donde lo más importante es:

- **Min:** Muestra el valor mínimo registrado en cada columna. En todas las estaciones, el mínimo es 0, lo que sugiere que hay días sin precipitaciones registradas.
- **25% (Primer Cuartil):** El 25% de los datos son menores o iguales a este valor. La mayoría de las estaciones tienen un valor de 0 en este cuartil, lo que indica que al menos el 25% de los días registraron no precipitaciones o cantidades muy bajas.
- **50% (Mediana):** La mediana divide el conjunto de datos en dos partes iguales. Nuevamente, la mayoría de las estaciones muestran una mediana de 0, sugiriendo que más de la mitad de los días tuvieron precipitaciones nulas o mínimas.
- **75% (Tercer Cuartil):** El 75% de los datos son menores o iguales a este valor. Vemos que para algunas estaciones este valor es ligeramente superior a 0, lo que indica que un 25% de los días registraron precipitaciones más significativas.
- **Max:** El valor máximo registrado en cada columna. Estos valores representan los días con las mayores precipitaciones en cada estación. Por ejemplo, la estación 5701001 registró un máximo de 148.797670 mm de lluvia.
- **Std (Desviación Estándar):** Muestra la variabilidad o dispersión de los datos. Una desviación estándar alta indica una mayor variabilidad en las precipitaciones. Las desviaciones estándar varían entre las estaciones, lo que refleja diferencias en la variabilidad climática en cada ubicación.

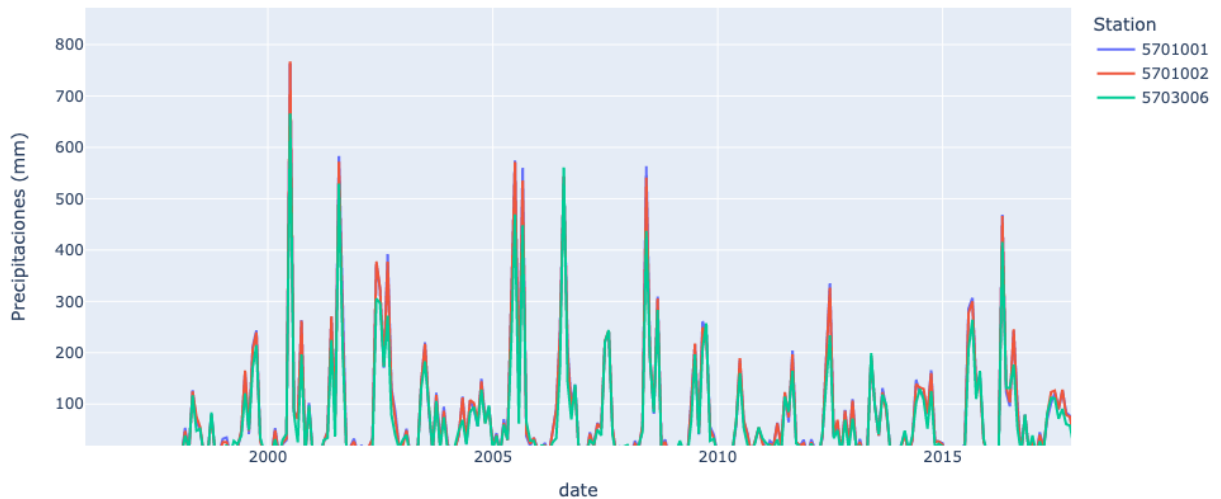


Fig. 2 Precipitaciones acumuladas mensuales por estación.

Estacionalidad: Se observa claramente un patrón estacional en las precipitaciones. Parece haber períodos del año con precipitaciones más altas y otros con precipitaciones más bajas. Esta estacionalidad es típica en muchos climas donde hay estaciones húmedas y secas definidas.

Variabilidad entre Estaciones: Las diferentes estaciones (líneas de colores) muestran tendencias similares en términos de estacionalidad, pero hay diferencias en los niveles de precipitación. Esto podría deberse a la ubicación geográfica de las estaciones, a los microclimas locales, o a otros factores.

Se eligieron 2 estaciones al azar para visualizar mediante un heatmap (mapa de calor) las precipitaciones acumuladas mensualmente a lo largo de todo el registro del dataset.

Precipitaciones acumuladas mensuales

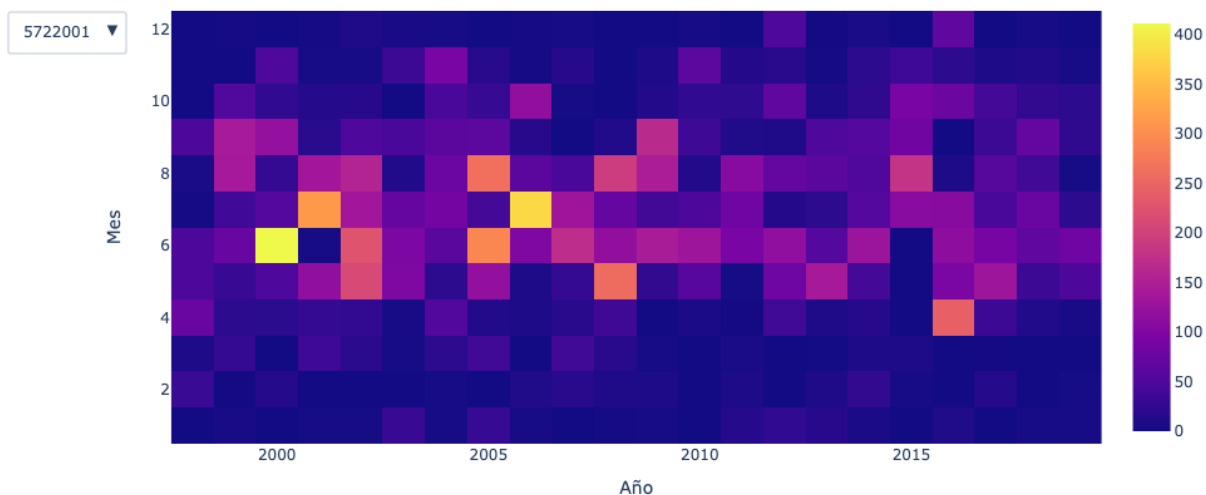


Fig. 3 Precipitaciones acumuladas mensuales estación 5722001

La estación 5722001 Estero Arrayán en la Montosa cuenta con dos meses particularmente altos en precipitaciones con 411 mm para el mes Junio del año 2000 y 379 mm para el mes de Mayo del año 2006. Los demás meses tienen una cantidad considerablemente menor de precipitaciones.

