

INSTITUTO DE EDUCACIÓN SUPERIOR



DEPARTAMENTO DE TECNOLOGÍA DIGITAL

BIG DATA Y CIENCIA DE DATOS

PROYECTO DE TESIS

**ANÁLISIS DE SERIES TEMPORALES Y PREDICCIÓN DE VARIABLES
AMBIENTALES ATMOSFÉRICAS EN EL OBSERVATORIO DE
MARCAPOMACOCHA**

AUTOR:

- Zegarra Pantoja, Alberto Cesar

ASESOR:

- Paraguay Arzapalo, Luis Enrique

LIMA - PERÚ

2025

ABSTRACT

This study aimed to analyze and forecast the temporal trends of key atmospheric variables: ozone (O_3), ultraviolet index (UV_MAX), and precipitation (PP), recorded at the Marcapomacocha Atmospheric Monitoring Observatory. Time series techniques and advanced predictive models were applied to historical data provided by SENAMHI, with the purpose of understanding the evolution and behavior of these variables over time.

A quantitative, non-experimental, and longitudinal research design was used. Exploratory data analysis tools were employed to characterize the time series, identify seasonal patterns, and detect potential anomalies. Subsequently, statistical forecasting models such as ARIMA and SARIMAX were implemented and evaluated using metrics including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to assess their accuracy and reliability.

Among the main findings, the SARIMAX model demonstrated strong performance in forecasting ozone levels, achieving a MAE of 2.27 DU and a relative error of 0.93%, highlighting its effectiveness in capturing the variable's behavior. For the UV index, the model also yielded acceptable predictions (MAE: 1.55; relative MAE: 14.33%) despite greater variability. In contrast, the precipitation model faced notable challenges (MAE: 3.08; relative MAE: 148.39%), reflecting the inherent complexity of this variable due to its extreme and discontinuous values.

In summary, the research objectives were achieved, successfully characterizing and modeling the time series of the studied atmospheric variables and generating useful predictions that can enhance environmental monitoring in high Andean regions. These findings lay the groundwork for future research incorporating hybrid models or machine learning techniques to strengthen atmospheric surveillance and response in similar contexts.

RESUMEN

El presente estudio tuvo como objetivo general analizar y predecir las tendencias temporales de variables atmosféricas clave tales como ozono (O_3), índice UV (UV_MAX) y precipitación (PP) registradas en el Observatorio de Vigilancia Atmosférica Marcapomacocha, mediante la aplicación de técnicas de series temporales y modelos predictivos avanzados. La investigación se centró en el análisis de datos históricos proporcionados por el SENAMHI, sin intervención directa en el fenómeno, con el propósito de comprender la evolución y comportamiento de estas variables en el tiempo.

La metodología empleada respondió a un enfoque cuantitativo con un diseño no experimental y de tipo longitudinal. Se utilizaron herramientas de análisis exploratorio de datos para caracterizar las series temporales, identificar patrones estacionales y detectar posibles anomalías. Posteriormente, se implementaron modelos estadísticos de predicción como ARIMA y SARIMAX, evaluados mediante métricas como el error cuadrático medio (RMSE) y el error absoluto medio (MAE), lo que permitió valorar su precisión y fiabilidad.

Entre los principales resultados, se destaca el buen desempeño del modelo SARIMAX en la predicción del ozono, con un MAE de 2.27 DU y un error relativo de 0.93%, lo que evidencia su capacidad para capturar adecuadamente el comportamiento de esta variable. En el caso del índice UV, el modelo también generó predicciones aceptables (MAE: 1.55; MAE relativo: 14.33%), aunque con una mayor variabilidad. Por otro lado, el modelo aplicado a la precipitación presentó mayores dificultades predictivas (MAE: 3.08; MAE relativo: 148.39%), reflejando la complejidad inherente a esta variable debido a sus valores extremos y discontinuos.

En síntesis, se cumplieron los objetivos propuestos, logrando no solo caracterizar y modelar las series temporales de las variables atmosféricas estudiadas, sino también generar predicciones útiles que pueden contribuir a mejorar el monitoreo ambiental en zonas altoandinas. Los hallazgos sientan las bases para futuras investigaciones que integren modelos híbridos o técnicas de aprendizaje automático, ampliando así las capacidades de vigilancia y respuesta frente a fenómenos atmosféricos en contextos similares.

ÍNDICE

INTRODUCCIÓN.....	8
CAPÍTULO I: DIAGNÓSTICO DE LA SITUACIÓN.....	10
1.1 PLANTEAMIENTO DEL PROBLEMA.....	10
1.1.1 Contextualización.....	10
1.1.2 Definición del problema.....	11
1.2 OBJETIVOS DE LA INVESTIGACIÓN.....	11
1.2.1 Objetivo General.....	11
1.2.2 Objetivos Específicos.....	12
1.3 JUSTIFICACIÓN E IMPORTANCIA.....	12
1.3.1 Conveniencia.....	12
1.3.2 Relevancia Social.....	13
1.4 DELIMITACIÓN Y ALCANCE DE LA SOLUCIÓN.....	14
CAPÍTULO II: MARCO CONTEXTUAL.....	15
2.1 DESCRIPCIÓN DE LA EMPRESA O ENTORNO DE APLICACIÓN.....	15
2.2 ESTADO DEL ARTE O ANTECEDENTES.....	17
2.3 FUNDAMENTOS TEÓRICOS.....	18
2.3.1 Series Temporales.....	18
2.3.2 Modelos ARIMA.....	18
2.3.3 Modelos SARIMA.....	19
2.3.4 Métricas de Evaluación.....	19
2.3.5 Ozono Atmosférico.....	20
2.3.6 Índice UV.....	21
2.3.7 Precipitación.....	22
2.3.8 Cambio Climático.....	23
2.3.9 Análisis Exploratorio de Datos.....	24
2.3.10 Machine Learning y Modelos Predictivos.....	24
2.4 DESCRIPCIÓN DE LOS INTERESADOS DE LA INVESTIGACIÓN.....	24
2.5 ANÁLISIS DE ALTERNATIVAS PARA LA SOLUCIÓN.....	25
2.5.1 Árbol de problemas.....	25
2.5.2 Árbol de objetivos.....	28
2.5.3 Alternativas de solución.....	30
2.5.4 Conclusión comparativa.....	32
CAPÍTULO III: PROPUESTA.....	33
3.1 MODALIDAD DE LA INVESTIGACIÓN.....	33
3.2 TÉCNICAS DE RECOLECCIÓN DE DATOS.....	33
3.3 INSTRUMENTOS DE RECOLECCIÓN DE DATOS.....	34
3.4 DESCRIPCIÓN DE LA METODOLOGÍA DE INVESTIGACIÓN.....	34
3.4.1 Metodología.....	34
3.4.2 Fases metodológicas.....	35
CAPÍTULO IV: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS.....	36
4.1 Análisis Descriptivo.....	36
4.2 Modelado Predictivo:.....	53
CAPÍTULO V: EVALUACIÓN ECONÓMICA-FINANCIERA.....	63

CONCLUSIONES.....	65
RECOMENDACIONES.....	66
BIBLIOGRAFÍA.....	67
ANÁLISIS DESCRIPTIVO.....	69
MODELADO PREDICTIVO.....	73

ÍNDICE DE TABLAS

Tabla 1. Comparación de tipos de estaciones meteorológicas y su equipamiento en el período 1963–2016.....	17
Tabla 2. Clasificación del Índice UV según nivel de riesgo y recomendaciones de protección.....	24
Tabla 3. Comparación de alternativas para la implementación de sistemas predictivos en el Observatorio Marcapomacocha.....	34
Tabla 4. Estadísticas descriptivas de las variables meteorológicas y ambientales registradas en la estación de monitoreo (2015–2024).....	38
Tabla 5. Estadísticas descriptivas de las variables O ₃ , UV_MAX y PP.....	42
Tabla 6. Vista tabular del conjunto de datos utilizado en el análisis.....	49
Tabla 7. Resultados de la prueba ADF para la serie de Ozono (O ₃).....	51
Tabla 8. Resultados de la prueba ADF para la serie del Índice UV Máximo (UV_MAX).....	51
Tabla 9. Resultados de la prueba ADF para la serie de Precipitación (PP).....	52
Tabla 10. SARIMAX Results de la variable O ₃	55
Tabla 11. Parámetros estimados del modelo SARIMAX(1, 0, 1) × (1, 1, 1, 91) para la variable O ₃	56
Tabla 12. Pruebas de diagnóstico del modelo SARIMAX para la variable O ₃	56
Tabla 13. Métricas de evaluación del modelo SARIMAX para la variable O ₃	58
Tabla 14. Resultados SARIMAX de la variable UV_MAX.....	59
Tabla 15. Parámetros estimados del modelo SARIMAX(1, 0, 1) × (1, 1, 1, 91) para la variable UV_MAX.....	59
Tabla 16. Pruebas de diagnóstico del modelo SARIMAX para la variable UV_MAX.....	60
Tabla 17. Métricas de evaluación del modelo SARIMAX para la variable UV_MAX.....	62
Tabla 18. SARIMAX Results de la variable PP.....	62
Tabla 19. Parámetros estimados del modelo SARIMAX(1, 0, 1) × (1, 1, 1, 91) para la variable PP....	63
Tabla 20. Pruebas de diagnóstico del modelo SARIMAX para la variable PP.....	63
Tabla 21. Métricas de evaluación del modelo SARIMAX para la variable PP.....	65
Tabla 22. Estimación de costos asociados al desarrollo del proyecto de tesis.....	66
Tabla 23. Estimación de costos por categoría en el desarrollo del proyecto.....	66

ÍNDICE DE FIGURAS

Figura 1. Espectrofotómetro Dobson Nº087 en operación en el OVA Marcapomacocha.....	21
Figura 2. Capacidad limitada del Observatorio Marcapomacocha para anticipar variaciones atmosféricas mediante el uso de modelos predictivos.....	25
Figura 3. Distribución temporal de los valores faltantes en la serie de datos.....	37
Figura 4. Distribución temporal de los valores faltantes luego de la imputación con KNN.....	37
Figura 5. Matriz de correlación entre las variables O3, UV_MAX y PP.....	38
Figura 6. Evolución temporal de los valores de O3.....	38
Figura 7. Evolución temporal del índice máximo de radiación UV (UV_MAX).....	39
Figura 8. Evolución temporal de la precipitación (PP).....	39
Figura 9. Distribución de la variable O3.....	40
Figura 10. Distribución de la variable UV_MAX.....	41
Figura 11. Distribución de la variable PP.....	41
Figura 12. Boxplots de la variable UV_MAX para detección de valores atípicos.....	42
Figura 13. Boxplots de la variable O3 para detección de valores atípicos.....	42
Figura 14. Boxplots de la variable PP para detección de valores atípicos.....	42
Figura 15. Valores diarios O ₃ , UV_MAX y PP durante el periodo de estudio.....	43
Figura 16. Distribución mensual de variables atmosféricas en la serie temporal analizada.....	43
Figura 17. Descomposición estacional aditiva del Ozono (O3).....	44
Figura 18. Descomposición estacional aditiva del Índice UV Máximo (UV_MAX).....	44
Figura 19. Descomposición estacional aditiva de la Precipitación (PP).....	45
Figura 20. Función de Autocorrelación (ACF) del Ozono (O3).....	45
Figura 21. Función de Autocorrelación (ACF) del Índice UV Máximo (UV_MAX).....	46
Figura 22. Función de Autocorrelación (ACF) de la Precipitación (PP).....	46
Figura 23. Serie temporal del conjunto de entrenamiento (Train).....	47
Figura 24. Descomposición aditiva de la serie temporal de Ozono (O3).....	48
Figura 25. Descomposición aditiva del componente UV_MAX en la serie de entrenamiento.....	48
Figura 26. Descomposición aditiva de la serie temporal de precipitación (PP).....	49
Figura 27. Funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para la serie de Ozono (O3).....	50
Figura 28. Funciones de autocorrelación y autocorrelación parcial para la serie UV_MAX.....	51
Figura 29. Función de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) para la serie de Precipitación (PP).....	52
Figura 30. Serie temporal observada y pronosticada de O ₃	55
Figura 31. Pronóstico de O ₃ con intervalo de confianza utilizando SARIMAX.....	56
Figura 32. Serie temporal observada y pronosticada de UV_MAX.....	59
Figura 33. Pronóstico de UV_MAX con intervalo de confianza utilizando SARIMAX.....	59
Figura 34. Serie temporal observada y pronosticada de PP.....	62
Figura 35. Pronóstico de PP con intervalo de confianza utilizando SARIMAX.....	62

INTRODUCCIÓN

El análisis de series temporales se ha consolidado como una herramienta fundamental en la comprensión y predicción de fenómenos atmosféricos, especialmente en el contexto del cambio climático y la gestión ambiental. En la Sierra Central del Perú, el Observatorio de Vigilancia Atmosférica Marcapomacocha desempeña un papel clave en la recolección y monitoreo de variables atmosféricas como el ozono, el índice de radiación ultravioleta y la precipitación. Estos datos resultan esenciales para evaluar la calidad del aire, prever riesgos ambientales y apoyar la toma de decisiones informadas a nivel regional.

A pesar de contar con registros sistemáticos, actualmente el Observatorio enfrenta limitaciones significativas debido a la falta de modelos predictivos avanzados que permitan anticipar con precisión las tendencias de estas variables. Esta carencia restringe la capacidad institucional para responder de manera oportuna ante cambios atmosféricos, dificultando así la planificación ambiental, la gestión de recursos hídricos y la implementación de políticas públicas adaptadas al contexto climático.

En este escenario, la presente investigación propone el análisis y modelado de series temporales con el objetivo de predecir variables clave como el ozono, el índice UV y la precipitación, utilizando modelos estadísticos como ARIMA y SARIMA, así como técnicas exploratorias y métricas de evaluación como RMSE y MAE. También considera los fundamentos teóricos del análisis temporal, el impacto del cambio climático en las variables estudiadas, y la posibilidad de integrar enfoques de machine learning para mejorar la precisión de las predicciones.

El presente documento se organiza en cinco capítulos que desarrollan de forma estructurada el proceso investigativo. El Capítulo I, Diagnóstico de la Situación, introduce el problema central que da origen al estudio, describiendo su aparición, causas y relevancia dentro de un contexto específico; en este apartado se delimitan los objetivos generales y específicos, se justifica la pertinencia del proyecto y se establecen tanto sus limitaciones como el alcance propuesto.

Capítulo II, Marco Contextual, se enfoca en la caracterización del entorno en el que se llevará a cabo la investigación, incluyendo la descripción de la empresa u organización involucrada, una revisión crítica de antecedentes y estudios previos vinculados al tema, la exposición de los fundamentos teóricos que sustentan la propuesta, el perfil de los participantes en la investigación y un análisis comparativo de posibles soluciones al problema identificado.

El Capítulo III, Propuesta, define el enfoque metodológico adoptado, precisando la modalidad de investigación, las técnicas e instrumentos empleados para la recolección de datos, así como el marco teórico que respalda dichas elecciones metodológicas.

En el Capítulo IV, Análisis e Interpretación de los Resultados, se presenta la aplicación de la metodología previamente establecida, estructurando el contenido según las fases del desarrollo y ofreciendo una interpretación clara y fundamentada de los hallazgos obtenidos, acompañada de los recursos necesarios para su comprensión y validación.

Por último, el Capítulo V, Evaluación Económica-Financiera, analiza la factibilidad del proyecto en términos de costos y recursos disponibles, detallando un presupuesto estimado que contempla insumos, servicios, equipamiento y posibles fuentes de financiamiento, con el fin de garantizar la viabilidad y sostenibilidad de la propuesta.

Series temporales, predicción atmosférica, ozono, índice UV, precipitación, ARIMA, SARIMA, machine learning, cambio climático, Perú.

CAPÍTULO I: DIAGNÓSTICO DE LA SITUACIÓN

1.1 PLANTEAMIENTO DEL PROBLEMA

El presente proyecto surge a partir de la necesidad crítica de fortalecer las capacidades del Observatorio de Vigilancia Atmosférica (OVA) de Marcapomacocha en cuanto al análisis y predicción de variables atmosféricas clave como el ozono, el índice UV y la precipitación. Actualmente, el observatorio no emplea herramientas analíticas ni modelos predictivos avanzados que permitan anticipar con precisión el comportamiento de dichas variables, lo que limita su capacidad de generar información oportuna y útil para la toma de decisiones en materia de gestión ambiental y adaptación al cambio climático. Esta necesidad se hace aún más evidente a la luz de lo señalado por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), que en una nota de prensa publicada en febrero de 2025 informa que el Observatorio de Vigilancia Atmosférica de Marcapomacocha ha sido oficialmente reconocido como estación regional por la Organización Meteorológica Mundial e incorporado al Programa de Vigilancia Atmosférica Global.

Este reconocimiento resalta la importancia científica e institucional del observatorio en la recopilación de datos sobre variables atmosféricas como el ozono y la radiación UV en zonas altoandinas. Sin embargo, este avance también pone en evidencia la necesidad de fortalecer sus capacidades analíticas mediante el desarrollo de herramientas predictivas que permitan transformar estos datos en información útil para anticipar cambios y apoyar la toma de decisiones en gestión ambiental.

1.1.1 Contextualización

El problema identificado se manifiesta en un contexto donde el OVA de Marcapomacocha, ubicado en la Sierra Central del Perú, cumple un rol fundamental en la vigilancia y monitoreo de las condiciones atmosféricas. A pesar de contar con una base de datos considerable sobre variables como ozono, radiación UV y niveles de precipitación, no se aplican métodos de análisis de series temporales ni modelos estadísticos o computacionales avanzados para proyectar su comportamiento futuro.

Esta situación se origina en parte por la falta de incorporación de tecnologías analíticas modernas y capacidades técnicas en modelamiento predictivo, lo cual deja a la institución en desventaja frente a los desafíos ambientales que enfrenta la región. La ausencia de estos modelos impide una planificación proactiva y una respuesta adecuada frente a fenómenos atmosféricos que pueden tener impactos significativos en la salud pública, los ecosistemas y el uso de los recursos naturales.

1.1.2 Definición del problema

El problema que se aborda en esta investigación es la ausencia de un enfoque predictivo sistemático basado en series temporales que permita al Observatorio de Marcapomacocha anticipar con precisión las tendencias del ozono, el índice UV y la precipitación. Esta deficiencia metodológica limita la efectividad de las acciones de monitoreo y gestión ambiental en una región particularmente vulnerable a los efectos del cambio climático.

Se trata de un problema concreto, localizado en el ámbito geográfico del observatorio, con implicancias directas en la toma de decisiones ambientales, y cuya solución es viable mediante la implementación de técnicas avanzadas de análisis y modelado de datos. Abordar este vacío permitirá fortalecer las capacidades técnicas del OVA, optimizar la gestión de los recursos atmosféricos y contribuir a una mejor preparación frente a eventos climáticos adversos en el corto y mediano plazo.