Precipitaciones acumuladas mensuales

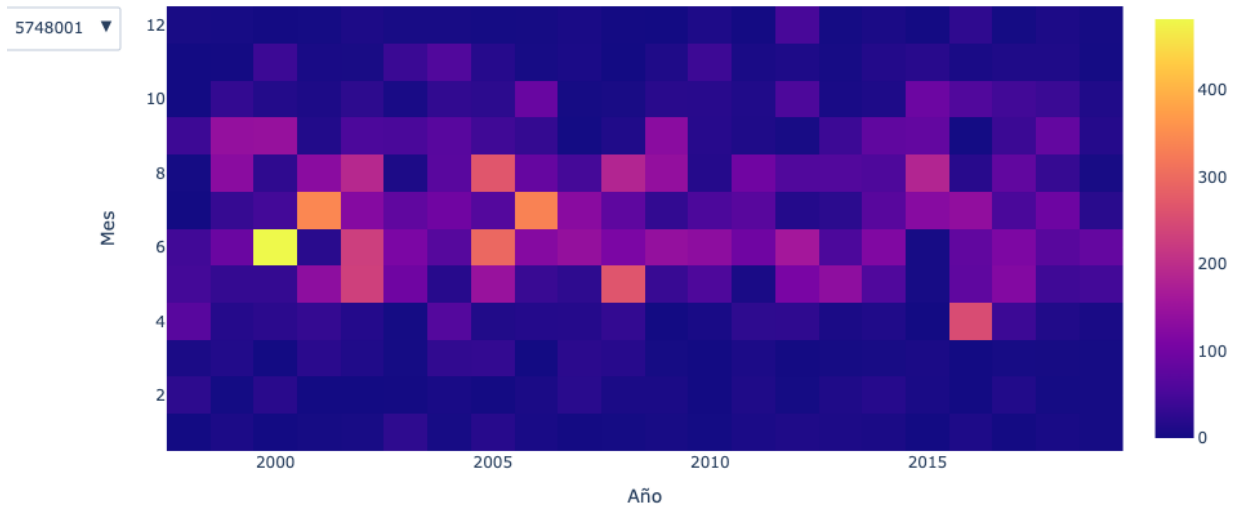


Fig. 4 Precipitaciones acumuladas mensuales estación 5748001

Se identifican patrones levemente iguales a la estación anterior, esta vez corresponde a Cuenca 5748001 - Río Maipo En Cabimbao, esta cuenca geográficamente cubre una extensa área que incluye otras cuencas más pequeñas.

En el contexto de análisis de series temporales, un patrón "estacional" se refiere a fluctuaciones que ocurren en intervalos regulares y predecibles, como cambios en la temperatura y la precipitación que siguen el ciclo de las estaciones. Estos patrones pueden tener un impacto significativo en el análisis de los datos y en cualquier modelo predictivo que se desarrolle.

La descomposición estacional de las series temporales permitiría visualizar y entender mejor estos componentes estacionales, además de las tendencias subyacentes y otros patrones en los datos.

Para esto se utilizó la librería `statsmodels.tsa.seasonal` que nos permite extraer esa información y visualizar.

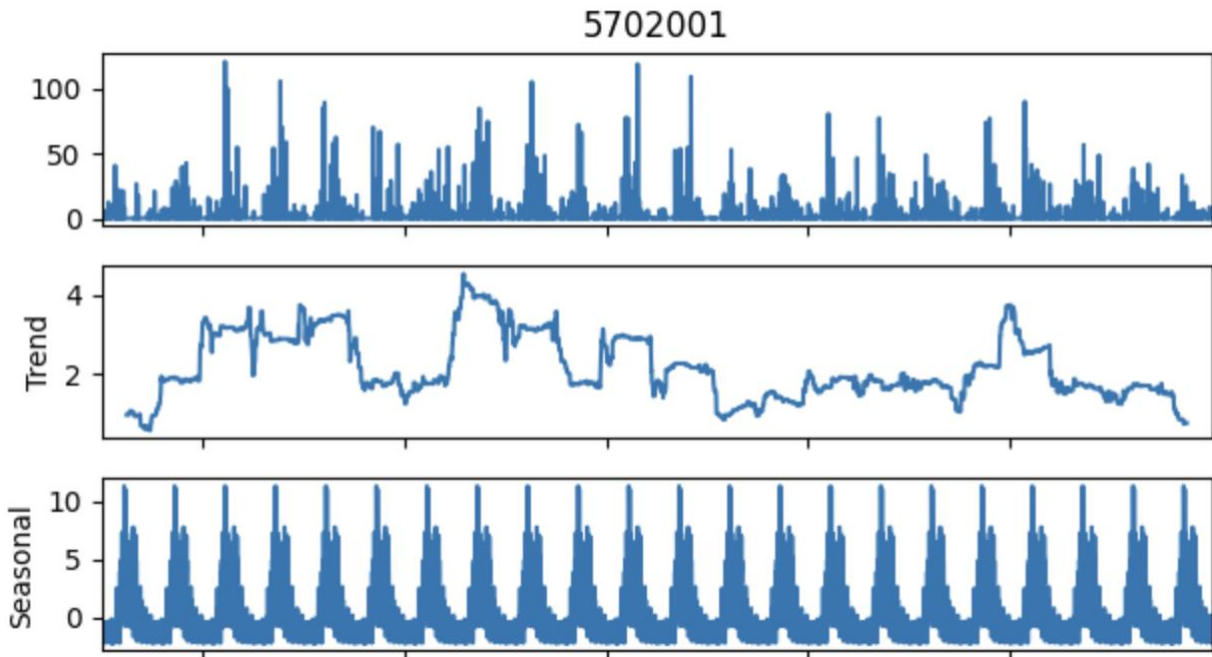


Fig. 5 Patrones estacionales estación 5702001

Tendencia (Trend):

- Representa la dirección o curso a largo plazo de la serie temporal.
- Hay un aumento notable en la tendencia en algún punto en la mitad de la serie temporal, lo que podría indicar un período con mayor promedio de precipitaciones.
- Después de alcanzar un pico, la tendencia parece disminuir, lo que sugiere una disminución en las precipitaciones hacia el final de la serie temporal.

Estacionalidad (Seasonal):

- Refleja las fluctuaciones o patrones que se repiten a intervalos regulares debido a factores estacionales.
- En el contexto climático, por ejemplo, podría reflejar patrones que se repiten anualmente con las estaciones.