Si el problema identificado no se aborda con urgencia y precisión, las repercusiones serán múltiples y de largo alcance, tanto a nivel local como nacional. La ausencia de modelos predictivos robustos en el Observatorio de Vigilancia Atmosférica de Marcapomacocha implica que las vastas cantidades de datos recopilados seguirán siendo subutilizadas, limitándose a ofrecer diagnósticos pasivos del estado atmosférico sin capacidad de anticipación. Esta falta de visión prospectiva reduce drásticamente la eficacia de las políticas ambientales, debilita los mecanismos de alerta temprana ante eventos extremos tales como episodios de radiación UV intensificada o precipitaciones inusuales, y retrasa la capacidad de reacción ante riesgos que podrían afectar la salud humana, los ecosistemas de altura y la seguridad hídrica de poblaciones vulnerables.

Por consiguiente, el no desarrollar estas capacidades científicas internamente perpetúa la dependencia tecnológica de fuentes externas, lo que limita la autonomía del país en la toma de decisiones ambientales informadas. En un escenario de acelerado cambio climático, no actuar representa ceder terreno frente a amenazas ambientales cada vez más complejas y dinámicas, y desperdiciar una oportunidad clave para consolidar un centro de excelencia regional que, con las herramientas adecuadas, podría convertirse en un referente de vigilancia y resiliencia climática en la región andina.

1.2 OBJETIVOS DE LA INVESTIGACIÓN

1.2.1 Objetivo General

Analizar y predecir las tendencias temporales de variables atmosféricas clave (ozono, índice UV, y precipitación) en el Observatorio de Vigilancia Atmosférica Marcapomacocha, utilizando técnicas de series temporales y modelos predictivos avanzados.

1.2.2 Objetivos Específicos

- Explorar y describir las series temporales de las variables atmosféricas (ozono, UV, precipitación) recolectadas en el Observatorio de Marcapomacocha.
- Identificar patrones temporales como estacionalidades, tendencias y ciclos en las series temporales de las variables atmosféricas.
- Aplicar y evaluar modelos de series temporales (como ARIMA, SARIMA) para predecir las futuras mediciones de las variables atmosféricas, utilizando métricas de error como RMSE y MAE.

1.3 JUSTIFICACIÓN E IMPORTANCIA

1.3.1 Conveniencia

En primer lugar, este proyecto se plantea como una propuesta altamente conveniente debido a su enfoque en el análisis y predicción de variables atmosféricas clave como el ozono, el índice UV y la precipitación. Estas variables tienen un papel fundamental en la evaluación de la calidad del aire y en la vigilancia de los efectos del cambio climático. Utilizando técnicas de series temporales y modelos predictivos, se busca anticipar las variaciones en dichas variables con el fin de mejorar la planificación y gestión ambiental. Por ejemplo, al prever incrementos inusuales en los niveles de radiación UV, se podrían implementar campañas preventivas de salud pública con anticipación, protegiendo así a las poblaciones más vulnerables.

En segundo lugar, desde el punto de vista técnico, la viabilidad del proyecto es sólida. La recolección y limpieza de datos atmosféricos, así como la aplicación de modelos estadísticos como ARIMA y SARIMA, pueden realizarse eficazmente utilizando herramientas computacionales disponibles como Python o R. Asimismo, la metodología incluye fases claras como el análisis exploratorio de datos, el ajuste de modelos, su validación y la interpretación de resultados, lo que garantiza un tratamiento riguroso de la información.

Desde la perspectiva económica, también es pertinente destacar que el proyecto requiere una inversión razonable en términos de infraestructura computacional y capacitación del personal, pero ofrece beneficios que superan con creces esos costos. Al mejorar la capacidad de predicción atmosférica, se reducen los costos asociados a respuestas tardías frente a eventos climáticos extremos, como lluvias intensas o picos de radiación UV. Así, la inversión se justifica por el valor agregado en la toma de decisiones y la reducción de impactos negativos.

Por último, el proyecto presenta una notable viabilidad socioambiental, ya que contribuirá al entendimiento de los cambios atmosféricos en la Sierra Central del Perú, una región altamente expuesta a variaciones climáticas. La generación de modelos predictivos no solo fortalece la capacidad técnica del Observatorio de Marcapomacocha, sino que también apoya políticas públicas sostenibles de adaptación al cambio climático. De este modo, la iniciativa trasciende lo técnico y se proyecta como una herramienta de transformación social y ambiental.

1.3.2 Relevancia Social

En cuanto a la relevancia social, este proyecto tiene el potencial de generar un impacto positivo directo en la gestión ambiental y en la planificación de recursos naturales. La capacidad de predecir con antelación eventos asociados al ozono, la radiación UV o la precipitación permite que autoridades locales, instituciones académicas y organismos ambientales tomen decisiones informadas y actúen de forma preventiva. Por ejemplo, si se anticipa un periodo de lluvias intensas en una zona determinada, se podrían activar planes de contingencia para prevenir deslizamientos o afectaciones a la infraestructura crítica.

Por otro lado, es importante subrayar el carácter sostenible del proyecto. Las herramientas desarrolladas estarán diseñadas para funcionar a largo plazo, lo que significa que su utilidad no se limita al presente estudio, sino que podrán integrarse de forma continua en las prácticas del observatorio y en otras iniciativas relacionadas con la vigilancia ambiental. Esta proyección futura garantiza una base técnica sólida para la planificación ambiental y la adaptación progresiva a los efectos del cambio climático.

Finalmente, la propuesta se sustenta en un enfoque basado en evidencias, lo que fortalece su confiabilidad. Todo el análisis y las predicciones se realizarán a partir de datos reales obtenidos por el propio observatorio, y se aplicarán metodologías validadas a nivel internacional. Esto asegura que los resultados obtenidos no solo sean precisos y relevantes, sino también aplicables para la formulación de políticas y estrategias ambientales fundamentadas en conocimiento científico.

1.4 DELIMITACIÓN Y ALCANCE DE LA SOLUCIÓN

La presente investigación se delimita al análisis de datos registrados por el Observatorio de Vigilancia Atmosférica (OVA) Marcapomacocha, específicamente relacionados con las variables de ozono atmosférico (O_3), índice de radiación ultravioleta (IUV) y precipitación. Este conjunto de datos, validado y proporcionado por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), cubre el período comprendido entre los años 2015 y 2024, y está geográficamente restringido a la localidad de Marcapomacocha, en el distrito homónimo, provincia de Yauli, departamento de Junín, a una altitud aproximada de 4,500 m s.n.m. El recurso, de acceso público, se encuentra disponible en formato CSV, en su primera versión publicada (v1.0), bajo licencia Open Data Commons Attribution License, con contacto institucional a través del correo info.sgd@senamhi.gob.pe.

Las principales limitaciones del proyecto incluyen la presencia de registros vacíos que deben ser depurados antes del análisis, así como la imposibilidad de extrapolar directamente los resultados a otras zonas del país debido a la naturaleza localizada del monitoreo. Adicionalmente, aunque el dataset abarca una década de observaciones, el modelado se restringió a un subconjunto de datos comprendido entre junio de 2020 y mayo de 2023 (dos años para entrenamiento y uno para prueba), debido a la alteración sustancial de las condiciones climáticas provocadas por El Niño Costero 2023–2024, clasificado por el SENAMHI como el más intenso de las últimas dos décadas. Esta anomalía climatológica generó distorsiones significativas en los patrones de temperatura y precipitación, lo cual compromete la representatividad estadística de los datos recientes para la formulación de modelos predictivos generalizables.

En cuanto al alcance, el proyecto tiene como objetivo la construcción de modelos de predicción basados en series temporales que permitan anticipar el comportamiento de las variables atmosféricas mencionadas en horizontes temporales de corto y mediano plazo. No obstante, las herramientas que se construyan estarán diseñadas para ser escalables y sostenibles, permitiendo su incorporación continua a las prácticas institucionales del OVA, y sirviendo como base técnica a largo plazo para la vigilancia ambiental regional, reforzando así el compromiso del Perú con los estándares del Programa de Vigilancia Atmosférica Global de la OMM.

CAPÍTULO II: MARCO CONTEXTUAL

2.1 DESCRIPCIÓN DE LA EMPRESA O ENTORNO DE APLICACIÓN

El Observatorio de Vigilancia Atmosférica (OVA) Marcapomacocha, situado en la Sierra Central del Perú, específicamente en el distrito de Marcapomacocha, provincia de Yauli, departamento de Junín, opera bajo la supervisión del Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). Este observatorio desempeña un papel clave en la recolección, monitoreo y validación de datos atmosféricos fundamentales como el ozono atmosférico, el índice de radiación ultravioleta (UV) y la precipitación. A una altitud aproximada de 4,500 metros sobre el nivel del mar, su ubicación estratégica permite obtener registros representativos del comportamiento de la atmósfera en una zona altamente sensible a los cambios climáticos.

Según el Servicio Nacional de Meteorología e Hidrología del Perú (2016), el Observatorio de Vigilancia Atmosférica (OVA) de Marcapomacocha dispone de dos configuraciones de estación meteorológica convencional y automática, complementadas por equipamiento especializado para el seguimiento de la capa de ozono y la radiación ultravioleta. Cada sistema se apoya en un conjunto de instrumentos diseñados para garantizar tanto la continuidad histórica de los registros como la precisión necesaria en estudios de estacionalidad, tendencias y pronóstico climático.

Tabla 1.

Comparación de tipos de estaciones meteorológicas y su equipamiento en el período 1963–2016.

Tipo de estación / Equipamiento	Período	Instrumentos principales	Variables medidas
Convencional	1963 – 2016	<ul style="list-style-type: none">– Termómetros– Pluviómetros manuales y de cubeta– Veleta y anemómetro de cazoletas– Tanque de evaporación (evaporímetro de tanque abierto)– Pirómetro– Biometer	<ul style="list-style-type: none">– Temperatura– Precipitación– Velocidad y dirección del viento– Evaporación– Radiación solar– Radiación UV

Automática (EMA)	2000 – 2016	<ul style="list-style-type: none"> – Sensor de temperatura (termistor/termopar) – Pluviómetro de cubeta basculante con registro digital – Anemómetro ultrasónico – Sensor de humedad relativa – Evaporímetro digital – Piranómetro 	<ul style="list-style-type: none"> – Temperatura – Precipitación – Velocidad y dirección del viento – Humedad relativa – Evaporación – Radiación solar
Equipamiento especializado	2000 – 2016	<ul style="list-style-type: none"> – Espectrofotómetro Dobson (medición de la capa de ozono) – Satélite Aura OMI (sensor Ozone Monitoring Instrument) – Radiómetro Biometer (UV continuo) 	<ul style="list-style-type: none"> – Concentración de ozono total en columna atmósfera – Radiación ultravioleta

Nota. La tabla presenta los principales instrumentos utilizados y las variables medidas por diferentes tipos de estaciones meteorológicas entre 1963 y 2016, distinguiendo entre estaciones convencionales, automáticas y aquellas con equipamiento especializado.

El proyecto propuesto tiene múltiples áreas de aplicación dentro del observatorio. En el ámbito de la gestión ambiental, los modelos predictivos que se desarrolle permitirán anticipar cambios en las variables atmosféricas monitoreadas, fortaleciendo la capacidad del OVA para tomar decisiones informadas y oportunas en beneficio del medio ambiente. En el campo de la investigación científica, los resultados del análisis servirán como base para estudios climáticos, meteorológicos y de cambio climático, tanto a nivel nacional como internacional. A su vez, la generación de predicciones confiables contribuirá a la formulación de políticas públicas, al proporcionar insumos científicos sólidos que respalden la creación de normativas ambientales y estrategias de adaptación.

Por otra parte, el proyecto tendrá un impacto significativo en los programas de educación y sensibilización ambiental, ya que los hallazgos y herramientas desarrolladas podrán ser utilizadas como recursos didácticos en escuelas, universidades y campañas de concientización ciudadana. Finalmente, en el campo de la planificación de recursos hídricos, el análisis de patrones de precipitación contribuirá a una gestión más eficiente y sostenible del agua en una región donde este recurso es vital para la población y los ecosistemas.

2.2 ESTADO DEL ARTE O ANTECEDENTES

- ❖ Aplicación de series temporales en el estudio de precipitaciones en la ciudad del Cusco, 1964 – 2018. Universidad Andina del Cusco. (Colque Candia, 2020-09-11)

El estudio tuvo como objetivo analizar el comportamiento histórico de las precipitaciones en dicha ciudad mediante un enfoque cuantitativo y descriptivo, utilizando un diseño no experimental. Se aplicó la metodología Box-Jenkins, desarrollando un modelo ARIMA a través de las fases de identificación, estimación, validación y pronóstico. Los resultados mostraron que la serie no presentaba tendencias ni ciclos, pero sí una clara estacionalidad, lo cual permitió ajustar un modelo ARIMA(2,0,2)(1,1,0)[12], considerado el más adecuado. El estudio concluyó que las precipitaciones en Cusco presentan un comportamiento regular y estacional, demostrando la utilidad de los modelos ARIMA para analizar fenómenos climáticos.

- ❖ Pronóstico de precipitaciones en Baños, Ecuador mediante técnicas de Machine Learning: análisis del período 2014 – 2023. Universidad Central del Ecuador. (De la Cruz Córdoba & Fernández Carrión, 2024)

Por su parte, el trabajo buscó mejorar la predicción de precipitaciones en una zona altamente vulnerable al cambio climático. Frente a las limitaciones de los modelos tradicionales, se exploraron diversas técnicas de aprendizaje automático como Árboles de Decisión, SVM, Redes Neuronales y LSTM, evaluadas mediante métricas de error estándar. Aunque los Árboles de Decisión presentaron los resultados más precisos, el modelo LSTM fue destacado por su capacidad para captar dependencias temporales en los datos. El estudio concluyó que el Machine Learning representa una herramienta poderosa para anticipar eventos climáticos extremos y contribuir a la gestión del riesgo y los recursos hídricos.

- ❖ Monitoreo de la atmósfera en el observatorio de vigilancia atmosférica Marcapomacocha. Servicio Nacional de Meteorología e Hidrología del Perú (2022-2023)

Finalmente, el informe técnico evidenció el trabajo del OVA en la recopilación y análisis diario de variables atmosféricas como ozono, precipitación y radiación UV. La metodología incluyó la validación de datos mediante el filtrado de registros vacíos y el análisis mensual de tendencias. Los resultados destacaron una disminución general de aerosoles, aunque se registraron picos significativos de PM2.5 y PM10 en diciembre, así como niveles elevados de radiación UV durante el último trimestre del año. Si bien el

estudio mostró avances en la comprensión de la calidad del aire, también subrayó la necesidad de fortalecer el monitoreo constante e implementar estrategias preventivas para mitigar los efectos de la contaminación. La investigación pone en evidencia la importancia de contar con herramientas predictivas para anticipar variaciones atmosféricas, lo cual resulta clave para proteger la salud pública y la sostenibilidad ambiental.

2.3 FUNDAMENTOS TEÓRICOS

2.3.1 Series Temporales

Las series temporales constituyen un conjunto de observaciones cuantitativas organizadas cronológicamente, recolectadas en intervalos de tiempo regulares. Su análisis permite identificar patrones como tendencias, estacionalidades y ciclos, lo cual es fundamental para la predicción de fenómenos que evolucionan en el tiempo. En palabras de George E. P. Box, uno de los pioneros en este campo, “una serie temporal es un conjunto de observaciones ordenadas en el tiempo, donde el propósito del análisis es comprender la estructura subyacente y prever su comportamiento futuro”.

En el marco de la presente investigación, el análisis de series temporales resulta crucial para modelar y anticipar el comportamiento de variables atmosféricas observadas en el Observatorio de Vigilancia Atmosférica Marcapomacocha, tales como el ozono, el índice de radiación ultravioleta (UV) y la precipitación. El uso de modelos estadísticos como ARIMA y SARIMA permite no solo caracterizar la dinámica de estas variables, sino también generar predicciones confiables que puedan sustentar decisiones estratégicas en gestión ambiental y planificación ante el cambio climático.

2.3.2 Modelos ARIMA

Los modelos ARIMA (AutoRegressive Integrated Moving Average) fueron introducidos y desarrollados por George E. P. Box y Gwilym M. Jenkins como una de las metodologías más robustas para el análisis de series temporales. Esta clase de modelos combina tres componentes esenciales: la autorregresión (AR), que incorpora relaciones lineales entre los valores actuales y pasados de la serie; la integración (I), que transforma series no estacionarias en estacionarias mediante diferenciación; y la media móvil (MA), que ajusta la serie en función de errores residuales de observaciones previas.

Gracias a esta estructura flexible, los modelos ARIMA son especialmente eficaces para capturar patrones complejos en datos que presentan variaciones irregulares o tendencias sin una estacionalidad definida. En el contexto de la presente investigación, su aplicación permite modelar el

comportamiento temporal de variables atmosféricas como el ozono y la radiación UV, facilitando predicciones confiables aún en escenarios donde la dinámica de los datos no es estrictamente cíclica. Esto representa una herramienta clave para fortalecer las capacidades predictivas del Observatorio Marcapomacocha en su labor de monitoreo ambiental y toma de decisiones.

2.3.3 Modelos SARIMA

Cuando una serie temporal presenta patrones estacionales claramente definidos, los modelos ARIMA tradicionales pueden resultar insuficientes para capturar adecuadamente la estructura de los datos. En estos casos, se emplean los modelos SARIMA (Seasonal ARIMA), los cuales integran componentes estacionales adicionales de autorregresión (SAR), integración (I) y media móvil (SMA) en su formulación. Según Hyndman y Athanasopoulos (2018), este tipo de modelos es particularmente útil para datos que muestran comportamientos cíclicos, como las variaciones anuales del índice de radiación ultravioleta (UV) o los niveles de precipitación, influenciados por factores climáticos y astronómicos recurrentes.

La ventaja del modelo SARIMA radica en su capacidad para capturar simultáneamente las dinámicas de corto plazo y los efectos estacionales persistentes, lo que lo convierte en una herramienta idónea para el análisis y la predicción de variables atmosféricas en contextos como el del Observatorio Marcapomacocha, donde los fenómenos climáticos tienden a seguir ciclos anuales bien definidos.

2.3.4 Métricas de Evaluación:

La evaluación del desempeño de los modelos predictivos requiere el uso de métricas que cuantifiquen la precisión de las estimaciones respecto a los valores observados. El Root Mean Square Error (RMSE) mide la desviación cuadrática media entre los valores predichos y observados, siendo especialmente sensible a errores grandes, lo cual resulta útil cuando se desea penalizar severamente las desviaciones extremas.

Por su parte, el Mean Absolute Error (MAE) calcula el promedio de los errores absolutos, ofreciendo una medida más intuitiva de la precisión general del modelo. Según Willmott y Matsuura (2005), ambos indicadores son herramientas fundamentales en la validación de modelos de predicción climática y ambiental, dado que permiten valorar la fiabilidad y robustez de las proyecciones generadas.

2.3.5 Ozono Atmosférico

El ozono estratosférico (O_3) constituye una capa esencial situada entre los 15 y 35 km de altitud en la estratosfera, cuya función principal es absorber la radiación ultravioleta tipo B (UV-B), atenuando su llegada a la superficie terrestre y protegiendo así a los seres vivos de sus efectos perjudiciales. No obstante, esta capa ha sido progresivamente degradada por la acción de compuestos contaminantes de origen antropogénico, entre los que destacan los clorofluorocarbonos (CFCs), el metano (CH_4), el óxido nitroso (N_2O) y el monóxido de cloro (ClO).

La disminución de la concentración de ozono permite una mayor penetración de radiación ultravioleta, lo cual puede alterar procesos biológicos esenciales como la fotosíntesis en las plantas, afectar el desarrollo de los ecosistemas y aumentar significativamente el riesgo de enfermedades en humanos y animales, como el cáncer de piel y las cataratas (Sánchez Vega, 2016).

Para efectos de la presente investigación, se empleará como variable el ozono atmosférico promedio diario, expresado en Unidades Dobson (DU). Una Unidad Dobson es una medida estándar utilizada para cuantificar la cantidad total de ozono presente en una columna vertical de la atmósfera. Técnicamente, 1 DU equivale a una capa de ozono puro de 0.01 milímetros de espesor a temperatura y presión estándar ($0^{\circ}C$ y 1 atm). La relación matemática que representa esta unidad es:

$$\diamond \quad 1 \text{ DU} = 2.69 \times 10^{16} \text{ moléculas de } O_3 \text{ por cm}^2$$

Esta métrica permite interpretar con claridad la variación del ozono a lo largo del tiempo y evaluar su impacto sobre la radiación ultravioleta y otras variables atmosféricas. La utilización de Unidades Dobson como medida estándar permitirá analizar de manera precisa las fluctuaciones del ozono atmosférico y su relación con las variables ambientales consideradas en este estudio.

En el Observatorio de Vigilancia Atmosférica Marcapomacocha (OVA), la medición de la Columna Total de Ozono (CTO) se realiza mediante el uso de un Espectrofotómetro Dobson, instrumento óptico especializado que permite determinar de manera directa la cantidad total de ozono presente en una columna vertical de la atmósfera. Este dispositivo, operado manualmente por personal técnico capacitado, forma parte del sistema de vigilancia atmosférica implementado por el SENAMHI.

Asimismo, el monitoreo del ozono también se lleva a cabo mediante sensores remotos instalados en satélites. En particular, el satélite AURA, equipado con el instrumento OMI (Ozone

Monitoring Instrument), permite observar continuamente la concentración de ozono en la tropósfera superior, estratosfera y mesósfera. El OMI continúa el legado de más de 30 años de observaciones satelitales del ozono iniciadas por los instrumentos BUV (Backscatter Ultraviolet Detector, 1970) y TOMS (Total Ozone Mapping Spectrometer, 1978). Este dispositivo mide la luz solar reflejada y retrodispersada en el espectro ultravioleta y visible, y gracias a sus capacidades hiperespectrales, mejora significativamente la precisión y exactitud en la estimación del ozono total (SENAMHI, 2025).

Figura 1.

Espectrofotómetro Dobson N°087 en operación en el OVA Marcapomacocha.



Nota. Monitoreo de la atmósfera en el Observatorio de Vigilancia Atmosférica Marcapomacocha. Tomado de SENAMHI (2024).

2.3.6 Índice UV

El Índice Ultravioleta (UV) es un indicador adimensional que cuantifica la intensidad de la radiación UV solar (longitudes de onda entre 280 y 400 nm) que alcanza la superficie terrestre en un instante dado, y está directamente relacionado con la dosis de radiación que incide sobre la piel humana. Valores elevados del índice UV (superiores a 6) implican un riesgo alto o muy alto de quemaduras solares, envejecimiento prematuro de la piel y desarrollo de patologías oculares, como cataratas. Por ello, su monitoreo continuo permite diseñar esquemas de alertas sanitarias y adoptar medidas de protección (uso de bloqueadores solares, gafas UV, ropa adecuada) para minimizar el daño cutáneo y ocular (Servicio Nacional de Meteorología e Hidrología del Perú, 2016).

Tabla 2.

Clasificación del Índice UV según nivel de riesgo y recomendaciones de protección.

Índice UV	Nivel de riesgo	Recomendaciones generales
0 – 2	Bajo	Protección mínima requerida.
3 – 5	Moderado	Usar sombrero, gafas de sol y protector solar.
6 – 7	Alto	Buscar sombra al mediodía, protección obligatoria.
8 – 10	Muy alto	Evitar la exposición directa al sol.
11 o más	Extremo	Riesgo extremo: exposición solar no recomendada.

Nota. La tabla muestra la clasificación del Índice UV en función del nivel de riesgo para la salud humana y las recomendaciones generales de protección solar, conforme a estándares internacionales de seguridad frente a la radiación ultravioleta.

En el OVA Marcapomacocha, la cuantificación del índice UV se realiza mediante un radiómetro Biometer, instrumento que mide la irradiancia UV global incidente de forma continua, registrando datos en intervalos típicos de 1 hora. El radiómetro está instalado en un soporte orientado hacia el cielo abierto, a fin de captar todo el espectro UV y minimizar sombras o reflejos. Periódicamente, el equipo es calibrado con lámparas estándar de laboratorio y verificado contra un dosímetro de referencia, cumpliendo los protocolos de calidad de la World Meteorological Organization (WMO) y la Global Atmosphere Watch (GAW) (Servicio Nacional de Meteorología e Hidrología del Perú, 2016).

2.3.7 Precipitación

Según Sánchez San Román, la precipitación atmosférica incluye todas las formas de agua (líquida o sólida) que descienden desde la atmósfera hasta la superficie terrestre: lluvia, nieve, granizo, escarcha y rocío. Su clasificación según el mecanismo de formación distingue:

- ❖ Precipitaciones ciclónicas, asociadas a sistemas frontales y ciclones extratropicales.
- ❖ Precipitaciones convectivas, típicas de tormentas de calentamiento diurno que generan ascensos rápidos de aire caliente.
- ❖ Precipitaciones orográficas, producidas por el levantamiento forzado de masas de aire húmedo al chocar con relieves montañosos.

En Marcapomacocha, la cuantificación de la precipitación se efectúa con una Estación Meteorológica Automática (EMA) equipada con un pluviómetro de cubeta basculante. Cada vez que se acumulan 0,2 mm de lluvia, la cubeta se vuelca y envía un pulso eléctrico al registrador, lo que permite un registro temporal con resolución de minutos. Además, la EMA mide simultáneamente temperatura, velocidad y dirección del viento, humedad relativa, evaporación y radiación solar, generando un conjunto integrado de variables meteorológicas (Servicio Nacional de Meteorología e Hidrología del Perú, 2016).

La alta frecuencia de muestreo y los procesos de validación internos (filtrado de pulsos espurios, comparación con pluviómetros manuales de control) garantizan la consistencia de la serie histórica de precipitación. Estos datos son fundamentales para la modelación hidrológica de cuencas andinas, la estimación de la recarga de acuíferos y la gestión de riesgos asociados a inundaciones y sequías.

El estudio de la precipitación es esencial dentro de la hidrología regional, ya que constituye la principal fuente de ingreso de agua a una cuenca. Resulta vital para la planificación de obras públicas, la gestión de recursos hídricos, la prevención de eventos extremos como inundaciones o sequías, y el análisis de fenómenos como la erosión. Para modelar y pronosticar su comportamiento, el análisis de series temporales es una herramienta clave, permitiendo anticipar tendencias y variaciones estacionales relevantes.

2.3.8 Cambio Climático

El cambio climático constituye un fenómeno de alcance global que provoca alteraciones en diversas variables atmosféricas como el ozono, el índice de radiación ultravioleta (UV) y la precipitación. Estas modificaciones, tanto naturales como inducidas por actividades humanas, deben ser consideradas en el diseño de modelos predictivos, ya que influyen en las tendencias y comportamientos observados a largo plazo.

Según la Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC), este fenómeno se define como una alteración del clima atribuible, directa o indirectamente, a la actividad humana que modifica la composición de la atmósfera mundial, sumándose a la variabilidad climática natural existente en períodos de tiempo comparables.

Incorporar el impacto del cambio climático en los modelos de predicción resulta fundamental para mejorar la precisión de los análisis atmosféricos y apoyar estrategias de adaptación ambiental.

2.3.9 Análisis Exploratorio de Datos

El Análisis Exploratorio de Datos (EDA) es una fase fundamental en todo proceso de modelamiento estadístico, cuyo objetivo es comprender la estructura y las características internas de los datos antes de aplicar técnicas predictivas. Según (Turkey, 1977), precursor del EDA, este enfoque enfatiza el uso de métodos gráficos y estadísticos para identificar patrones, tendencias, estacionalidades y anomalías en los conjuntos de datos, facilitando así una interpretación más profunda y adecuada del fenómeno estudiado.

En el contexto de la presente investigación, el EDA resulta esencial para analizar las series temporales de variables atmosféricas como el ozono, el índice UV y la precipitación, permitiendo detectar comportamientos irregulares o estacionales que deben ser considerados en la construcción de modelos predictivos robustos. Al realizar un análisis preliminar exhaustivo, se garantiza que los modelos aplicados sean más pertinentes y que las predicciones resultantes tengan una mayor precisión y fiabilidad.

2.3.10 Machine Learning y Modelos Predictivos

El machine learning constituye un enfoque moderno en el análisis de datos que permite construir modelos predictivos altamente precisos mediante el uso de grandes volúmenes de información. Según (James et al., 2017), las técnicas de aprendizaje automático incluyen métodos de regresión, clasificación y redes neuronales, los cuales son capaces de identificar patrones complejos en los datos y generar predicciones más confiables que los modelos estadísticos tradicionales.

En el marco de la presente investigación, la aplicación de modelos de machine learning resulta estratégica para mejorar la predicción de variables atmosféricas monitoreadas en el Observatorio de Vigilancia Atmosférica Marcapomacocha. El empleo de técnicas avanzadas permite capturar mejor las dinámicas no lineales y los comportamientos atípicos que podrían escapar a los modelos clásicos, incrementando así la calidad de las alertas ambientales y fortaleciendo la gestión regional ante los impactos del cambio climático.

2.4 DESCRIPCIÓN DE LOS INTERESADOS DE LA INVESTIGACIÓN

El presente proyecto de investigación es desarrollado por Alberto Cesar Gregorio Zegarra Pantoja, estudiante del sexto y último ciclo de la carrera de Big Data y Ciencia de Datos en el Instituto TECSUP. Su formación académica abarca un sólido conjunto de cursos técnicos y analíticos vinculados al análisis de datos, minería de datos, estadística aplicada, modelamiento de negocios, tecnologías emergentes, sistemas de gestión de datos, analítica empresarial y ciencia de datos en general.

Aunque no cuenta con experiencia laboral previa, el investigador ha adquirido habilidades clave a lo largo de su formación académica, desarrollando competencias que incluyen la construcción de modelos predictivos, el procesamiento de grandes volúmenes de datos, y la aplicación de técnicas de visualización y análisis estadístico para la toma de decisiones. Además, ha participado en cursos orientados al desarrollo profesional, liderazgo y sostenibilidad, lo cual refuerza su capacidad para asumir un proyecto de esta envergadura de forma integral.

Al ser un proyecto individual, el estudiante asume todos los roles clave en el desarrollo de la investigación: desde la recolección y limpieza de datos, la aplicación de modelos de series temporales, hasta el análisis de resultados y la formulación de conclusiones. Su compromiso con el rigor académico y la innovación tecnológica constituye una fortaleza fundamental para el cumplimiento de los objetivos planteados.

Luis Paraguay Arzapalo, docente e investigador de TECSUP, cumple el rol de guía metodológico y académico en el desarrollo del presente trabajo. Su experiencia en Dirección de Tecnologías de la Información, desarrollo de software, inteligencia de negocios y ciencia de datos, así como su conocimiento en ingeniería de datos, aprendizaje automático y minería de datos, contribuyen a garantizar la calidad científica y técnica de la investigación.

2.5 ANÁLISIS DE ALTERNATIVAS PARA LA SOLUCIÓN

2.5.1 Árbol de problemas

Un árbol de problemas permite clarificar de manera estructurada las causas y los efectos de una situación problemática. En este caso, se aplica para analizar la capacidad limitada del Observatorio Marcapomacocha para anticipar variaciones atmosféricas mediante el uso de modelos predictivos.

Figura 2.

Capacidad limitada del Observatorio Marcapomacocha para anticipar variaciones atmosféricas mediante el uso de modelos predictivos.



Nota. El mapa mental presenta el árbol de problemas desarrollado para mostrar las causas y efectos del problema identificado. Elaboración propia.

❖ Problema central

- Capacidad limitada del Observatorio Marcapomacocha para anticipar variaciones atmosféricas mediante el uso de modelos predictivos.

Radica en la ausencia de un enfoque predictivo sistemático basado en el análisis de series temporales que permita anticipar con precisión las tendencias del ozono, el índice UV y la precipitación. Esta carencia metodológica limita la capacidad del observatorio para generar pronósticos confiables y oportunos, afectando directamente la eficacia de las acciones de monitoreo y gestión ambiental en una región vulnerable a los efectos del cambio climático.

❖ Causas

- Aprovechamiento insuficiente de los datos históricos atmosféricos.

El Observatorio de Vigilancia Atmosférica (OVA) Marcapomacocha utiliza sus datos principalmente en informes descriptivos con análisis básicos, sin aplicar modelado predictivo, inteligencia artificial ni análisis avanzado de tendencias. No cuenta con sistemas automatizados de procesamiento en tiempo real ni detección de anomalías, y la validación sigue siendo manual. Su infraestructura actual no permite un análisis automatizado, limitando el uso de los datos para aplicaciones avanzadas.

- Débil articulación institucional con el ámbito académico y científico.

Aunque el OVA forma parte de redes internacionales como GAW, no desarrolla proyectos colaborativos con universidades ni produce publicaciones científicas en conjunto. Sus convenios son principalmente técnicos y no hay mecanismos para transferencia tecnológica o innovación conjunta. Además, sus datos se publican en formatos estáticos, sin protocolos que faciliten su integración en investigaciones académicas o desarrollos externos.

- Aplicación restringida de modelos predictivos basados en series temporales.

El observatorio dispone de sensores satelitales como el GOES 16, pero no aplica modelos predictivos avanzados como LSTM o SARIMA para capturar la estacionalidad y variabilidad atmosférica. Esta limitación se agrava por la escasa formación técnica del personal en modelado predictivo, lo que afecta la precisión de los pronósticos y limita la capacidad para anticipar eventos extremos o gestionar eficazmente recursos como el agua.

❖ Efectos

- Baja capacidad operativa ante fenómenos climáticos extremos.

La ausencia de modelos predictivos avanzados y sistemas automatizados limita la capacidad del observatorio para emitir alertas tempranas y planificar respuestas eficaces frente a eventos climáticos extremos. Esto afecta directamente la preparación de autoridades y comunidades ante fenómenos como radiación ultravioleta elevada, precipitaciones intensas o variaciones térmicas abruptas. Como resultado, se incrementan los riesgos para la salud pública, la seguridad de la población y la integridad de la infraestructura local, ya que no se cuenta con información oportuna para implementar medidas preventivas.

- Deficiencias en la planificación ambiental y gestión de recursos.

La limitada capacidad del OVA Marcapomacocha para anticipar condiciones atmosféricas reduce la eficacia en la formulación de estrategias ambientales y la gestión sostenible de recursos como el agua. En contextos de sequías o lluvias intensas, la falta de predicciones precisas impide tomar decisiones proactivas para mitigar riesgos. Asimismo, la ausencia de evidencia anticipada en las políticas públicas ambientales lleva a una planificación reactiva, basada en eventos ya ocurridos, lo cual limita su capacidad de adaptación frente a la variabilidad climática y el cambio global.

➤ Reducción del impacto institucional y científico del observatorio.

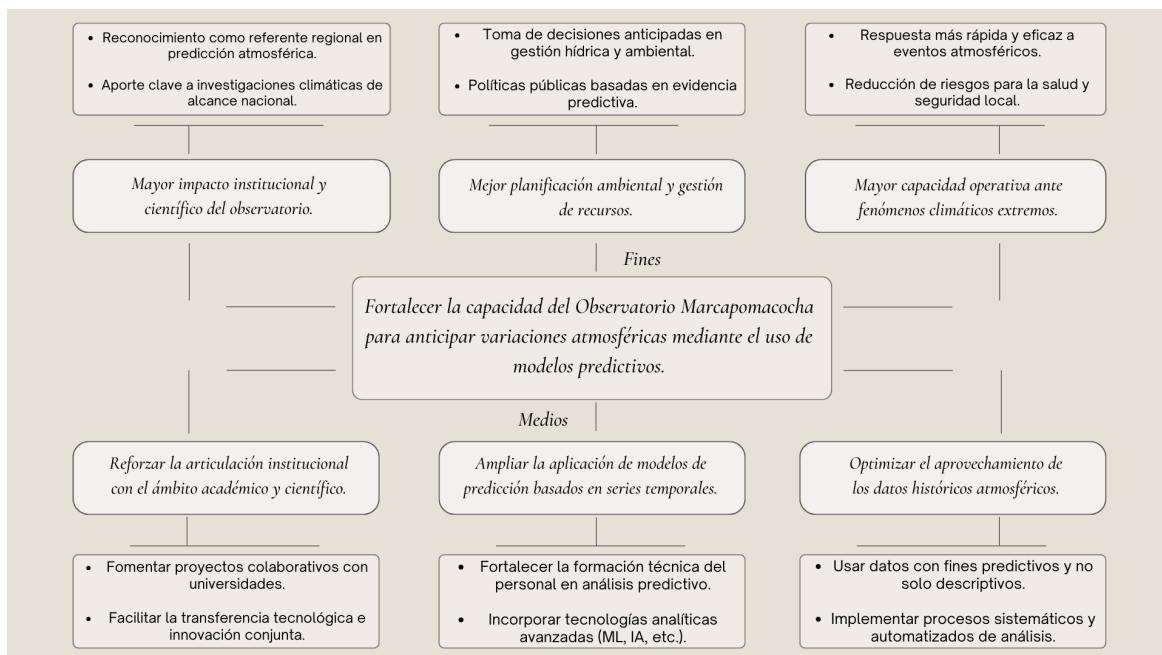
La escasa articulación con el ámbito académico y el bajo uso de sus datos en estudios colaborativos restringen el impacto científico del observatorio. Aunque pertenece a redes internacionales como GAW, su participación en publicaciones indexadas y proyectos de investigación es mínima, lo que reduce su visibilidad y liderazgo en la región. Además, el poco aprovechamiento de datos clave —como los de aerosoles, ozono y gases de efecto invernadero— en investigaciones climáticas globales representa una oportunidad perdida para posicionarse como referente técnico en el monitoreo atmosférico.

2.5.2 Árbol de objetivos

Un árbol de objetivos es la contraparte del árbol de problemas. Mientras el árbol de problemas parte de una situación negativa y la descompone en causas y efectos, el árbol de objetivos transforma esa situación en un estado deseado y lo descompone en medios y fines.

Figura 3.

Fortalecer la capacidad del Observatorio Marcapomacocha para anticipar variaciones atmosféricas mediante el uso de modelos predictivos.



Nota. El mapa mental presenta el árbol de objetivos desarrollado para mostrar los medios y fines. Elaboración propia.

❖ Objetivo central

- Fortalecer la capacidad del Observatorio Marcapomacocha para anticipar variaciones atmosféricas mediante el uso de modelos predictivos.