Residuo (Residual):

- Comprende la variación en los datos que no es explicada por la tendencia o la estacionalidad.
- Son esencialmente los errores o los componentes aleatorios que no siguen un patrón predecible.
- En estudios climatológicos y meteorológicos, es importante analizar los residuos para entender eventos extremos y su impacto.
- Los residuos también pueden ser analizados para detectar cambios en la variabilidad de las precipitaciones que podrían estar relacionados con cambios climáticos o factores antropogénicos.

Interpretación de los Resultados del Test de Mann-Kendall

Los resultados del test de Mann-Kendall aplicados a las series temporales de precipitación de diversas estaciones en la cuenca del río Maipo indican lo siguiente:

Se deja en claro y en evidencia los posibles sesgos y limitaciones si se hubiera considerado hasta el final del dataset Abril de 2020.

Representatividad Estacional

El estudio se enfoca en tendencias a lo largo de todo el año, es ideal tener datos que cubran todos los meses para cada año. Al detenerse en abril de 2020, se está omitiendo parte del año, lo que podría afectar la representación de las tendencias estacionales, especialmente si esos meses ausentes suelen tener características climáticas distintivas.

Consistencia en el Análisis de Tendencias

Para un análisis de tendencias, como el test de Mann-Kendall, es preferible tener una serie temporal completa y consistente. La falta de datos para una parte del año podría introducir un sesgo, ya que los patrones de precipitación o temperatura pueden variar significativamente a lo largo de los meses.

Resultados

Tendencia General

Los resultados del test de Mann-Kendall indican que la mayoría de las estaciones analizadas (23 de 23) no presentan una tendencia significativa en las precipitaciones durante el período de estudio (1998-2019). Esto se refleja en los valores de 'Z' que no superan el umbral de

significancia y los valores de 'p' mayores a 0.05, sugiriendo la ausencia de una tendencia marcada en las precipitaciones en estas estaciones.(Vease tabla 3)

Interpretación de los Valores de 'Tau'

Los valores de Tau de Kendall en todos los casos son bajos, reflejando una ausencia de tendencias fuertes y marcadas en las precipitaciones. La consistencia en la dirección y magnitud de estos valores sugiere que, aunque no son estadísticamente significativos, hay una leve inclinación hacia tendencias negativas, pero su impacto es limitado.(Vease tabla 3)

Tabla 3

Estaciones al azar donde refleja según el Test dichos resultados

Estación: 5701001
Trend: no trend
Z Value: -0.1156462912047216
p Value: 0.9079328739332353
h Value: False
s Value: -24848.0
Tau Value: -0.000770037436748419

Estación: 5701002
Trend: no trend
Z Value: -0.3278465770336318
p Value: 0.7430276684476431
h Value: False
s Value: -70738.0
Tau Value: -0.002192164689339571

Estación: 5706001
Trend: no trend
Z Value: -0.644740927208035
p Value: 0.519095089100585
h Value: False
s Value: -124367.0
Tau Value: -0.0038541229030944393

Estación: 5707002
Trend: no trend
Z Value: -0.4021281620608116
p Value: 0.6875897078601314
h Value: False
s Value: -85189.0
Tau Value: -0.002639999967770487

Significado Ecológico y Ambiental

La falta de tendencias significativas en la mayoría de las estaciones sugiere una relativa estabilidad en los patrones de precipitaciones durante el período analizado. Sin embargo, la variabilidad observada subraya la necesidad de considerar la heterogeneidad local en cualquier planificación de manejo de recursos hídricos y estrategias de adaptación al cambio climático.

Implicaciones para la Investigación Futura

Estos hallazgos enfatizan la importancia de continuar con el monitoreo y análisis detallado de las tendencias climáticas en la cuenca del río Maipo. La variabilidad entre las estaciones y la sensibilidad de los resultados a los cambios en el período de análisis resaltan la necesidad de un enfoque detallado y localizado en la gestión de los recursos hídricos y en las estrategias de adaptación al cambio climático.

Interpretación de los resultados del SDII

Análisis del Índice de Intensidad Diaria Simple (SDII) de las Precipitaciones (1998-2019)

Visión General de los Resultados

Los resultados actualizados del SDII indican que, en general, hay una tendencia decreciente en la intensidad de las precipitaciones diarias en la mayoría de las estaciones estudiadas. Sin embargo, es importante notar que los valores de pendiente (slope) son generalmente bajos y los valores de 'p' en muchos casos son superiores a 0.05, lo que implica que estas tendencias no son estadísticamente significativas en todas las estaciones.

Tendencias Observadas

Tendencias Levemente Decrecientes: La mayoría de las estaciones muestran pendientes negativas, sugiriendo una tendencia hacia una menor intensidad en las precipitaciones diarias. Aunque estas tendencias son leves, reflejan un cambio sutil en la intensidad de las precipitaciones a lo largo del tiempo.

Significancia Estadística

Unas pocas estaciones, como la 5702001 y la 5707002, muestran valores de 'p' inferiores a 0.05, indicando que las tendencias decrecientes en estas estaciones son estadísticamente significativas. Sin embargo, en la mayoría de las estaciones, los valores de 'p' no alcanzan este umbral, sugiriendo que las tendencias observadas podrían no ser consistentes o influenciadas por variabilidad a corto plazo. (Vease tabla 4)

Tabla 4

Tabla de resultados Intensidad de Precipitación Diaria Simple (SDII)

Estación	Pendiente (Slope)	P-value
5701001	-0.124776	0.187409
5701002	-0.141430	0.121489
5702001	-0.129742	0.049907
5703006	-0.136391	0.087006
5704002	-0.134745	0.104262
5705001	-0.096912	0.099858
5706001	-0.129730	0.084119
5707002	-0.100881	0.036675
5710001	-0.113406	0.104228
5715001	-0.103305	0.161756
5716001	-0.144224	0.077383
5720001	-0.087528	0.170417
5721001	-0.121465	0.076651
5721016	-0.103132	0.156163
5722001	-0.108351	0.173551
5722002	-0.094774	0.097661
5730005	-0.068124	0.220830
5730008	-0.057949	0.297905

5734001	-0.042158	0.520146
5737002	-0.101532	0.084144
5741001	-0.043158	0.625669
5746001	-0.048873	0.478504
5748001	-0.085987	0.154457

Implicaciones Ecológicas y Ambientales

Estos resultados apuntan a una tendencia general hacia precipitaciones diarias de menor intensidad en la región. Aunque la mayoría de estas tendencias no son estadísticamente significativas, sugieren una posible modificación en los patrones de lluvia que podría tener implicancias para la gestión de los recursos hídricos, la agricultura y la planificación ambiental en la cuenca del río Maipo.