Este objetivo busca transformar la actual limitación en la anticipación de fenómenos atmosféricos en una capacidad sistemática y precisa para generar pronósticos confiables. Para ello, se propone incorporar el análisis de series temporales y tecnologías innovadoras, lo que permitirá mejorar la gestión ambiental y fomentar una toma de decisiones proactiva frente a eventos climáticos extremos. De esta manera, se incrementará la eficacia operativa del observatorio y se fortalecerá su impacto tanto científico como institucional.

❖ Medios

- Optimizar el aprovechamiento de los datos históricos atmosféricos.

La transformación de datos descriptivos en predictivos es clave para superar la limitación actual del Observatorio, que prioriza informes estáticos y análisis básicos. Al implementar procesos sistemáticos y automatizados, se podrá explotar el potencial de décadas de datos atmosféricos almacenados. Esto permitirá identificar patrones ocultos, correlaciones entre variables climáticas y precursores de eventos extremos. La validación automatizada, complementada con técnicas de imputación de datos faltantes, aumentará la confiabilidad de los insumos para modelos avanzados.

- Ampliar la aplicación de modelos de predicción basados en series temporales.

La incorporación de modelos resolverá la actual brecha en el análisis de estacionalidad y variabilidad atmosférica. Estos algoritmos, entrenados con los datos históricos del observatorio y enriquecidos con información del sensor GOES 16, pueden prever cambios en parámetros críticos como presión barométrica o humedad relativa con ventanas de anticipación.

- Reforzar la articulación institucional con el ámbito académico y científico.

Fundamental para potenciar la capacidad del Observatorio Marcapomacocha. Establecer alianzas estratégicas con universidades reconocidas, permitirá desarrollar soluciones tecnológicas conjuntas adaptadas a los microclimas andinos. Esta colaboración puede concretarse mediante proyectos de investigación aplicada que ajusten modelos de aprendizaje automático (ML) a las condiciones locales, así como a través de la implementación de plataformas de datos abiertos en formatos estándar.

❖ Fines

- Mayor capacidad operativa ante fenómenos climáticos extremos.

La implementación de modelos predictivos avanzados y sistemas automatizados permitirá al Observatorio Marcapomacocha emitir alertas tempranas con mayor precisión y anticipación. Esto se traducirá en una respuesta más rápida y eficaz frente a eventos atmosféricos adversos como radiación ultravioleta elevada, precipitaciones intensas o cambios térmicos abruptos.

- Mejor planificación ambiental y gestión de recursos.

Contar con predicciones atmosféricas confiables y oportunas facilita la toma de decisiones anticipadas en la gestión hídrica y ambiental, aspectos fundamentales para la sostenibilidad de la región andina donde se ubica el observatorio. La capacidad de anticipar condiciones como sequías o lluvias intensas permite diseñar estrategias proactivas que mitiguen riesgos y optimicen el uso de recursos naturales, en especial el agua. Asimismo, la disponibilidad de evidencia predictiva robusta contribuye a fundamentar políticas públicas ambientales más efectivas y adaptativas, superando la tradicional planificación reactiva basada en eventos pasados.

- Mayor impacto institucional y científico del observatorio.

Al fortalecer sus capacidades técnicas y establecer vínculos con el ámbito académico y científico, el Observatorio Marcapomacocha podrá consolidarse como un referente regional en predicción atmosférica. La incorporación de modelos avanzados y la generación de datos de alta calidad incrementarán su visibilidad y prestigio en redes internacionales como la GAW, así como su participación en publicaciones indexadas y proyectos de investigación colaborativos. Este mayor impacto institucional se traducirá en un aporte clave a estudios climáticos de alcance nacional, posicionando al observatorio como un actor fundamental en el monitoreo y análisis de variables atmosféricas críticas, como aerosoles, ozono y gases de efecto invernadero.

2.5.3 Alternativas de solución

Con base en el árbol de objetivos elaborado, se identificaron tres alternativas viables para abordar el problema central del proyecto. Estas opciones buscan fortalecer la capacidad predictiva del Observatorio Marcapomacocha a través de diferentes estrategias técnico-institucionales. A continuación, se describen brevemente:

❖ **Alternativa 1: Plataforma Analítica Interna**

Esta opción plantea el desarrollo de una herramienta informática propia orientada al análisis predictivo de variables atmosféricas, utilizando los datos históricos disponibles en el observatorio. Contempla, además, la automatización de procesos y la capacitación del personal, con el objetivo de fortalecer la autonomía técnica institucional. Si bien supone desafíos iniciales en cuanto a capacidades instaladas y recursos humanos especializados, también permite sentar las bases para una infraestructura analítica sostenible en el mediano y largo plazo.

❖ **Alternativa 2: Alianza Estratégica con Instituciones Académicas**

Esta propuesta se enfoca en establecer convenios de colaboración con universidades o centros de investigación, con el fin de co-diseñar e implementar modelos predictivos. Sus pilares son la transferencia de conocimiento, la integración metodológica y la formación conjunta. Si bien esta alternativa puede enriquecer significativamente las capacidades institucionales, su ejecución depende de factores exógenos, como la disposición de los socios estratégicos, la disponibilidad académica y la gestión administrativa de los convenios, lo cual puede generar demoras considerables y limitar el control directo del observatorio sobre los procesos clave.

❖ **Alternativa 3: Implementación de Modelos de Series Temporales**

Esta alternativa constituye la propuesta central de la presente investigación y se posiciona como la más sólida, factible y estratégicamente alineada con los recursos técnicos y operativos actualmente disponibles. Se fundamenta en la aplicación de Modelos de Series Temporales, una técnica estadística ampliamente validada para el análisis y pronóstico de variables atmosféricas como el ozono, el índice UV y la precipitación. Aprovecha eficientemente los registros históricos existentes, transformándolos en herramientas predictivas de alto valor operativo para una gestión ambiental proactiva, autónoma y basada en evidencia científica. Su implementación resulta técnicamente viable, escalable y adaptable a futuras necesidades del observatorio, y permite impulsar mejoras inmediatas en la capacidad institucional. Esta solución no solo responde con precisión al problema identificado, sino que proyecta al observatorio hacia un modelo de vigilancia atmosférica resiliente, tecnológicamente avanzado y alineado con estándares internacionales en la gestión climática.

A continuación, se presenta un análisis comparativo de las tres alternativas identificadas, evaluadas con base en criterios clave como el impacto predictivo, la factibilidad de implementación, la sostenibilidad operativa, el tiempo estimado para su ejecución, el nivel de autonomía institucional que otorgan y su potencial de transferibilidad hacia otros contextos o estaciones similares. Esta

comparación tiene como propósito identificar la opción más eficaz y estratégicamente alineada con los objetivos del Observatorio Marcapomacocha.

2.5.4 Conclusión comparativa

Tras el análisis comparativo de las alternativas propuestas, se concluye que la implementación de modelos de series temporales, en particular ARIMA y SARIMA, constituye la opción más eficaz, sostenible y estratégicamente alineada con los objetivos del Observatorio Marcapomacocha. Esta solución no solo responde de manera directa al problema identificado, sino que aprovecha al máximo los registros históricos disponibles, permitiendo convertirlos en herramientas predictivas de alto valor operativo. Su viabilidad técnica, basada en métodos estadísticos validados y replicables, garantiza una mejora significativa en la autonomía institucional y en la capacidad de respuesta ante variaciones atmosféricas.

Tabla 3.

Comparación de alternativas para la implementación de sistemas predictivos en el Observatorio Marcapomacocha.

Criterio	Alternativa 1 <i>Plataforma Analítica Interna</i>	Alternativa 2 <i>Alianza Académica</i>	Alternativa 3 <i>Modelos de Series Temporales</i>
Impacto predictivo	Alto (depende del desarrollo de capacidades internas y la calidad del software creado)	Alto (potenciado por el conocimiento externo, pero sujeto a su disponibilidad)	Muy alto (uso directo de técnicas validadas con resultados reproducibles)
Tiempo de implementación	Medio-largo (etapas de diseño, prueba y capacitación)	Corto-medio (depende de los tiempos de coordinación interinstitucional)	Medio (los modelos están disponibles y listos para adaptación)
Costo	Alto (inversión en infraestructura, formación técnica y soporte)	Bajo (uso de recursos compartidos o convenios sin costo directo)	Moderado (uso de software libre y recursos humanos internos)
Sostenibilidad	Media (requiere soporte técnico constante y actualización de sistemas)	Baja-media (alta rotación académica o cambios institucionales pueden afectar)	Alta (metodología replicable y sostenida por personal formado localmente)
Autonomía institucional	Alta (control completo sobre datos, desarrollo y uso de la herramienta)	Baja (dependencia metodológica y tecnológica de socios externos)	Alta (fortalece las competencias internas sin depender de terceros)
Transferibilidad	Media (adaptación requiere personal capacitado y soporte técnico)	Alta (modelo académico puede escalarse mediante nuevos convenios)	Alta (modelo técnico replicable en otras estaciones con datos similares)

Nota. La alternativa 3 (Modelos de Series Temporales) incluye técnicas ARIMA y SARIMA aplicables con software libre, lo que reduce costos y mejora la replicabilidad del modelo en otras estaciones hidrometeorológicas. Elaboración propia.

CAPÍTULO III: PROPUESTA

3.1 MODALIDAD DE LA INVESTIGACIÓN

La presente investigación se enmarca en la modalidad documental, la cual, según Hernández Sampieri et al. (2014), consiste en la recopilación, análisis e interpretación sistemática de información proveniente de documentos escritos, informes técnicos, artículos científicos y otras fuentes secundarias que abordan el fenómeno de estudio. Esta modalidad es adecuada para el proyecto porque se basa en la revisión exhaustiva de literatura especializada sobre análisis de series temporales y modelos predictivos aplicados a variables atmosféricas, así como en el análisis de datos secundarios proporcionados por el Observatorio de Vigilancia Atmosférica Marcapomacocha. De esta forma, se evita la manipulación directa del fenómeno y se aprovechan fuentes confiables para fundamentar teóricamente la investigación y desarrollar el análisis cuantitativo correspondiente.

Esta modalidad permite realizar un estudio riguroso y sistemático sin necesidad de intervención directa en el campo, apoyándose en la información ya generada y disponible, lo que es coherente con el enfoque cuantitativo y el diseño no experimental que se emplean en la investigación.

3.2 TÉCNICAS DE RECOLECCIÓN DE DATOS

La recolección de datos en esta investigación se fundamenta principalmente en técnicas documentales y de uso de datos secundarios, conforme a lo establecido por Hernández Sampieri et al. (2014).

- ❖ Investigación documental: Esta técnica consiste en la búsqueda, selección, análisis y síntesis de información contenida en documentos escritos, tales como artículos científicos, informes técnicos, libros y bases de datos especializadas. Su propósito es obtener un marco teórico y metodológico sólido que sustente el estudio, permitiendo conocer el estado actual del conocimiento sobre análisis de series temporales y modelos predictivos aplicados a variables atmosféricas.
- ❖ Uso de datos secundarios: Se emplean datos atmosféricos históricos previamente recolectados por el Observatorio de Vigilancia Atmosférica Marcapomacocha. La utilización de datos secundarios es una técnica válida cuando se dispone de registros confiables y pertinentes, lo que evita la necesidad de realizar una recolección primaria y permite un análisis cuantitativo riguroso.

3.3 INSTRUMENTOS DE RECOLECCIÓN DE DATOS

En esta investigación, los instrumentos para la recolección de datos están orientados a la obtención y organización de información documental y de registros históricos, conforme a lo establecido por Hernández Sampieri et al. (2014) y la literatura especializada en investigación documental.

- ❖ Fichas de contenido y fichas bibliográficas: Se emplean para registrar y organizar de manera sistemática la información extraída de documentos, artículos científicos, informes técnicos y otras fuentes académicas. Las fichas de contenido permiten anotar datos relevantes, ideas principales y comentarios críticos, facilitando la posterior interpretación y análisis (Hernández Sampieri et al., 2014, p. 388). Las fichas bibliográficas contienen los datos de referencia de cada fuente, asegurando un adecuado control y citación de las mismas.
- ❖ Bases de datos digitales: Se utilizan para almacenar y procesar los datos secundarios proporcionados por el Observatorio de Vigilancia Atmosférica Marcapomacocha. Estos registros históricos de variables atmosféricas se organizan en formatos compatibles con software estadístico para su análisis cuantitativo.
- ❖ Documentos de archivo e informes técnicos: Constituyen fuentes primarias de información que complementan la revisión documental. La recopilación y análisis de estos documentos permiten contextualizar los resultados y validar la información obtenida de otras fuentes.

3.4 DESCRIPCIÓN DE LA METODOLOGÍA DE INVESTIGACIÓN

3.4.1 Metodología

La metodología de esta investigación se fundamenta en un enfoque cuantitativo y un diseño no experimental de tipo longitudinal, tal como lo define Hernández Sampieri et al. (2014).

El enfoque cuantitativo permite analizar y medir variables mediante técnicas estadísticas para establecer relaciones y patrones entre ellas, proporcionando resultados objetivos y generalizables. Por su parte, el diseño no experimental se caracteriza porque el investigador no manipula deliberadamente las variables, sino que observa y analiza los fenómenos en su contexto natural, sin intervenir en ellos.

El diseño longitudinal implica la recolección y análisis de datos en múltiples momentos a lo largo del tiempo, lo que facilita estudiar la evolución y los cambios en las variables de interés. En este caso, se analizarán datos históricos de variables atmosféricas ya recolectados por el Observatorio de Vigilancia Atmosférica Marcapomacocha, sin intervención directa en el fenómeno, permitiendo observar cómo estas variables cambian y se relacionan a lo largo del tiempo.

3.4.2 Fases metodológicas

- ❖ Análisis descriptivo: Se aplicarán técnicas de análisis exploratorio de datos (EDA) para resumir y visualizar las características de las variables atmosféricas. Se calcularán estadísticas descriptivas, además de graficar series temporales para evaluar el comportamiento de las variables a lo largo del tiempo.
- ❖ Modelado predictivo: Se implementarán modelos estadísticos de series temporales, como ARIMA y SARIMA, para predecir variables atmosféricas específicas (ozono, índice UV, precipitación). La validación de estos modelos se realizará mediante métricas estadísticas como el error cuadrático medio (RMSE) y el error absoluto medio (MAE), asegurando la precisión y confiabilidad de las predicciones.

CAPÍTULO IV: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

4.1 Análisis Descriptivo

En esta sección se presentan los resultados obtenidos mediante el análisis exploratorio de los datos atmosféricos registrados entre 2015 y 2024 en el Observatorio de Vigilancia Atmosférica Marcapomacocha. Se incluyen representaciones gráficas como series temporales, diagramas de caja, histogramas y mapas de calor que permiten identificar tendencias, patrones estacionales, anomalías y variabilidad en las variables de estudio: concentración de ozono, índice UV y precipitación.

Tabla 4.

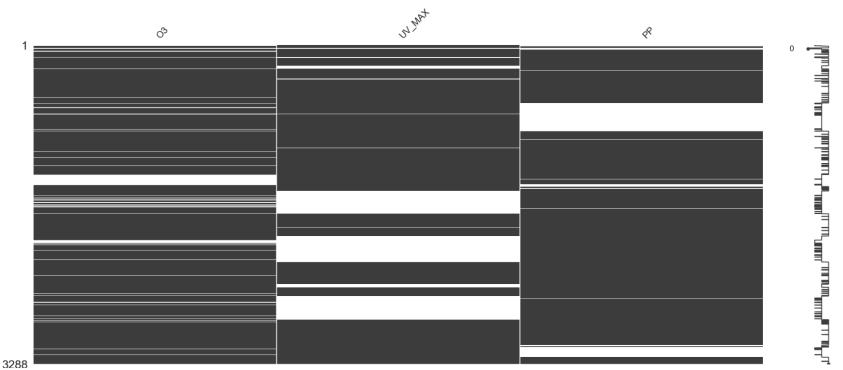
Estadísticas descriptivas de las variables meteorológicas y ambientales registradas en la estación de monitoreo (2015–2024).

Variable	Conteo	Media	Mínimo	25%	50%	75%	Máximo	Desv. estándar
ID	3,288	1,644.35	1.00	822.75	1644.50	2466.25	3287.00	949.05
FECHA	3,288	—	2015-06-01	2017-08-31	2019-11-30	2022-03-31	2024-05-31	0.0
LONGITUD	3,288	-76.3288	-76.3288	-76.3288	-76.3288	-76.3288	-76.3288	0.0
LATITUD	3288	-11.4404	-11.4404	-11.4404	-11.4404	-11.4404	-11.4404	0.0
ALTITUD	3288	4,443.0	4443.0	4443.0	4443.0	4443.0	4443.0	0.0
O3	2935	243.09	232.0	240.0	242.7	245.7	259.0	3.76
UV MAX	2423	11.58	2.0	10.0	11.0	13.0	19.0	2.77
PP	2807	2.61	0.0	0.0	0.0	2.8	27.8	4.40
UBIGEO	3288	120804	120804	120804	120804	120804	120804	0.0
FECHA CORTE	3288	—	2024-05-31	2024-05-31	2024-05-31	2024-05-31	2024-05-31	0.0

Nota. La tabla muestra las estadísticas descriptivas (conteo, media, mínimo, cuartiles, máximo y desviación estándar) de variables ambientales y meteorológicas medidas en la estación ubicada en el distrito correspondiente al UBIGEO 120804. Los valores de longitud, latitud y altitud permanecen constantes. Elaboración propia.

Figura 3.

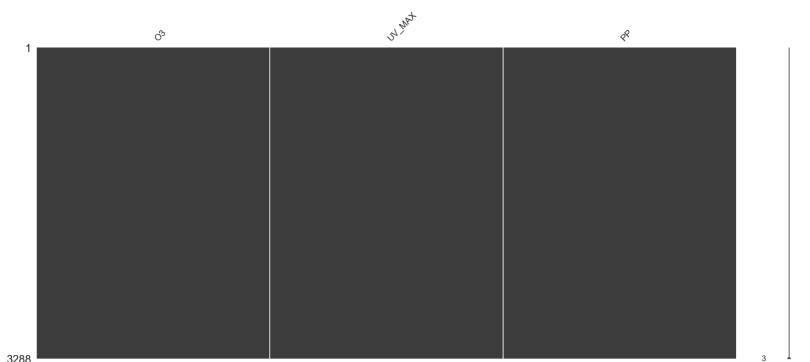
Distribución temporal de los valores faltantes en la serie de datos.



Nota. La figura muestra la presencia y distribución temporal de valores faltantes en las variables O3, UV_MAX y PP a lo largo del periodo de observación. Cada línea representa un registro; las líneas grises indican datos presentes y las líneas blancas indican datos ausentes. Elaboración propia.

Figura 4.

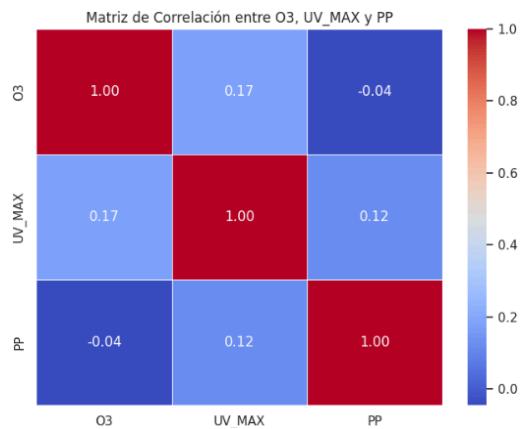
Distribución temporal de los valores faltantes luego de la imputación con KNN.