Recomendaciones para Futuras Investigaciones

La variabilidad en los resultados del SDII entre diferentes estaciones resalta la importancia de considerar factores locales y regionales en el estudio de las tendencias de precipitaciones. Futuras investigaciones deberían enfocarse en entender mejor las causas de estas tendencias y en evaluar su impacto a largo plazo en los ecosistemas y en la gestión de recursos hídricos.

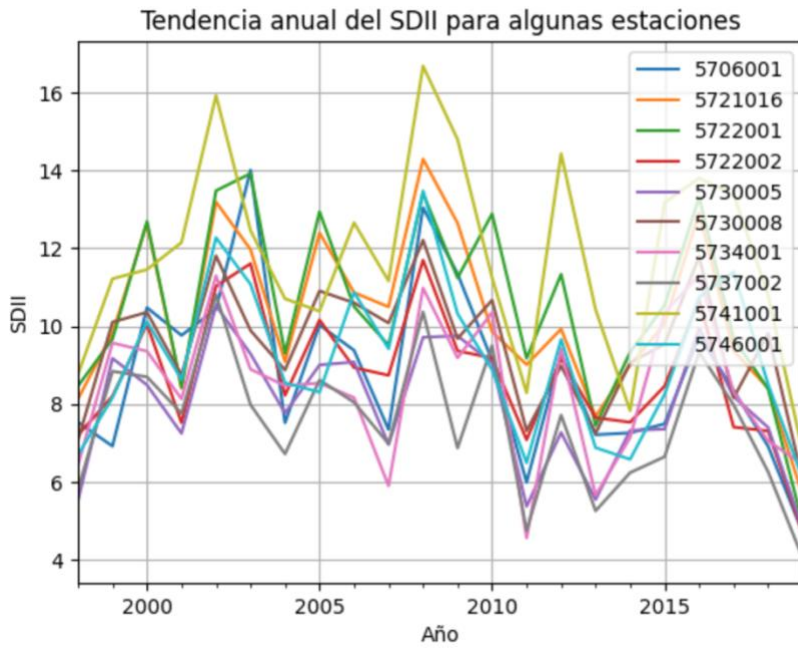


Fig. 6 Tendencia anual del SDII

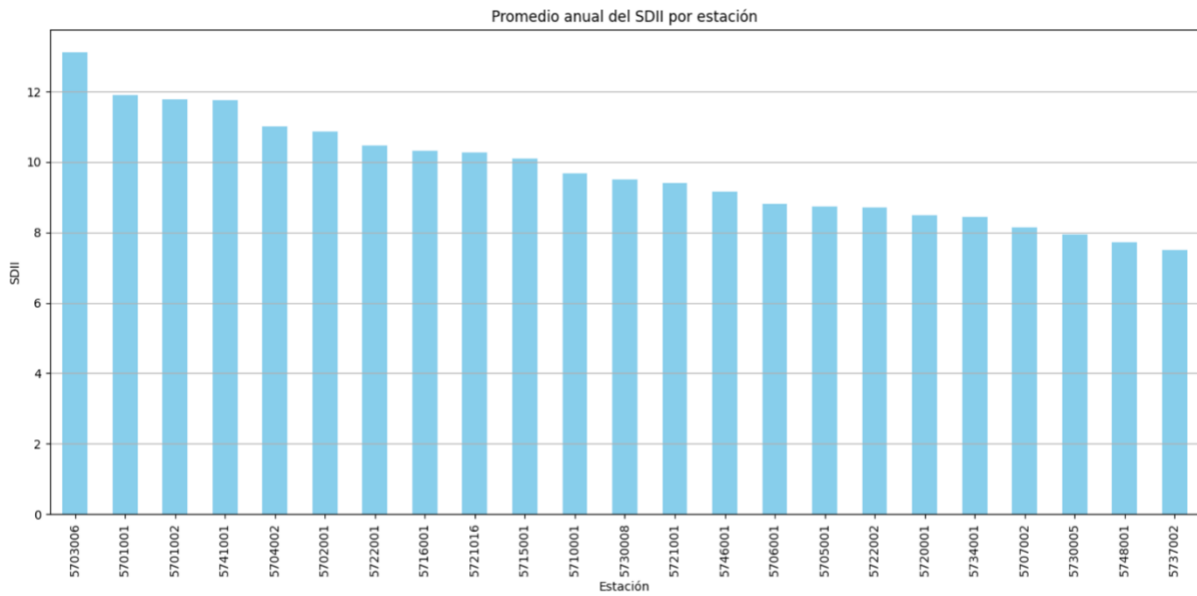


Fig. 7 Promedio anual del SDII por estación

El gráfico muestra el promedio anual del SDII para cada estación o localidad. Las estaciones están ordenadas de mayor a menor según su intensidad promedio de precipitación. Esto te permite ver rápidamente cuáles estaciones tienen una mayor intensidad promedio de precipitación (SDII más alto) y cuáles tienen una menor intensidad (SDII más bajo). Si se observa la barra más alta en el gráfico, que corresponde a la estación 5703000, el SDII está cerca de 13, esto significa que, en promedio, cada día lluvioso en esta estación durante el año tuvo alrededor de 13 milímetros de precipitación.

Las estaciones que tienen un SDII más bajo son las siguientes con identificación geográfica:

5720001 Río Molina Antes Junta San Francisco

5707002 Río Colorado Antes Junta Río Maipo

5730005 Zanjón De La Aguada

5706001 Río Olivares

5722002 Río Mapocho en los Almendros

5730008 Quebrada Ramon en Recinto Emos

5734001 Estero Polpaico en Chicaumo

5746001 Estero Puangue en Ruta 78

Desde esta visualización, se pueden identificar variaciones regionales en la intensidad de la precipitación. Por ejemplo, la estación a la izquierda del gráfico tiene el SDII promedio anual más alto, lo que indica que, en promedio, experimenta una mayor intensidad de precipitación en días húmedos en comparación con otras estaciones.¶

Interpretación de los Resultados del Índice Estandarizado de Precipitación SPI para la Estación Húmeda.

El SPI utilizado es el de 30 días.

Código 2.1

Definición de estaciones

```
# Definir las estaciones en el hemisferio sur
estacion_humeda_meses = [4, 5, 6, 7, 8, 9, 10] # Abril a Octubre
estacion_seca_meses = [11, 12, 1, 2, 3] # Noviembre a Marzo
```

Definición de estaciones

Humeda: A partir del mes de Abril hasta el mes de Octubre

Seca: A partir del mes de Noviembre hasta el mes de Marzo

Tabla 5

Resultados de estadísticas para la estación húmeda

Estadísticas ajustadas para la estación húmeda:

	SPI
count	154.000000
mean	0.451169
std	0.949893
min	-3.000000
25%	-0.135518
50%	0.479441
75%	0.947316
max	2.992386

Rango de Valores del SPI:

El rango de SPI va desde -3 hasta 2,992. Esto indica que durante el período de estudio, hubo años con condiciones extremas tanto de sequía como de exceso de precipitación.