Nota. La figura muestra la distribución temporal de los valores faltantes en las variables O3, UV_MAX y PP después de aplicar la imputación mediante el algoritmo KNN con 3 vecinos. La ausencia de líneas blancas indica que los valores faltantes fueron completamente imputados. Elaboración propia.

Figura 5.

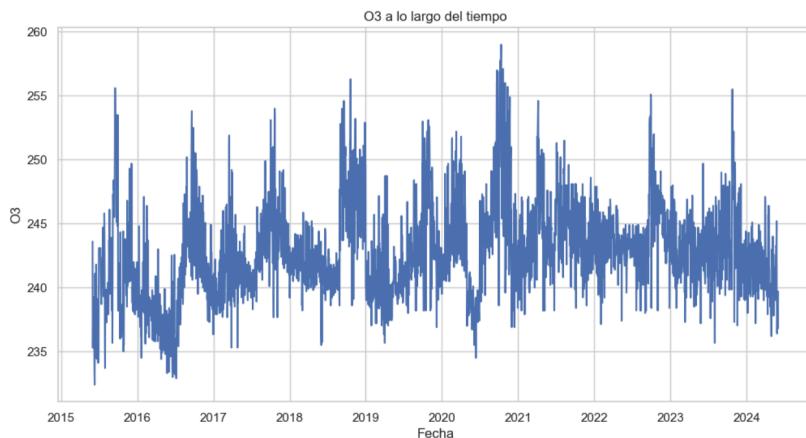
Matriz de correlación entre las variables O3, UV_MAX y PP.



Nota. La figura representa la correlación lineal entre las variables ambientales O3, UV_MAX y PP mediante un mapa de calor. Los valores anotados indican el coeficiente de correlación de Pearson. Elaboración propia.

Figura 6.

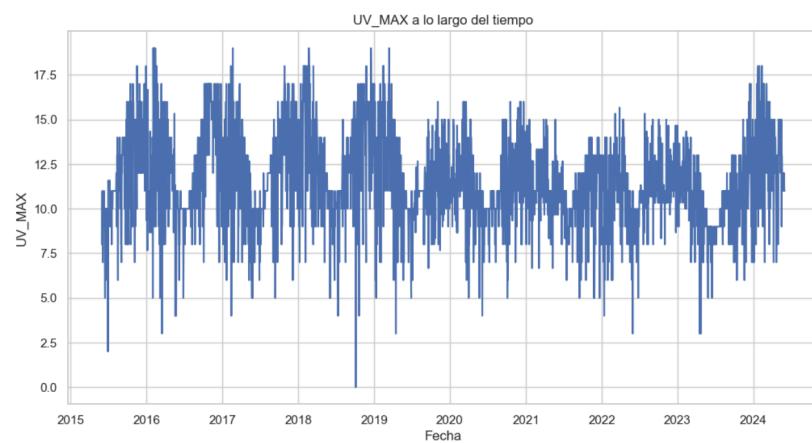
Evolución temporal de los valores de O3.



Nota. La figura muestra la variación del ozono (O3) a lo largo del tiempo según los registros de la estación. Elaboración propia.

Figura 7.

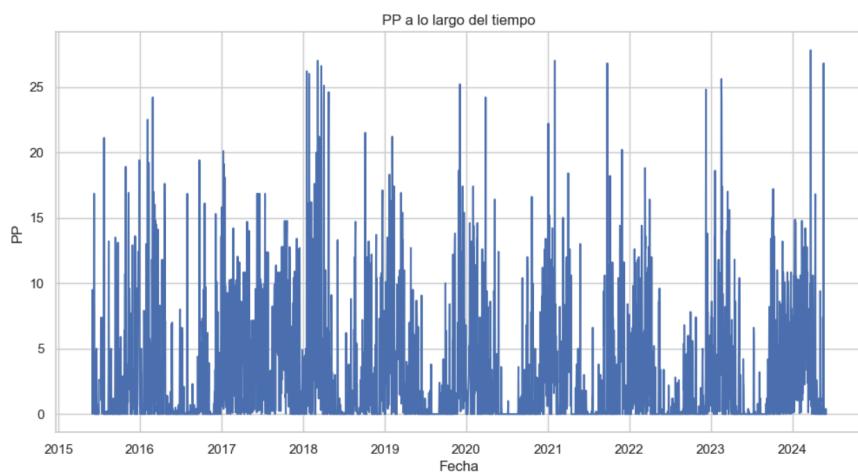
Evolución temporal del índice máximo de radiación UV (UV_MAX).



Nota. La figura representa los valores del índice máximo de radiación ultravioleta (UV_MAX) registrados a lo largo del periodo de análisis. Elaboración propia.

Figura 8.

Evolución temporal de la precipitación (PP).



Nota. La figura ilustra los registros de precipitación (PP) a lo largo del tiempo. Elaboración propia.

Tabla 5.

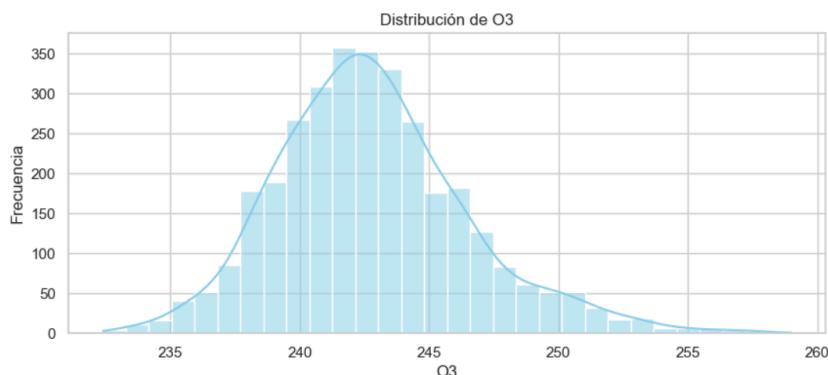
Estadísticas descriptivas de las variables O3, UV_MAX y PP.

	O3	UV_MAX	PP
Count	3288.000000	3288.000000	3288.000000
Mean	243.128859	11.531893	2.825315
Std	3.661525	2.551489	4.327758
Min	232.400000	0.000000	0.000000
25%	240.800000	10.000000	0.000000
50%	242.800000	11.000000	0.600000
75%	245.100000	13.000000	4.208333
Max	259.000000	19.000000	27.800000

Nota. La tabla presenta las estadísticas descriptivas de las variables O3, UV_MAX y PP. Elaboración propia.

Figura 9.

Distribución de la variable O3.

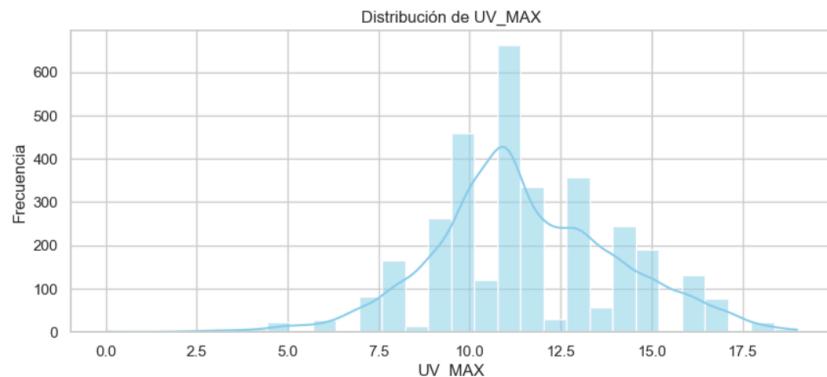


Nota. La figura muestra la distribución de frecuencias del ozono (O3) en la serie de datos. Elaboración propia.

La distribución del ozono presenta una forma aproximadamente simétrica, ligeramente sesgada a la izquierda, con la mayoría de valores concentrados entre 238 y 246 unidades. Esto sugiere una variabilidad moderada y una distribución cercana a la normal.

Figura 10.

Distribución de la variable UV_MAX.

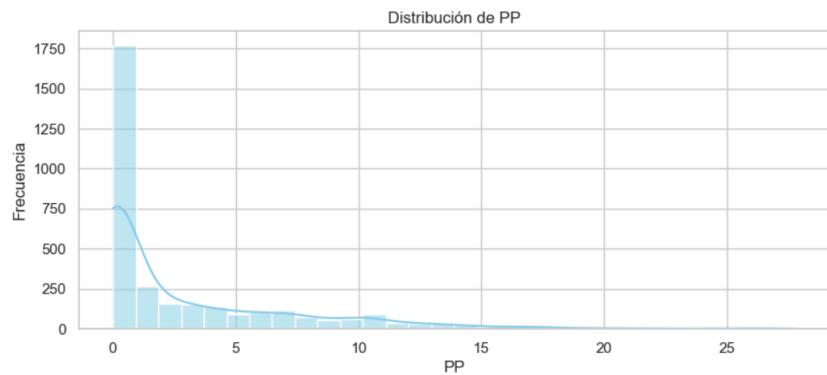


Nota. La figura representa la distribución del índice máximo de radiación ultravioleta (UV_MAX), permitiendo identificar su concentración y dispersión a lo largo del conjunto de datos. Elaboración propia.

Esta variable exhibe una distribución asimétrica con una ligera curtosis positiva, indicando la presencia de valores extremos hacia el lado derecho (índices UV más altos). La mayor frecuencia se concentra alrededor de valores de 10 a 13.

Figura 11.

Distribución de la variable PP.

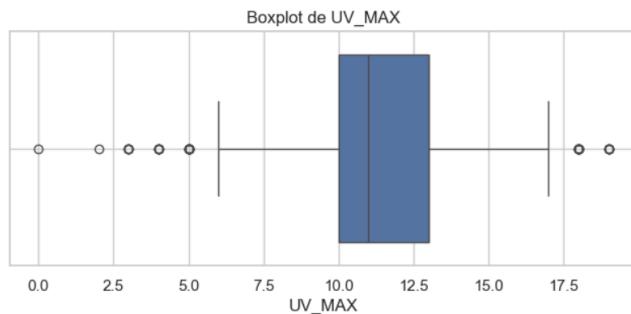


Nota. La figura ilustra la distribución de la precipitación (PP) registrada. Elaboración propia.

La distribución de la precipitación es altamente sesgada a la derecha (distribución exponencial), lo cual es característico de este tipo de variable. La mayoría de registros muestran valores cercanos a cero, indicando días sin lluvia o con muy baja precipitación, mientras que los valores altos son poco frecuentes pero posibles.

Figura 12.

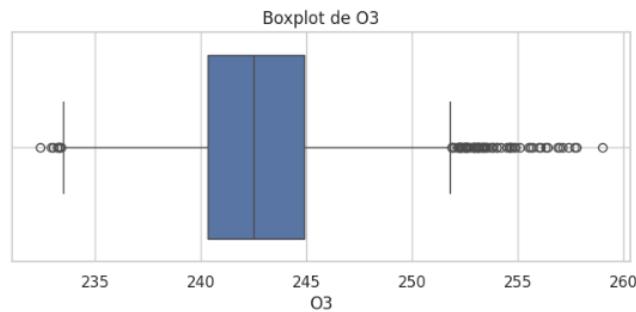
Boxplots de la variable UV_MAX para detección de valores atípicos.



Nota. Cada boxplot muestra la distribución de una variable numérica del conjunto de datos. Los puntos fuera de los bigotes representan posibles valores atípicos. Elaboración propia.

Figura 13.

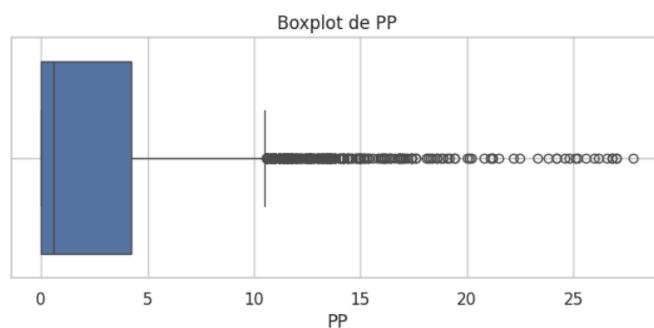
Boxplots de la variable O3 para detección de valores atípicos.



Nota. Cada boxplot muestra la distribución de una variable numérica del conjunto de datos. Los puntos fuera de los bigotes representan posibles valores atípicos. Elaboración propia.

Figura 14.

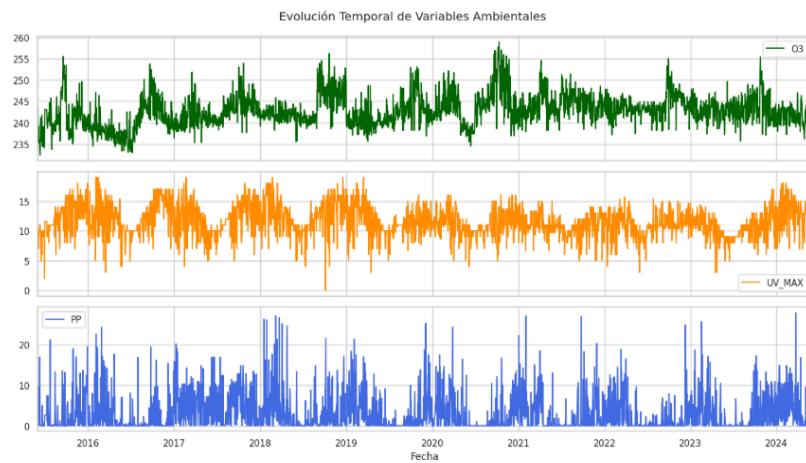
Boxplots de la variable PP para detección de valores atípicos.



Nota. Cada boxplot muestra la distribución de una variable numérica del conjunto de datos. Los puntos fuera de los bigotes representan posibles valores atípicos. Elaboración propia.

Figura 15.

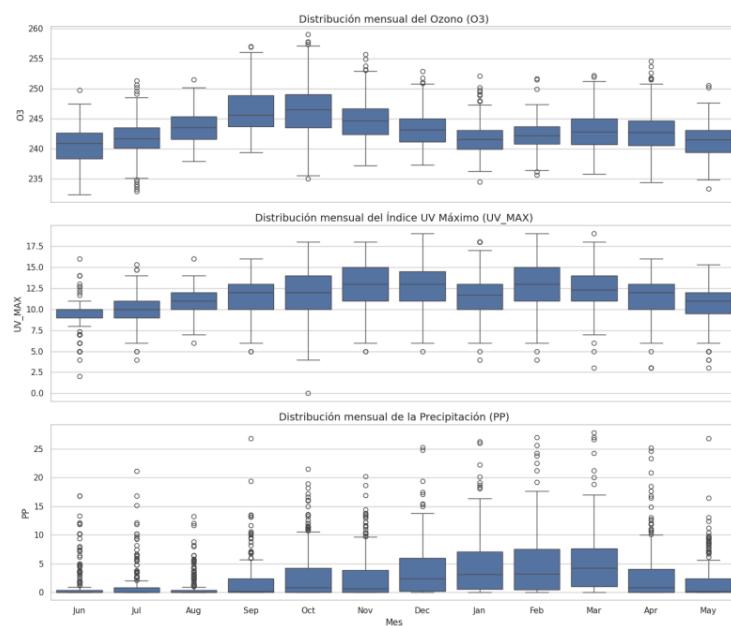
Valores diarios O_3 , UV_MAX y PP durante el periodo de estudio.



Nota. Muestra la variación diaria de las tres variables ambientales, permitiendo observar tendencias y estacionalidades. Elaboración propia.

Figura 16.

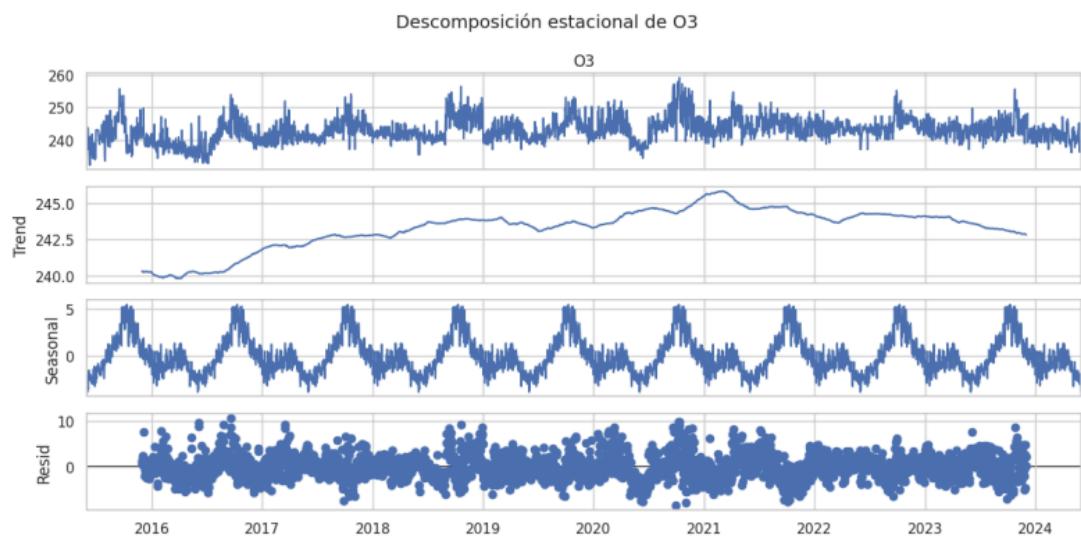
Distribución mensual de variables atmosféricas en la serie temporal analizada.



Nota. La figura muestra diagramas de caja para O3, UV_MAX y PP, desagregados por mes. Se utilizan abreviaturas en inglés para los nombres de los meses. La línea media en cada caja representa la mediana mensual, los límites de la caja indican los cuartiles y los puntos externos representan valores atípicos. Elaboración propia.

Figura 17.

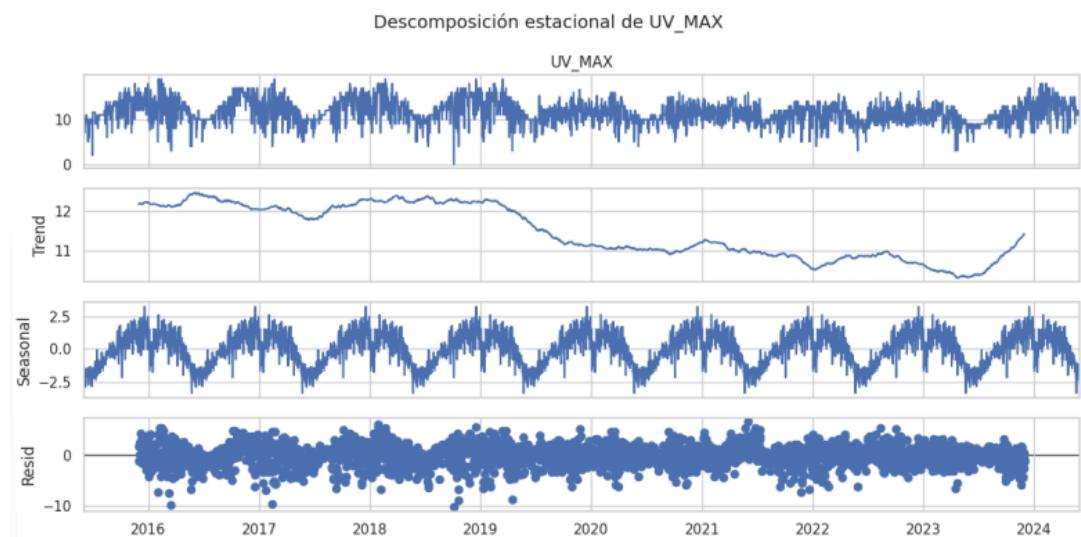
Descomposición estacional aditiva del Ozono (O3).



Nota. La figura presenta la descomposición temporal de la serie de ozono en sus componentes: tendencia, estacionalidad y residuo, utilizando un modelo aditivo con periodo anual (365 días). Elaboración propia.

Figura 18.

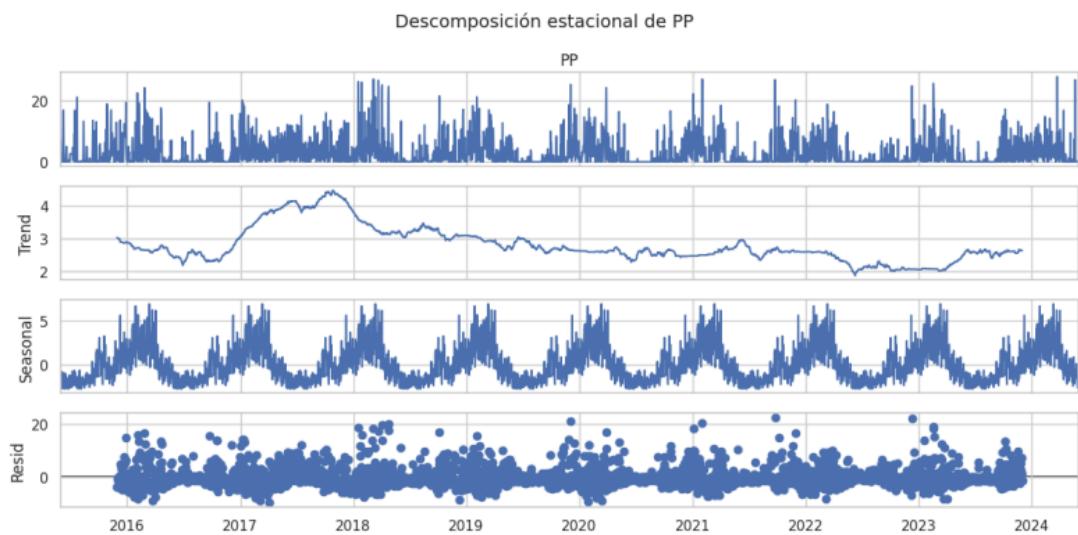
Descomposición estacional aditiva del Índice UV Máximo (UV_MAX).



Nota. La figura muestra la descomposición del índice UV máximo en sus componentes de tendencia, estacionalidad y residuo. Se emplea un modelo aditivo con un periodo de 365 días. Elaboración propia.

Figura 19.

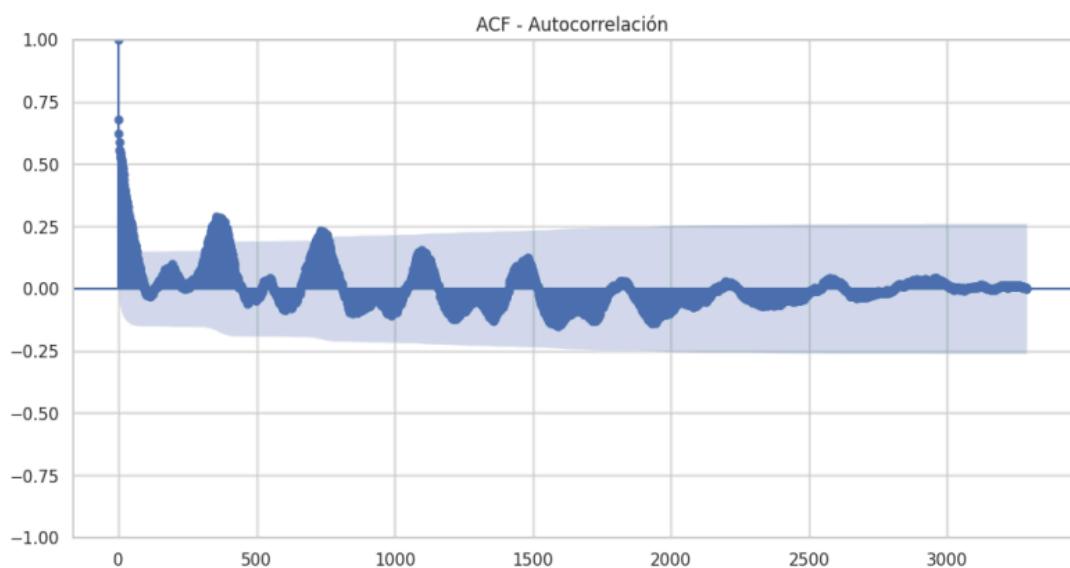
Descomposición estacional aditiva de la Precipitación (PP).



Nota. La figura descompone la serie de precipitación en componentes aditivos: tendencia, estacionalidad y residuo. Permite observar el comportamiento cíclico de la precipitación anual, bajo un modelo con periodicidad diaria. Elaboración propia.

Figura 20.

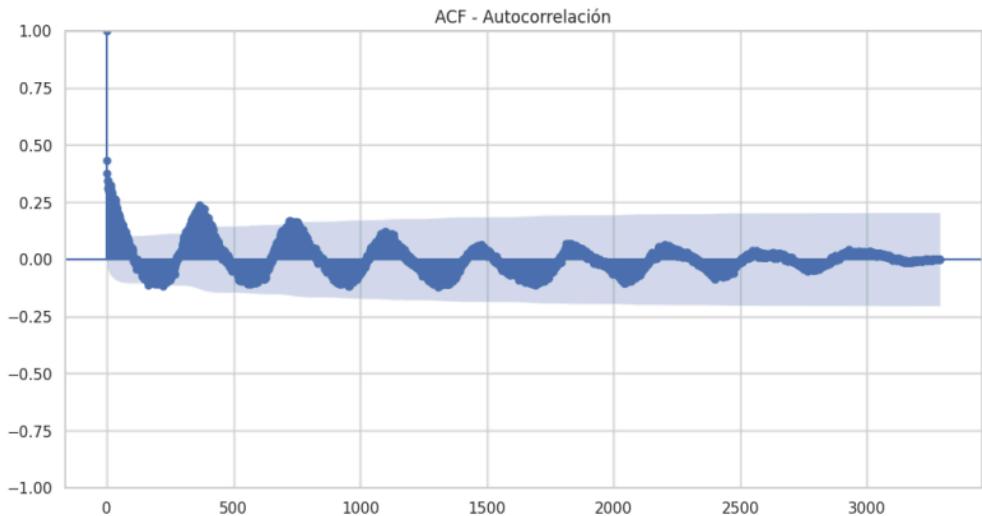
Función de Autocorrelación (ACF) del Ozono (O₃).



Nota. La gráfica muestra la autocorrelación de la concentración de ozono para 3.287 rezagos. Permite identificar patrones de dependencia temporal en la serie, indicando posible estacionalidad o persistencia. Elaboración propia.

Figura 21.

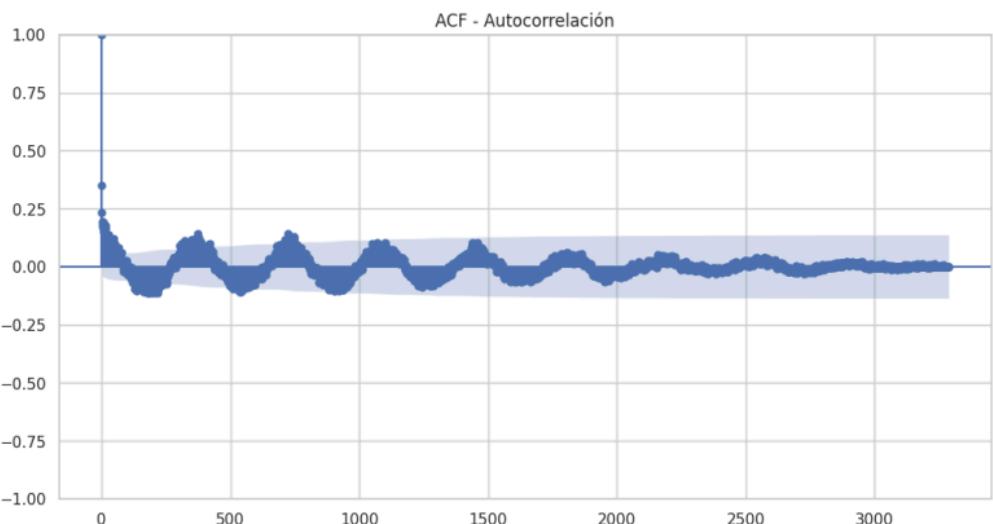
Función de Autocorrelación (ACF) del Índice UV Máximo (UV_MAX).



Nota. Se representa la autocorrelación del índice UV máximo a lo largo de 3.287 rezagos. El análisis revela la estructura temporal y posibles ciclos en la exposición máxima a radiación ultravioleta. Elaboración propia.

Figura 22.

Función de Autocorrelación (ACF) de la Precipitación (PP).



Nota. La gráfica detalla la autocorrelación de la serie de precipitaciones para 3.287 rezagos. Se utiliza para evaluar la dependencia temporal y la existencia de patrones recurrentes en la variable. Elaboración propia.

Tabla 6.

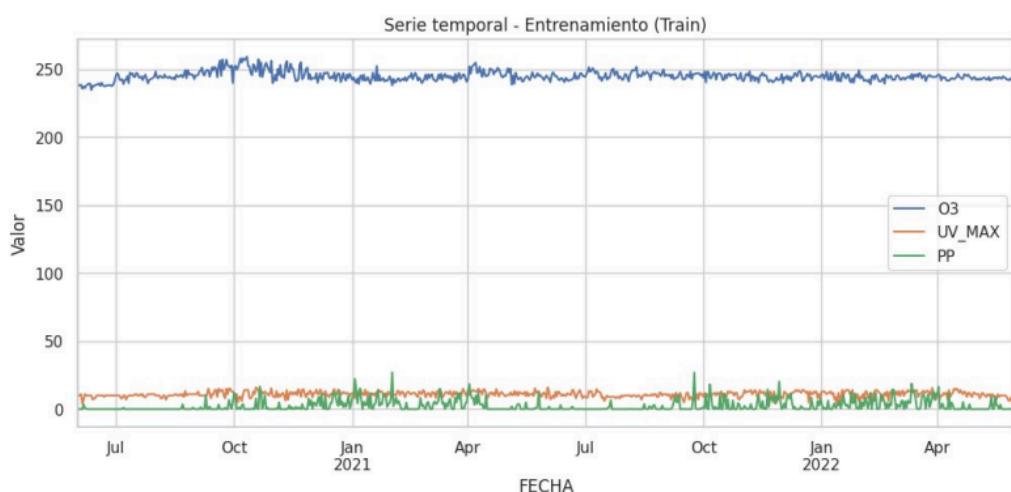
Vista tabular del conjunto de datos utilizado en el análisis.

FECHA	O3	UV_MAX	PP
2020-06-01	237.0	11.000000	0.0
2020-06-02	238.3	10.000000	0.0
2020-06-03	237.8	10.000000	0.0
2020-06-04	238.4	11.000000	0.0
2020-06-05	235.5	4.000000	0.0
...
2022-05-27	241.7	6.000000	0.0
2022-05-28	243.1	8.000000	0.2
2022-05-29	242.5	9.000000	0.0
2022-05-30	244.1	3.000000	0.4
2022-05-31	244.9	11.666667	0.0

Nota. La tabla muestra las observaciones crudas del conjunto de datos, con variables atmosféricas registradas en una serie temporal. Elaboración propia.

Figura 23.

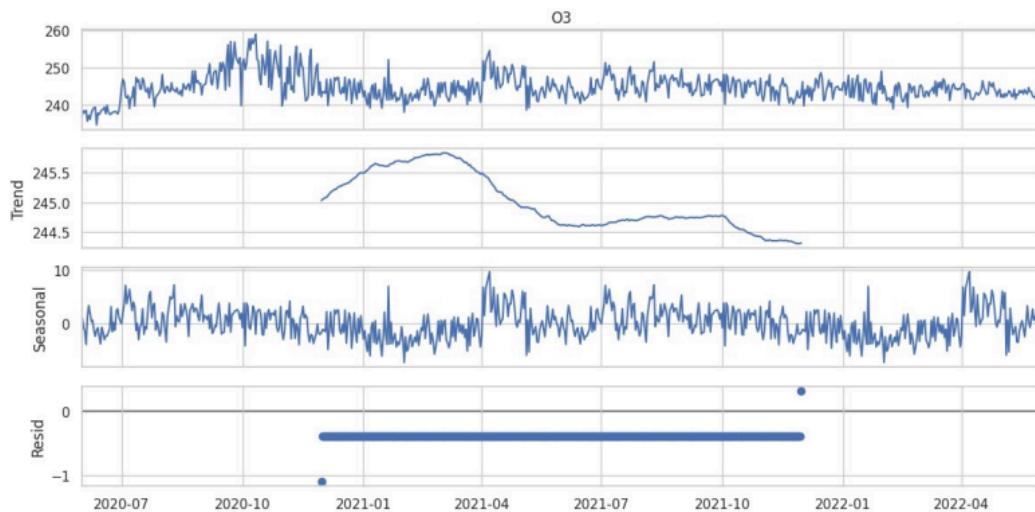
Serie temporal del conjunto de entrenamiento (Train).



Nota. La figura representa la evolución temporal de la variable objetivo durante el periodo asignado al conjunto de entrenamiento. Elaboración propia.

Figura 24.

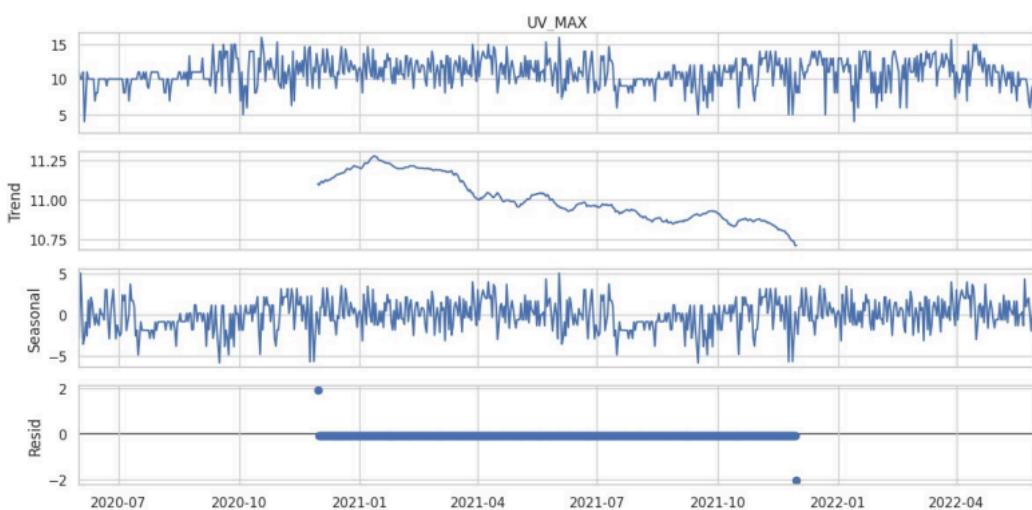
Descomposición aditiva de la serie temporal de Ozono (O3).



Nota. La figura presenta la descomposición de la serie temporal de O3 en sus componentes: tendencia, estacionalidad y residuo. Se utilizó un modelo aditivo con una periodicidad de 365 días. Elaboración propia.

Figura 25.

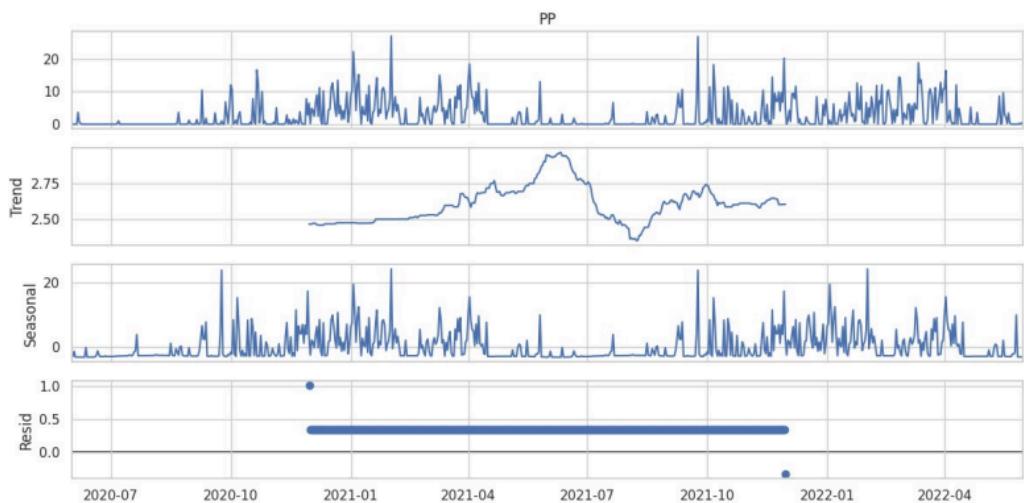
Descomposición aditiva del componente UV_MAX en la serie de entrenamiento.



Nota. La figura presenta la descomposición aditiva de la variable UV_MAX en tendencia, estacionalidad y residuo, utilizando una periodicidad anual (365 días). Elaboración propia.

Figura 26.

Descomposición aditiva de la serie temporal de precipitación (PP).



Nota. La figura presenta la descomposición aditiva de la precipitación diaria, separando la serie original en sus componentes de tendencia, estacionalidad y residuo. Se asume una periodicidad anual de 365 días. Elaboración propia.

Tabla 7.

Resultados de la prueba ADF para la serie de Ozono (O3).

ADF Statistic	-3.1913479104150104
p-value	0.02049099030216959

Nota. La estadística ADF es -3.191 y el valor p es 0.0205, menor que el umbral de significancia del 5%, lo que indica que la serie de tiempo del Ozono es estacionaria. Elaboración propia.

Tabla 8.

Resultados de la prueba ADF para la serie del Índice UV Máximo (UV_MAX).

ADF Statistic	-8.970580501299946
p-value	7.773685307283148e-15

Nota. La estadística ADF es -8.971 y el valor p es prácticamente cero, lo que evidencia una fuerte evidencia de estacionariedad en la serie del Índice UV Máximo. Elaboración propia.

Tabla 9.