Media del SPI:

La media del SPI es aproximadamente 0.451. Este valor positivo sugiere que, en promedio, la estación húmeda ha sido ligeramente más húmeda que el promedio histórico.

Desviación Estándar:

Una desviación estándar de 0.949 indica una variabilidad considerable en la precipitación durante la estación húmeda a lo largo de los años. Esto podría reflejar la influencia de factores como el cambio climático o la variabilidad interanual natural.

Análisis Cuantil:

El 25% de los valores de SPI están por debajo de -0.135, y el 75% están por debajo de 0.947.

Esto significa que la mayoría de los valores de SPI caen en un rango que va de ligeramente secos a moderadamente húmedos.

Extremos:

Los valores extremos de -3 y 3+ indican años de condiciones muy secas y muy húmedas, respectivamente. Estos extremos son cruciales para entender la variabilidad y la intensidad de los eventos de sequía y exceso de precipitación.

Pendiente: -0.00030586797811902775
Valor p para la pendiente: 0.8596838249236703

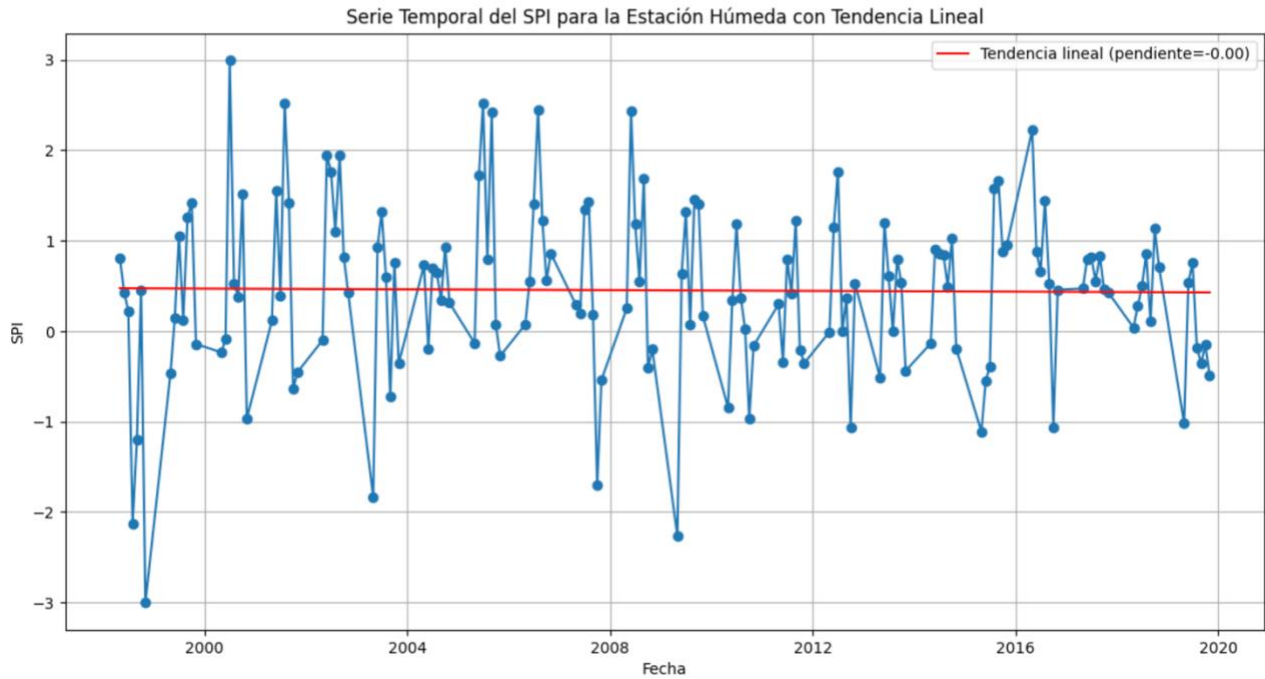


Fig. 8 Serie temporal del SPI Estación húmeda

El SPI muestra una tendencia general positiva normal durante la estación húmeda, pero los otros modelos indican una disminución en la intensidad de las precipitaciones, podría sugerir que, aunque la estación húmeda sigue siendo húmeda, la cantidad o intensidad de las precipitaciones puede estar disminuyendo. Esto sería relevante para entender cómo los patrones de precipitación están cambiando en la región.

Longitud de Dryspell

En un principio se utilizó la metrica estadistica del promedio, sin embargo este estadístico puede no ser el mejor indicador, en situaciones donde hay una alta variabilidad, como es el caso de los períodos secos en un clima mediterráneo. El uso de valores mínimos y máximos puede proporcionar una visión más clara de la gama completa de variabilidad en la duración de los dry spells. es por eso que se presenta un heat map con la duración máxima de los dryspell para cada estación y año.

Para este Heatmap, los colores amarillos representan una cantidas mas alta de periodos de dryspell mientras que los mas azulados representan una duración menor.