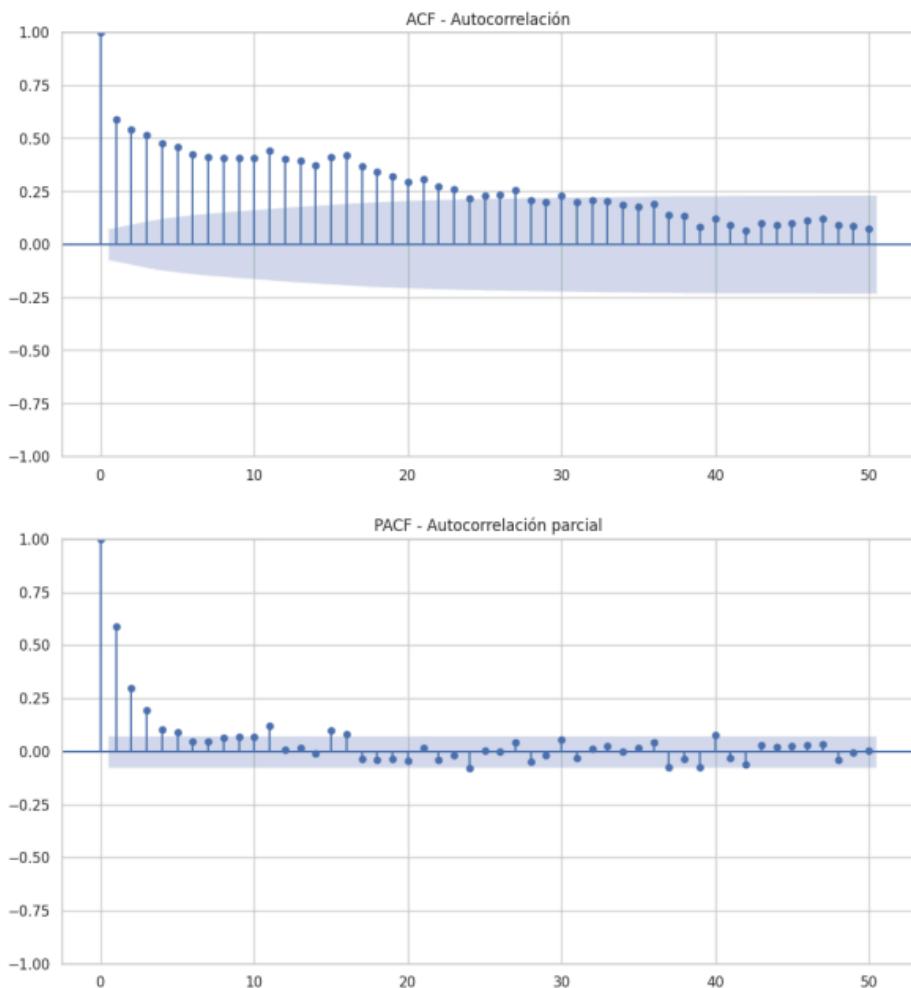
Resultados de la prueba ADF para la serie de Precipitación (PP).

ADF Statistic	-4.137596373530011
p-value	0.0008388470579614095

Nota. La estadística ADF es -4.138 con un valor p de 0.0008, menor al umbral del 5 %, por lo que se concluye que la serie de Precipitación es estacionaria. Elaboración propia.

Figura 27.

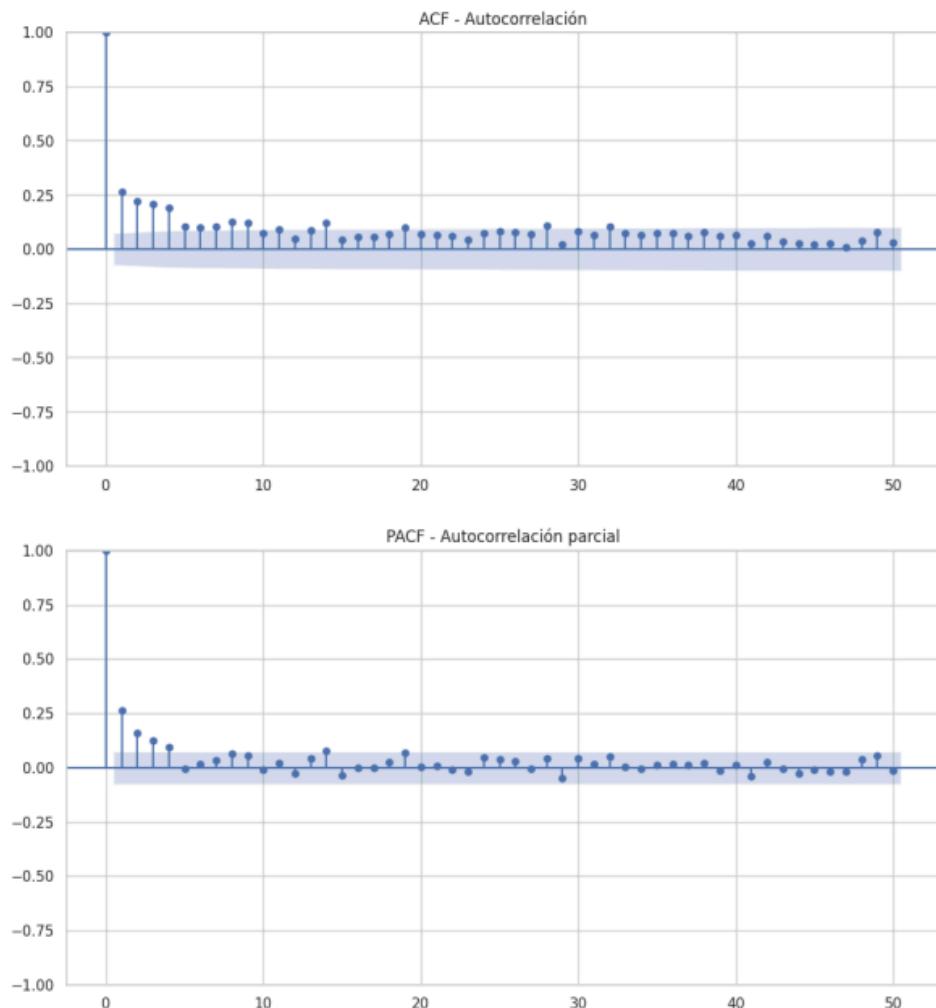
Funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para la serie de Ozono (O3).



Nota. Los gráficos muestran la estructura de dependencia temporal de la serie de Ozono hasta 50 rezagos. La ACF permite identificar la presencia de correlaciones globales, mientras que la PACF revela relaciones directas con rezagos específicos. Elaboración propia.

Figura 28.

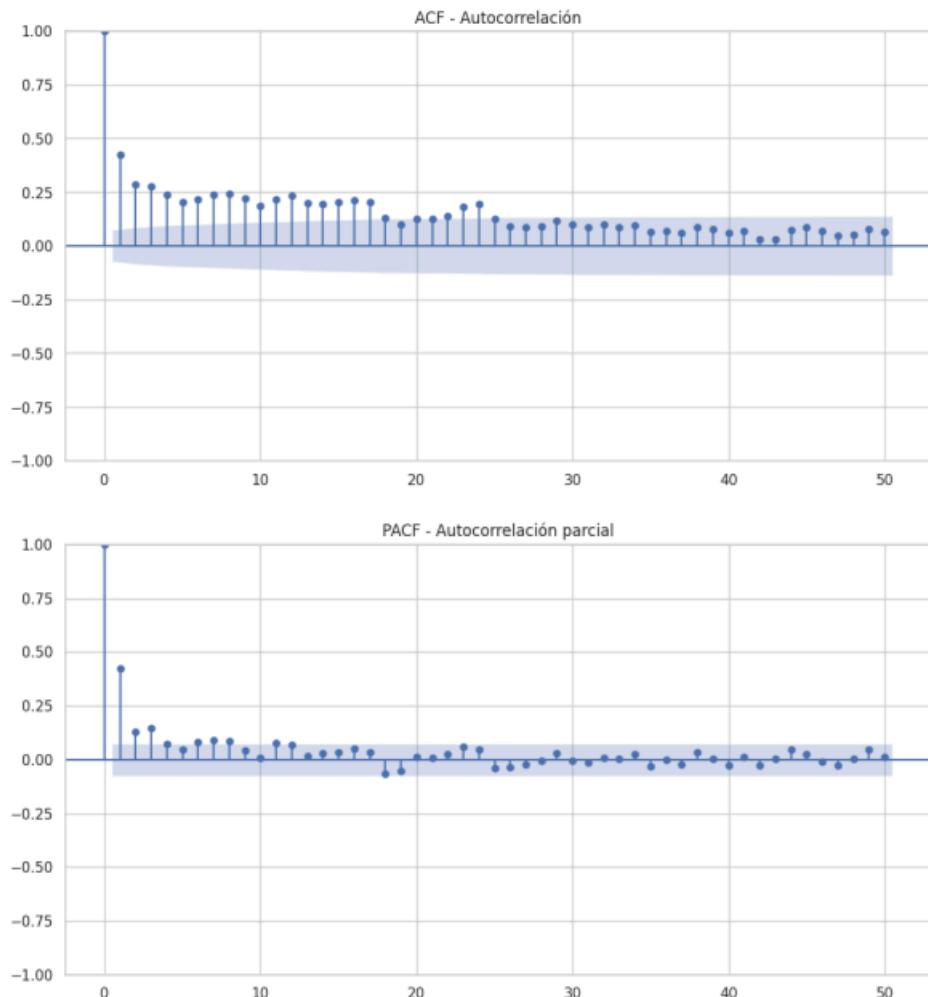
Funciones de autocorrelación y autocorrelación parcial para la serie UV_MAX.



Nota. La gráfica ACF muestra la dependencia serial a distintos rezagos, revelando correlaciones significativas principalmente en los primeros lags. La gráfica PACF permite identificar el número de rezagos relevantes para un modelo autorregresivo. Elaboración propia.

Figura 29.

Función de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) para la serie de Precipitación (PP).



Nota. Los gráficos ACF y PACF muestran la dependencia temporal de la serie de Precipitación hasta 50 rezagos. La ACF permite identificar la presencia de correlación global en distintos rezagos, mientras que la PACF muestra la correlación condicional. Elaboración propia.

4.2 Modelado Predictivo:

En esta parte se detalla la implementación de los modelos de series temporales ARIMA y SARIMA aplicados a las variables seleccionadas. Se documenta el proceso de identificación, estimación y validación de los modelos, incluyendo los criterios de selección como AIC, BIC y pruebas de diagnóstico de residuos. Además, se presentan las métricas de evaluación como el error cuadrático medio (RMSE) y el error absoluto medio (MAE), que permiten verificar la precisión de las predicciones. Finalmente, se incluyen visualizaciones de las proyecciones generadas, las cuales sustentan la viabilidad de los modelos para aplicaciones futuras en la vigilancia atmosférica.

Tabla 10.

SARIMAX Results de la variable O3.

Variable dependiente	O3
Número de observaciones	730
Modelo	SARIMAX(1, 0, 1) x (1, 1, 1, 91)
Log-verosimilitud	-1404.400
AIC	2818.800
BIC	2840.314
HQIC	2827.210
Muestra	0 - 730
Tipo de covarianza	opg

Nota. Modelo SARIMAX aplicado a la serie temporal de ozono (O3) con 730 observaciones. Elaboración propia.

La especificación incluye componentes autorregresivos y de medias móviles estacionales con periodicidad de 91 días. Los criterios de información (AIC, BIC, HQIC) permiten comparar la calidad del ajuste. La covarianza se estimó mediante el método de gradiente outer-product (opg). Elaboración propia.

Tabla 11.

Parámetros estimados del modelo SARIMAX($1, 0, 1$) \times ($1, 1, 1, 91$) para la variable O_3 .

	coef	std err	z	P> z 	[0.025	0.975]
ar.L1	0.9535	0.018	54.015	0.000	0.919	0.988
ma.L1	-0.7663	0.036	-21.431	0.000	-0.836	-0.696
ar.S.L91	-0.3919	0.043	-9.184	0.000	-0.475	-0.308
ma.S.L91	-0.3408	0.060	-5.639	0.000	-0.459	-0.222
sigma2	9.8343	0.551	17.859	0.000	8.755	10.914

Nota. Se presentan los coeficientes estimados (coef), errores estándar (std err), valores z, niveles de significancia (P>|z|) y los intervalos de confianza al 95 % para cada parámetro del modelo. Todos los parámetros resultaron estadísticamente significativos ($p < .001$). Elaboración propia.

Tabla 12.

Pruebas de diagnóstico del modelo SARIMAX para la variable O_3 .

Diagnostics	Value
Ljung-Box (L1) (Q)	0.00
Prob(Q)	0.95
Jarque-Bera (JB)	6.23
Prob(JB)	0.04
Heteroskedasticity (H)	0.42
Prob(H) (two-sided)	0.00
Skew	-0.11
Kurtosis	3.48

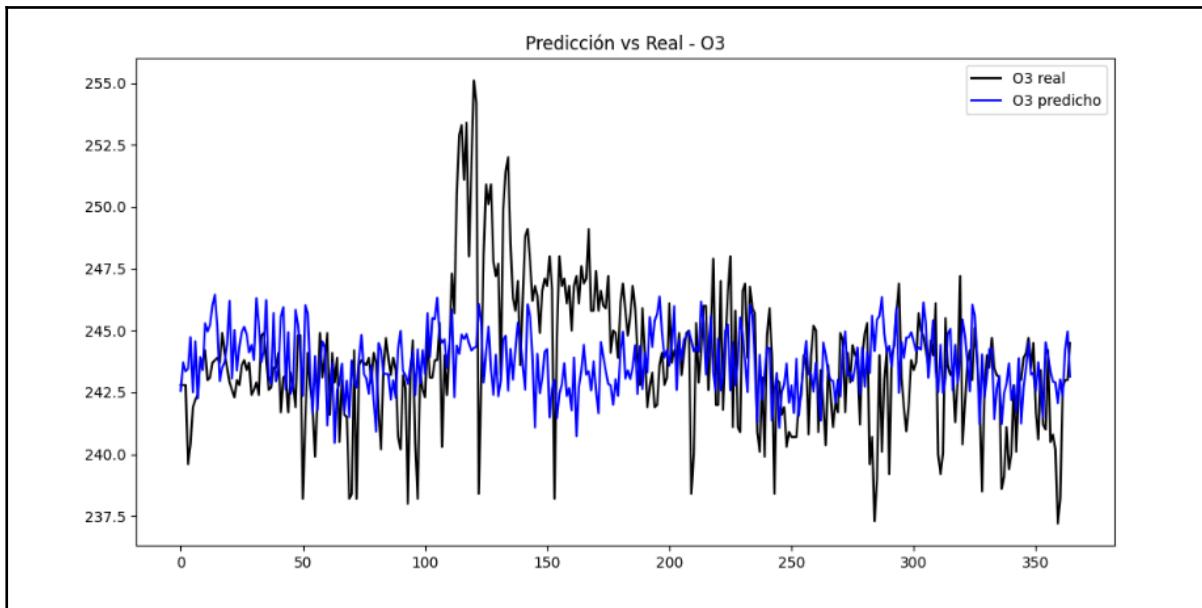
Nota. Se reportan estadísticas de diagnóstico sobre los residuos del modelo. Elaboración propia.

La prueba de Ljung-Box indica ausencia de autocorrelación ($p = .95$). La prueba de Jarque-Bera sugiere desviación de la normalidad ($p = .04$). La prueba de heterocedasticidad (H) revela

evidencia de varianza no constante ($p < .001$). Los valores de asimetría (Skew) y curtosis (Kurtosis) describen la forma de la distribución de los residuos.

Figura 30.

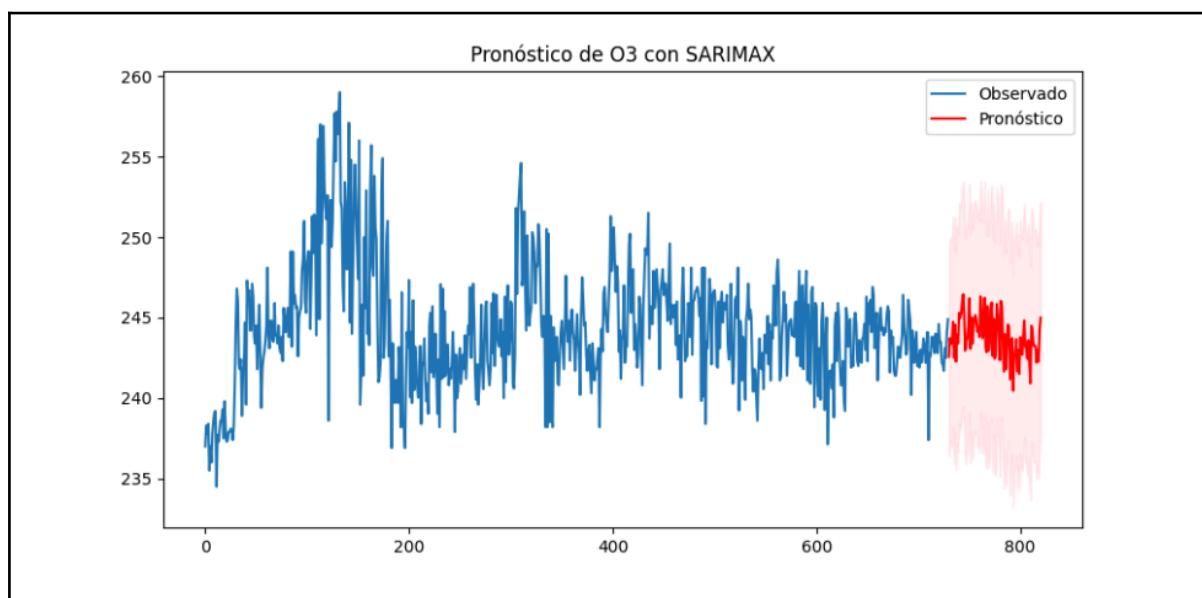
Serie temporal observada y pronosticada de O₃.



Nota. Comparación entre los valores reales de O₃ (línea negra) y los valores pronosticados por el modelo SARIMAX (línea azul) en el conjunto de prueba. Elaboración propia.

Figura 31.

Pronóstico de O₃ con intervalo de confianza utilizando SARIMAX.



Nota. Se muestra el pronóstico de 91 pasos para la concentración de O₃ (línea roja) a partir de los datos observados (línea azul) mediante el modelo SARIMAX. El área sombreada representa el intervalo de confianza al 95 %, indicando la incertidumbre asociada a la predicción. Elaboración propia.

Tabla 13.

Métricas de evaluación del modelo SARIMAX para la variable O₃.

	Valor
MAE	2.27
RMSE	2.98
R²	-0.11
MAE Relativo	0.93%

Nota. Se presentan las métricas de desempeño del modelo en el conjunto de prueba. Elaboración propia.

Tabla 14.

Resultados SARIMAX de la variable UV_MAX.

Variable dependiente	UV_MAX
Número de observaciones	730
Modelo	SARIMAX(1, 0, 1) x (1, 1, 1, 91)
Log-verosimilitud	-1204.101
AIC	2418.202
BIC	2439.715
HQIC	2426.612
Muestra	0 - 730
Tipo de covarianza	opg

Nota. Modelo SARIMAX aplicado a la serie temporal de radiación ultravioleta máxima (UV_MAX) con 730 observaciones. Elaboración propia.