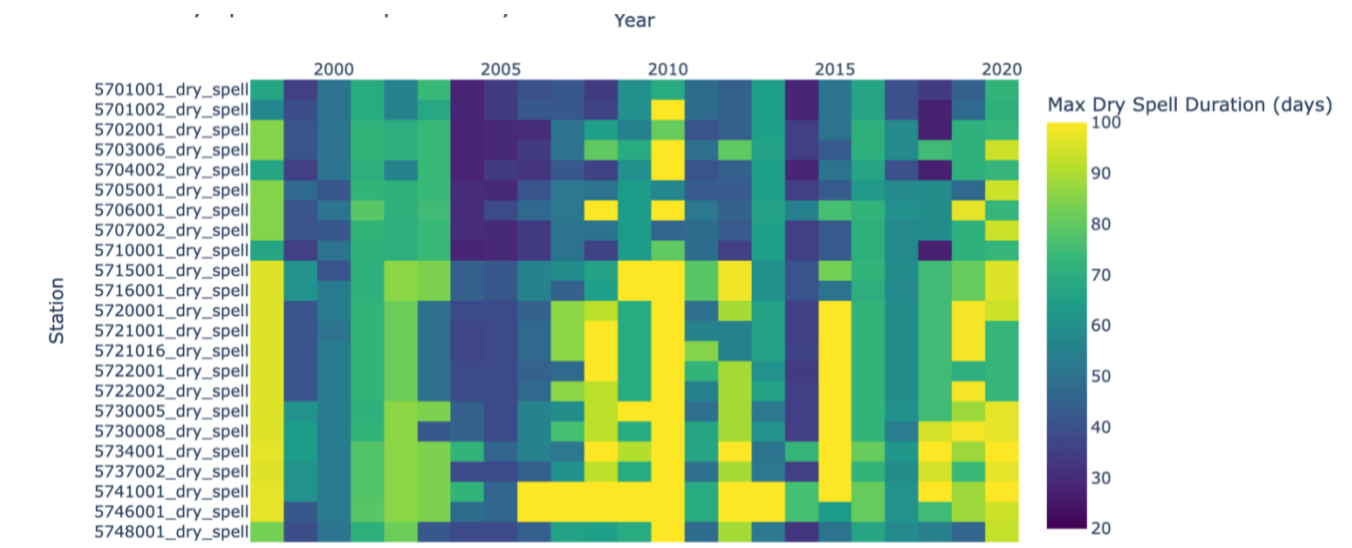


Fig. 9 Duración máxima de los dry-spell por año y estación.

Variabilidad en la Longitud de los Dry Spells: Los resultados muestran una variabilidad considerable en la longitud promedio de los dry spells entre las estaciones, lo que indica diferencias en los patrones climáticos a lo largo de la cuenca.(Véase figura 9)

Frecuencia de los Dry Spells: El número de períodos secos también varía entre las estaciones, lo que sugiere que algunas áreas pueden ser más propensas a experimentar dry spells que otras.(Véase figura 9).

Las estaciones que muestran las mayores duraciones medias de periodos secos (dry spells) son las siguientes:

- Estación **5741001**: Duración media de 90.48 días, mediana de 88 días.
- Estación **5746001**: Duración media de 84.39 días, mediana de 84 días.
- Estación **5734001**: Duración media de 81.30 días, mediana de 81 días.
- Estación **5730005**: Duración media de 73.70 días, mediana de 72 días.
- Estación **5715001**: Duración media de 73.35 días, mediana de 72 días.
- Estación **5730008**: Duración media de 72.35 días, mediana de 72 días.
- Estación **5716001**: Duración media de 71.65 días, mediana de 70 días.
- Estación **5737002**: Duración media de 70.74 días, mediana de 72 días.
- Estación **5721016**: Duración media de 69.91 días, mediana de 72 días.

Correlación de variable temperatura y “DrySpell”

En el análisis, se investigó la relación entre la temperatura diaria y la ocurrencia de días secos (dryspell) en diferentes estaciones meteorológicas. Para ello, aplicamos la correlación de Pearson, que mide la asociación lineal entre dos variables. Los resultados mostraron correlaciones positivas significativas en varias estaciones, indicando que a medida que la temperatura aumenta, también lo hace la probabilidad de experimentar días secos. Los coeficientes de correlación variaron entre las estaciones, reflejando diferencias en los patrones climáticos locales. La significancia estadística de estas correlaciones fue alta, con p-valores cercanos o iguales a cero, lo que sugiere que estas tendencias no son producto del azar, más bien, reflejan patrones reales.

Tabla 6

Resultados correlación temperatura y dry-spell

Out [24] :

	correlation	p_value
5720001	0.417186	0.000000e+00
5730008	0.458919	0.000000e+00
5730005	0.444669	0.000000e+00
5722002	0.407211	1.976263e-323
5701002	0.408461	0.000000e+00
5721001	0.360083	3.172359e-248
5746001	0.399727	1.232223e-310
5748001	0.445356	0.000000e+00
5721016	0.398027	8.872346e-308
5716001	0.473331	0.000000e+00
5706001	0.302210	7.678309e-172
5705001	0.297973	6.926896e-167
5707002	0.335421	1.225717e-213
5704002	0.399831	8.222884e-311
5702001	0.381981	1.274426e-281
5701001	0.392115	5.703196e-298
5715001	0.467856	0.000000e+00
5710001	0.391835	1.641799e-297
5734001	0.431135	0.000000e+00
5741001	0.413815	0.000000e+00
5722001	0.439668	0.000000e+00
5703006	0.396594	2.207641e-305
5737002	0.441698	0.000000e+00

Interpretación en el Contexto de Capstone project

Los resultados obtenidos en este estudio, muestran una correlación significativa entre el aumento de las temperaturas y la frecuencia de días secos en varias estaciones meteorológicas. Esto es indicativo de que los cambios en las temperaturas podrían ser un factor contribuyente a los patrones observados de sequía en la región. Estos hallazgos son consistentes con las expectativas teóricas y aportan evidencia empírica que apoya la teoría de que el cambio climático y el aumento de las temperaturas, pueden estar influenciando la frecuencia y la duración de los períodos de sequía.

Análisis de Resultados del Modelo DeepAR para la Predicción de Precipitaciones

Contexto y Objetivo del Modelo

En este proyecto, se aplicó el modelo DeepAR, un enfoque de redes neuronales recurrentes, para predecir las cantidades de precipitación en la cuenca del río Maipo. El objetivo era predecir estas cantidades para identificar posteriormente períodos de sequía o "dry spells"

Formato de entrada al modelo DeepAr

En el modelo DeepAR de GluonTS, `ListDataset` es una estructura de datos tipo diccionarios utilizada para almacenar y manipular series temporales para el entrenamiento y la predicción en modelos de aprendizaje profundo. Cada entrada del diccionario generalmente contiene claves como "start", "target" principalmente y adicionalmente cuenta con "feat_dynamic_real" que sirve para aplicación de covariables.

`ListDataset` proporciona una interfaz para iterar fácilmente sobre los datos durante el entrenamiento, lo que permite al modelo acceder a las series temporales y sus características asociadas. (Véase Código 3.1)

Código 3.1

Resultado del acceso a “ListDataset” en python

```
from gluonts.dataset.common import ListDataset
import matplotlib.pyplot as plt

# Asegurándonos de que el índice es de tipo DateTime
data.index = pd.to_datetime(data.index)
data['date'] = pd.to_datetime(data['date']) # Convertimos la columna de fecha en un objeto datetime
data.set_index('date', inplace=True)

# Visualización de las primeras filas del dataframe
print(data.head())

# Convertir el DataFrame de pandas en un ListDataset de GluonTS
time_series = []
for column in data.columns:
    time_series.append({"start": data.index[0], "target": data[column].tolist()})

# Creación del conjunto de datos (este paso es preparatorio y no se visualizará nada)
training_data = ListDataset(time_series, freq="D")
```

	5701001	5701002	5702001	5703006	5704002	5705001	5706001	\
date								
1998-01-01	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1998-01-02	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1998-01-03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1998-01-04	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1998-01-05	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

División de las fechas para entrenamiento, validación y prueba

Los tramos de corte para los datos fueron definidos por fecha:

Conjunto de Entrenamiento menor a 31 de Diciembre de 2016

Conjunto de Validación entre 1 de enero de 2017 hasta 31 de Diciembre de 2017

Conjunto de Prueba desde 1 de enero de 2018 hasta 31 de Diciembre de 2019

Configuración y Optimización del Modelo

El modelo se configuró utilizando el estimador DeepAREstimator de GluonTS, con parámetros clave optimizados mediante Optuna. Los hiperparámetros óptimos encontrados fueron: número de capas (num_layers) igual a 2, tasa de dropout (dropout_rate) de 0.3044 y tasa

de aprendizaje (lr) de aproximadamente $2.538e-05$. Estos parámetros fueron seleccionados para equilibrar la complejidad del modelo y evitar el sobreajuste.

Resultados y Métricas de Rendimiento

Para tener un contexto del rendimiento del modelo proporcionaré información de los parámetros de dichas métricas.

Las métricas de rendimiento indicaron desafíos significativos en la precisión de las predicciones del modelo. El error cuadrático medio (MSE) fue de 1.529 y el error absoluto medio (MAE) de 38.491, sugiriendo una discrepancia notable entre las predicciones y los valores reales. Además, los valores elevados en MAPE y sMAPE reafirmaron la dificultad del modelo para predecir con precisión las cantidades de precipitación.

Código 4.1

```
from gluonts.evaluation import Evaluator
import json

evaluator = Evaluator(quantiles=[0.1, 0.5, 0.9])
agg_metrics, item_metrics = evaluator(iter(tss), iter(forecasts), num_series=len(training_dataset))

print(json.dumps(agg_metrics, indent=4))
#print(agg_metrics)

Running evaluation: 100% | ██████████ | 23/23 [00:00<00:00, 352.54it/s]

{
  "MSE": 1.5291800413331342,
  "abs_error": 38.4917840371089,
  "abs_target_sum": 38.48895925283432,
  "abs_target_mean": 0.3346866021985593,
  "seasonal_error": 2.0712960183050177,
  "MASE": 0.13052940198354782,
  "MAPE": 0.998677220609453,
  "sMAPE": 1.99722437651261,
  "MSIS": 5.2197061795225785,
  "QuantileLoss[0.1]": 7.7043994313897235,
  "Coverage[0.1]": 0.034782608695652174,
  "QuantileLoss[0.5]": 38.49178383978628,
  "Coverage[0.5]": 0.5130434782608696,
  "QuantileLoss[0.9]": 69.26878516703437,
  "Coverage[0.9]": 0.7652173913043478,
  "RMSE": 1.2366001946195602,
  "NRMSE": 3.6948004087893644,
  "ND": 1.0000733920669567,
  "wQuantileLoss[0.1]": 0.20017167470752986,
  "wQuantileLoss[0.5]": 1.0000733869402236,
  "wQuantileLoss[0.9]": 1.7997053313914542,
  "mean_absolute_QuantileLoss": 38.48832281273679,
  "mean_wQuantileLoss": 0.9999834643464025,
  "MAE_Coverage": 0.35507246376811596,
  "OWA": NaN
}
```

Periodo de predicción o context length

En el contexto de predicción de la precipitación, se presentan varios problemas uno de ellos es la cantidad de días en los que se pretende predecir, eso implica multivariantes de datos que no se tienen en esta investigación, cómo afecta el microclima, la altura del área, la zona geográfica propiamente tal, para la gran amplitud de la zona, la gran variabilidad dificulta este proceso.

Se realizaron varias pruebas para ver el funcionamiento del modelo, es así como en la imagen tiene una ventana de contexto de 5 días el cual no se ve reflejado ningún resultado óptimo aún si hubiese precipitación, lo mismo sucedió con ventanas de contexto mayores indicando la poca precisión de las predicciones del modelo.

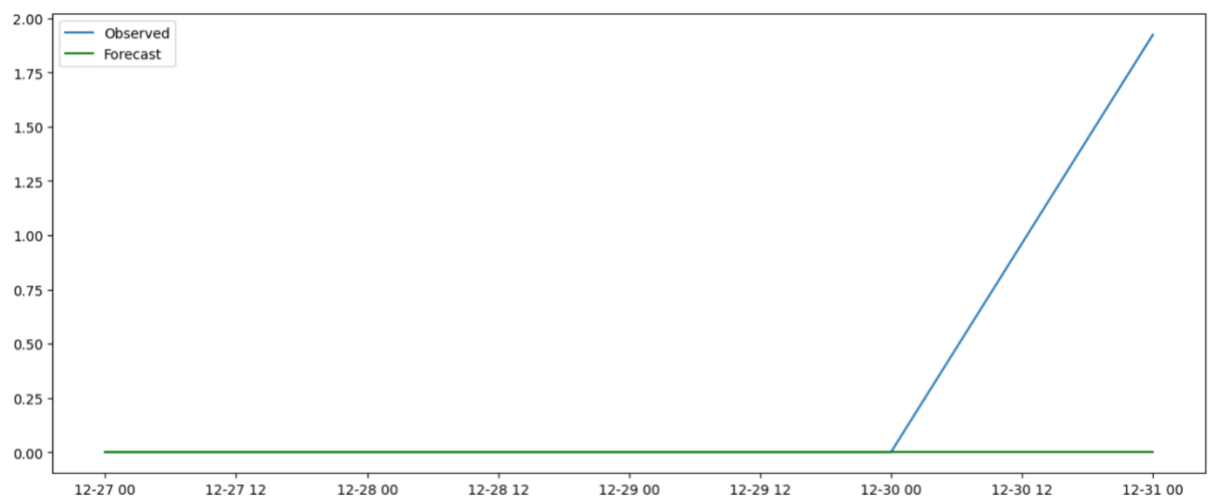


Fig. 10 Grafico predicción modelo DeepAr

Discusión y Reflexiones

Los resultados subrayan la complejidad inherente en la predicción del clima, un fenómeno que involucra numerosas variables e interacciones dinámicas. Esta complejidad, combinada con las limitaciones del modelo DeepAR para capturar la dinámica completa de las series temporales climáticas, puede haber contribuido a los desafíos observados. La longitud de la serie temporal y la calidad de los datos son factores adicionales que podrían haber influido en el rendimiento del modelo.

Conclusiones

El proyecto "AquaTendencias Maipo" ha realizado una contribución significativa al análisis de la variabilidad climática y su impacto en los patrones de sequía en la cuenca del río Maipo. Al centrarse en la hipótesis de que los patrones climáticos históricos y los datos meteorológicos pueden ser indicativos para la predicción de períodos secos. El estudio ha empleado metodologías avanzadas para desentrañar las complejas interacciones entre los factores climáticos y los eventos de sequía. Los resultados destacan una tendencia general hacia una menor intensidad de las precipitaciones diarias en la mayoría de las estaciones estudiadas. Un cambio en la intensidad y frecuencia de las precipitaciones, hace evidente que la cantidad de dryspell o periodos secos se vean con mayor frecuencia y esto puede afectar la disponibilidad de agua, la planificación de la infraestructura hídrica y las estrategias agrícolas.

Aunque las tendencias actuales pueden ser leves, es crucial considerarlas en una perspectiva a largo plazo, especialmente en el contexto del cambio climático, que podría exacerbar o alterar estas tendencias.

La variabilidad entre las estaciones y la ausencia de tendencias significativas en muchas de ellas subrayan la complejidad de predecir patrones de sequía en la región. Se ha demostrado que existe una correlación significativa entre ciertos patrones climáticos y la ocurrencia de períodos secos. Sin embargo, la investigación también ha revelado la alta variabilidad y complejidad inherente a los sistemas meteorológicos.

El proyecto a pesar de encontrarse con limitaciones técnicas y de una complejidad notable a los cambios meteorológicos se obtuvo una radiografía a lo que respecta al área geográfica de la cuenca del Río Maipo lo que nos permite desarrollar una idea clara para que desde ahora se tomen iniciativas con respecto al recurso hídrico y un enfoque mayor al estudio de los llamados dryspell, que bajo esta investigación tomaron una relevancia en como afecta al área de la agricultura, seguridad alimentaria, gestión de recursos y ecosistemas y biodiversidad.

Recomendaciones para Futuras Investigaciones

- **Abordar la Variabilidad Climática:** Dado que las variaciones climáticas y los eventos extremos pueden afectar significativamente la precisión del modelo, es crucial incorporar un análisis más profundo de estos factores. Esto podría incluir el uso de modelos climáticos más detallados o la colaboración con climatólogos para entender mejor estos efectos.
- **Incorporación de Modelos Climáticos en la Predicción:** Integrar modelos climáticos específicos en el proceso de predicción de DeepAR podría ayudar a comprender y predecir mejor las variaciones en los periodos secos. Esto implica una fusión de modelos de machine learning con modelos climáticos tradicionales.
- **Uso de Técnicas de Big Data y Aprendizaje Automático Avanzado:** Dada la gran cantidad de datos y la complejidad de los patrones climáticos, el uso de técnicas de big data y aprendizaje automático avanzado, como redes neuronales convolucionales (CNN) y modelos de atención, puede proporcionar nuevas perspectivas.
- **Análisis de Series Temporales Multivariadas:** Considerar múltiples variables climáticas y sus interacciones en el modelo puede ayudar a capturar la complejidad de los

patrones de los periodos secos. Esto incluye variables como temperatura, presión atmosférica, humedad y datos de fenómenos climáticos extremos.

- **Simulaciones de Escenarios Futuros:** Utilizar el modelo para simular diferentes escenarios futuros de cambio climático y su impacto en los periodos secos puede proporcionar información valiosa para la planificación y la adaptación.
- **Evaluación de la Resiliencia del Modelo a Cambios Climáticos:** Probar cómo el modelo responde a los cambios rápidos y extremos en las condiciones climáticas puede ayudar a evaluar y mejorar su robustez.
- **Incorporación de Técnicas de Aprendizaje por Transferencia:** Utilizar el aprendizaje por transferencia para aplicar conocimientos adquiridos en un conjunto de datos o región a otra podría ser una forma efectiva de adaptar el modelo a diferentes contextos climáticos.
- **Fomentar la Colaboración Interdisciplinaria:** Trabajar junto con expertos en diferentes áreas, como climatología, meteorología y ciencias ambientales, para desarrollar un enfoque más holístico en la modelización de periodos secos.

Estas recomendaciones tienen como objetivo, no solo mejorar la precisión y la aplicabilidad del modelo DeepAR, sino también, reconocer y abordar la complejidad inherente al modelar fenómenos climáticos. A través de estos esfuerzos, se espera contribuir al desarrollo de herramientas de predicción más robustas y confiables para la gestión de recursos hídricos y la planificación agrícola en un contexto de cambio climático.

Bibliografía

Motamedi, A., Gohari, A., & Haghghi, A. T. (2023). Three-decade assessment of dry and wet spells change across Iran, a fingerprint of climate change. *Scientific Reports*, 13(1), 2888. <https://doi.org/10.1038/s41598-023-30040-0>

Villazón Gómez, Juan Alejandro, Noris Noris, Pavel, & García Reyes, Roberto Alejandro. (2021). Uso del índice de precipitación estandarizado para la evaluación de la intensidad de la sequía en la provincia de Holguín. *Idesia (Arica)*, 39(4), 103-109. <https://dx.doi.org/10.4067/S0718-34292021000400103>

Alencar Da Silva Alves, K. M., & Silva Nóbrega, R. (2017). Tendencia pluviométrica y concentración estacional de precipitación en la cuenca hidrográfica del río Moxotó Pernambuco—BRASIL. *Revista Geográfica de América Central*, 1(58), 295. <https://doi.org/10.15359/rgac.58-1.12>

Rivoire, P., Trambly, Y., Neppel, L., Hertig, E., & Vicente-Serrano, S. M. (2019). Impact of the dry-day definition on Mediterranean extreme dry-spell analysis. *Natural Hazards and Earth System Sciences*, 19(8), 1629–1638. <https://doi.org/10.5194/nhess-19-1629-2019>

Thoithi, W., Blamey, R. C., & Reason, C. J. C. (2021). Dry Spells, Wet Days, and Their Trends Across Southern Africa During the Summer Rainy Season. *Geophysical Research Letters*, 48(5), e2020GL091041. <https://doi.org/10.1029/2020GL091041>

Amazon. (2022). GluonTS - Probabilistic Time Series Modeling in Python. Sphinx.

<https://ts.gluon.ai/stable/>

(2) Impact of the dry-day definition on Mediterranean extreme dry-spell analysis

Rivoire, P., Trambly, Y., Neppel, L., Hertig, E., & Vicente-Serrano, S. M. (2019). Impact of the dry-day definition on Mediterranean extreme dry-spell analysis. *Natural Hazards and Earth System Sciences*, 19(8), 1629–1638. <https://doi.org/10.5194/nhess-19-1629-2019>