La especificación incluye componentes autorregresivos y de medias móviles estacionales con periodicidad de 91 días. Los criterios de información (AIC, BIC, HQIC) permiten evaluar la calidad del ajuste. La matriz de covarianza fue estimada mediante el método de gradiente outer-product (opg).

Tabla 15.

Parámetros estimados del modelo SARIMAX(1, 0, 1) × (1, 1, 1, 91) para la variable UV_MAX.

	coef	std err	z	P> z 	[0.025	0.975]
ar.L1	0.8908	0.039	22.956	0.000	0.815	0.967
ma.L1	-0.6935	0.058	-11.864	0.000	-0.808	-0.579
ar.S.L91	-0.2682	0.051	-5.278	0.000	-0.368	-0.169
ma.S.L91	-0.6207	0.059	-10.551	0.000	-0.736	-0.505
sigma2	4.4503	0.267	16.641	0.000	3.926	4.974

Nota. Se reportan los coeficientes estimados (coef), errores estándar (std err), valores z, niveles de significancia (P>|z|) e intervalos de confianza al 95 % para cada parámetro del modelo. Todos los parámetros son estadísticamente significativos ($p < .001$). Elaboración propia.

Tabla 16.

Pruebas de diagnóstico del modelo SARIMAX para la variable UV_MAX.

Diagnostics	Value
Ljung-Box (L1) (Q)	0.45
Prob(Q)	0.50
Jarque-Bera (JB)	4.47
Prob(JB)	0.11
Heteroskedasticity (H)	1.22
Prob(H) (two-sided)	0.17
Skew	-0.16

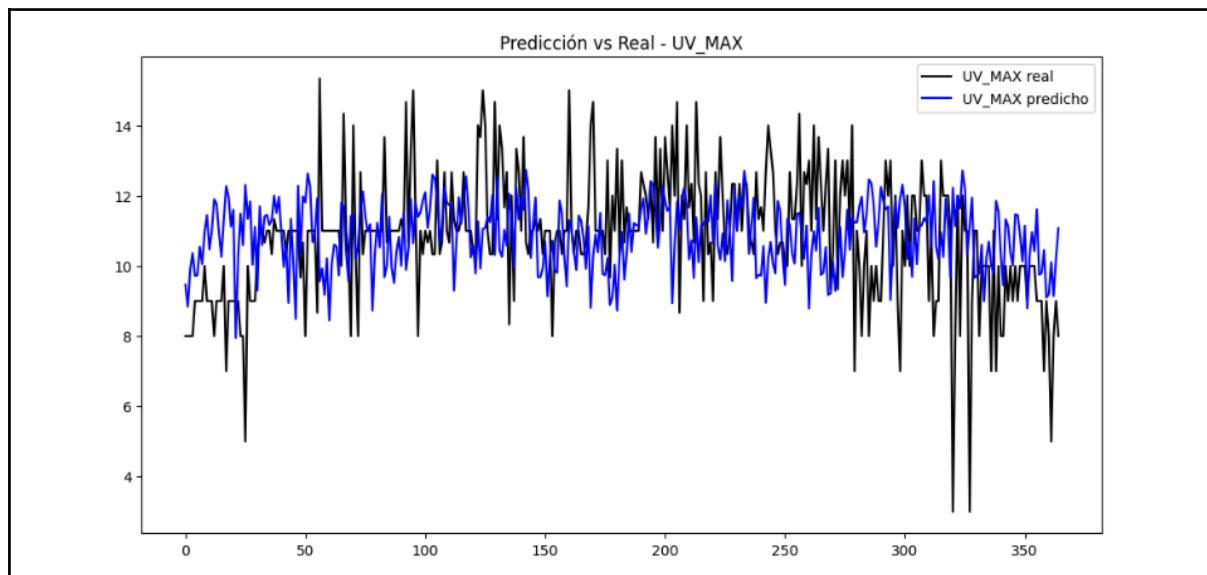
Kurtosis	3.32
-----------------	------

Nota. Se presentan estadísticas de diagnóstico aplicadas a los residuos del modelo. Elaboración propia.

La prueba de Ljung-Box indica ausencia de autocorrelación ($p = .50$). La prueba de Jarque-Bera no rechaza la normalidad de los residuos ($p = .11$). La prueba de heterocedasticidad (H) sugiere varianza constante ($p = .17$). Los valores de asimetría (Skew) y curtosis (Kurtosis) describen una distribución aproximadamente simétrica y mesocúrtica.

Figura 32.

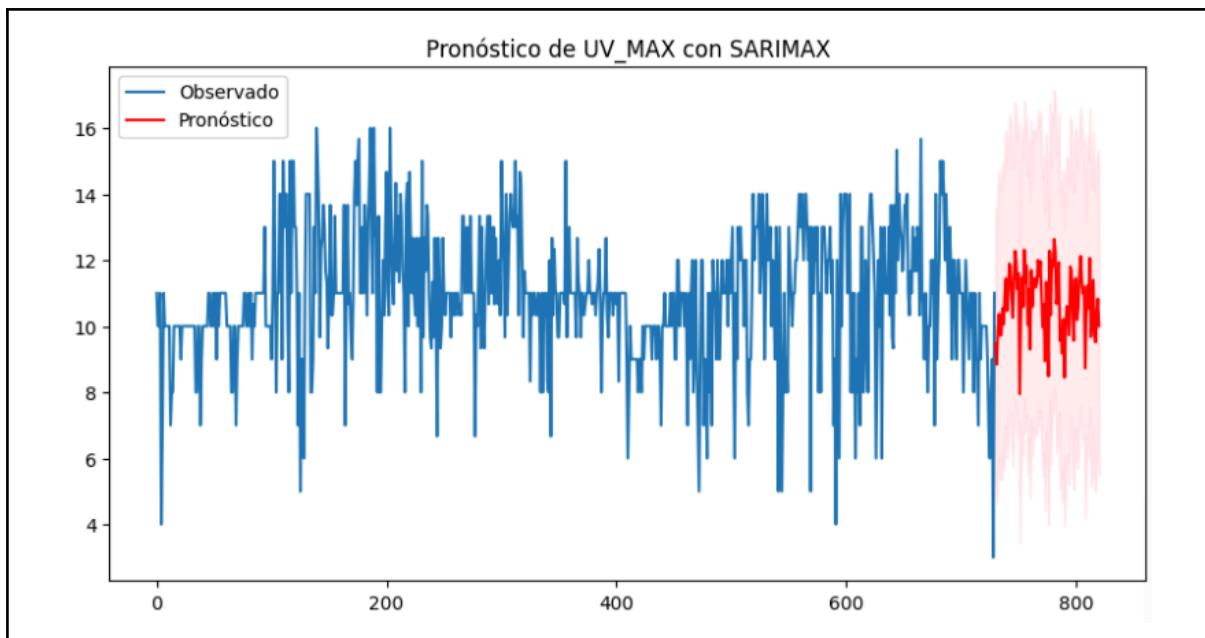
Serie temporal observada y pronosticada de UV_MAX.



Nota. Comparación entre los valores reales de UV_MAX (línea negra) y los valores pronosticados por el modelo SARIMAX (línea azul) en el conjunto de prueba. Elaboración propia.

Figura 33.

Pronóstico de UV_MAX con intervalo de confianza utilizando SARIMAX.



Nota. Se muestra el pronóstico de 91 pasos para la radiación ultravioleta máxima (UV_MAX) generado mediante el modelo SARIMAX. La línea roja representa los valores pronosticados, la línea azul los valores observados, y el área sombreada indica el intervalo de confianza al 95 %, lo cual refleja la incertidumbre asociada a la predicción. Elaboración propia.

Tabla 17.

Métricas de evaluación del modelo SARIMAX para la variable UV_MAX.

	Valor
MAE	1.55
RMSE	2.06
R ²	-0.30
MAE Relativo	14.33%

Nota. Se presentan las métricas de desempeño del modelo en el conjunto de prueba. Elaboración propia.

Tabla 18.*SARIMAX Results de la variable PP.*

Dep. Variable	PP
No. Observations	730
Model	SARIMAX(1, 0, 1)x(1, 1, 1, 91)
Log Likelihood	-1584.506
AIC	3179.012
BIC	3200.525
HQIC	3187.421
Sample	0 - 730
Covariance Type	opg

Nota. Modelo SARIMAX aplicado a la serie temporal de precipitación (PP) con 730 observaciones. Elaboración propia.

La especificación incluye componentes autorregresivos y de medias móviles estacionales con periodicidad de 91 días. Los criterios de información (AIC, BIC, HQIC) permiten evaluar la calidad del ajuste. La matriz de covarianza fue estimada mediante el método de gradiente outer-product (opg).

Tabla 19.*Parámetros estimados del modelo SARIMAX(1, 0, 1) × (1, 1, 1, 91) para la variable PP.*

	coef	std err	z	P> z 	 0.025	0.975
ar.L1	0.9281	0.026	35.022	0.000	0.876	0.980
ma.L1	-0.7159	0.049	-14.734	0.000	-0.811	-0.621
ar.S.L91	-0.3030	0.051	-5.969	0.000	-0.402	-0.203
ma.S.L91	-0.6586	0.061	-10.829	0.000	-0.778	-0.539
sigma2	17.6916	0.724	24.436	0.000	16.273	19.111

Nota. Se presentan los coeficientes estimados (coef), errores estándar (std err), valores z, niveles de significancia (P>|z|) e intervalos de confianza al 95 % para cada parámetro del modelo. Todos los coeficientes son estadísticamente significativos ($p < .001$), lo que indica su contribución al ajuste del modelo de precipitación (PP). Elaboración propia.

Tabla 20.

Pruebas de diagnóstico del modelo SARIMAX para la variable PP.

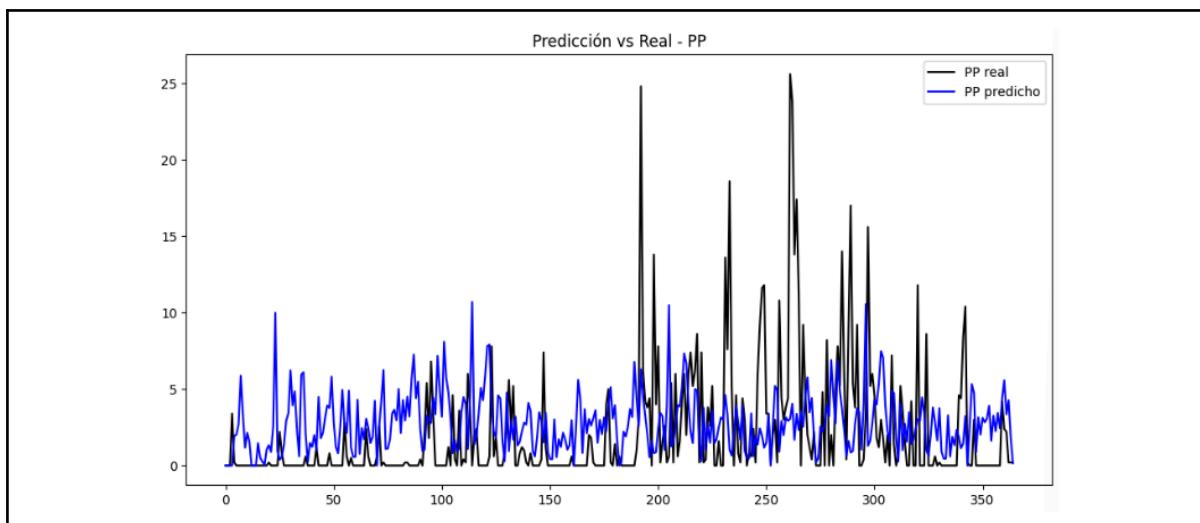
Diagnostics	Value
Ljung-Box (L1) (Q)	6.38
Prob(Q)	0.01
Jarque-Bera (JB)	309.82
Prob(JB)	0.00
Heteroskedasticity (H)	1.02
Prob(H) (two-sided)	0.91
Skew	1.02
Kurtosis	6.08

Nota. Se reportan estadísticas aplicadas a los residuos del modelo. Elaboración propia.

La prueba de Ljung-Box detecta autocorrelación significativa ($p = .01$). La prueba de Jarque-Bera indica que los residuos no siguen una distribución normal ($p < .001$), con sesgo positivo (Skew = 1.02) y curtosis elevada (Kurtosis = 6.08), lo que sugiere colas pesadas. La prueba de heterocedasticidad no muestra evidencia de varianza no constante ($p = .91$).

Figura 34.

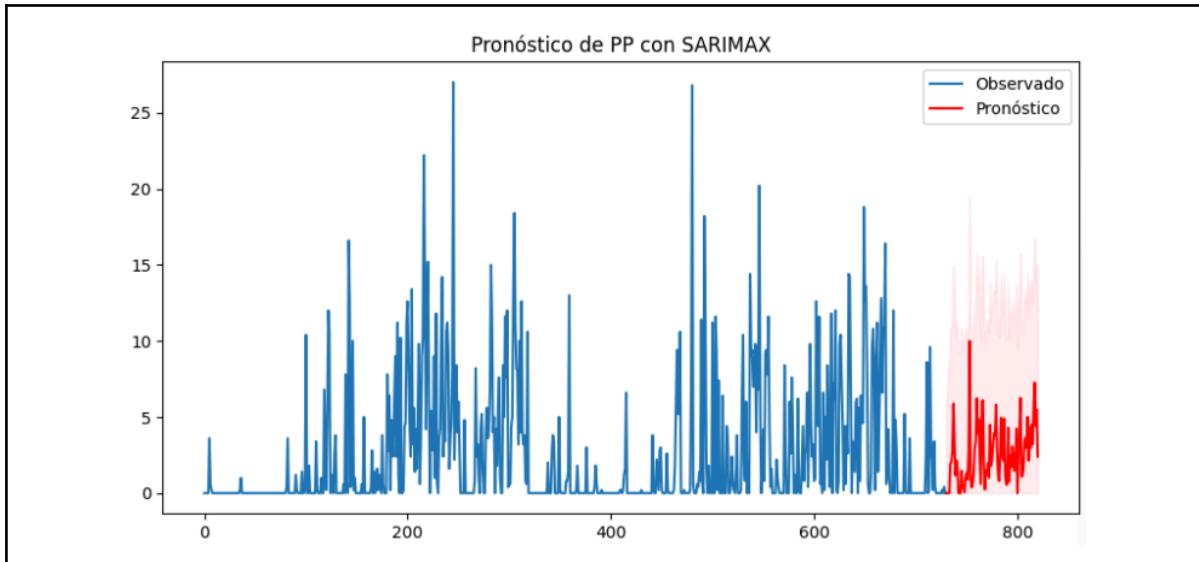
Serie temporal observada y pronosticada de PP.



Nota. Comparación entre los valores reales de precipitación (PP) en el conjunto de prueba (línea negra) y los valores pronosticados por el modelo SARIMAX (línea azul). La visualización permite evaluar el grado de ajuste del modelo en la estimación de valores futuros de precipitación.
Elaboración propia.

Figura 35.

Pronóstico de PP con intervalo de confianza utilizando SARIMAX.



Nota. Se muestra el pronóstico de 91 pasos para la precipitación (PP) generado mediante el modelo SARIMAX. La línea roja representa los valores pronosticados, ajustados para no ser negativos. La línea azul muestra los valores observados. El área sombreada indica el intervalo de confianza al 95 %, reflejando la incertidumbre en la estimación.

Tabla 21.

Métricas de evaluación del modelo SARIMAX para la variable PP.

	Valor
MAE	3.08
RMSE	4.33
R ²	-0.25
MAE Relativo	148.39%

Nota. Se presentan las métricas de desempeño del modelo en el conjunto de prueba.
Elaboración propia.

CAPÍTULO V: EVALUACIÓN ECONÓMICA-FINANCIERA

Este proyecto se basa principalmente en el análisis de fuentes secundarias y no requiere inversiones en infraestructura o adquisición de equipos costosos. No obstante, se identifican ciertos costos fijos y variables necesarios para su correcta ejecución, principalmente relacionados con servicios, herramientas digitales, alimentación y capacitación.

Tabla 22.

Estimación de costos asociados al desarrollo del proyecto de tesis.

Objetivo	Estimar los recursos económicos asociados al desarrollo de la tesis titulada "Análisis y predicción de variables atmosféricas mediante series temporales en el Observatorio de Vigilancia Atmosférica Marcapomacocha (OVA)", considerando los costos reales asumidos y los recursos utilizados durante el proceso de investigación.
Alcance: Duración del proyecto, componentes	Duración del proyecto: 5 meses (de marzo a agosto de 2025). Actividades principales: - Recolección y limpieza de datos atmosféricos (2015–2024). - Análisis exploratorio de datos (EDA). - Entrenamiento y evaluación de modelos de series temporales. - Redacción y validación de resultados.
Estructura de costos	Por categorías: Desarrollo (RRHH estimado). Infraestructura y servicios cloud. Software y licencias. Datos. Validación y documentación. Otros costos operativos. Total estimado.

Nota. La tabla presenta una desagregación de los costos estimados para el desarrollo de la tesis, clasificados por categoría de gasto. Todos los valores están basados en estimaciones reales incurridas durante el proceso. Elaboración propia.

Tabla 23.

Estimación de costos por categoría en el desarrollo del proyecto.

Categoría	Monto (S/)
Desarrollo (valoración del trabajo)	S/. 2.000,00
Infraestructura y servicios cloud	S/. 3.000,00
Software y licencias	S/. 80,00
Datos	S/. 0,00
Validación y documentación	S/. 0,00
Otros costos operativos	S/. 1.854,00
Total estimado	S/. 6.934,00

Nota. Se detallan los costos estimados en soles (S/) para cada categoría relevante en el desarrollo del proyecto. Valores reportados en función de la planificación presupuestaria. Elaboración propia.

La evaluación económico-financiera del presente proyecto revela que, debido a su enfoque en el análisis de datos secundarios y al uso de herramientas tecnológicas de libre acceso o bajo costo, los requerimientos económicos son moderados. Los costos fijos se limitan principalmente a servicios esenciales como el acceso a internet, suscripciones a plataformas digitales de apoyo y un curso de capacitación especializado que fortalecerá las competencias técnicas necesarias para el análisis. Por otro lado, los costos variables se asocian a necesidades básicas como alimentación y la impresión final del documento.

CONCLUSIONES

Esta investigación logró analizar y predecir las tendencias temporales de variables atmosféricas clave: ozono (O_3), índice UV (UV_MAX) y precipitación (PP). Registradas en el Observatorio de Vigilancia Atmosférica Marcapomacocha, mediante el uso de técnicas de series temporales, con énfasis en modelos predictivos SARIMAX.

El procedimiento metodológico siguió una secuencia clara: desde la recopilación y validación de los datos proporcionados por SENAMHI, pasando por un análisis exploratorio detallado que permitió identificar patrones estacionales y anomalías, hasta la implementación y evaluación de modelos estadísticos avanzados. Este enfoque riguroso permitió no solo describir las características de las series temporales, sino también generar predicciones con un grado aceptable de precisión para cada variable.

En el caso del ozono, el modelo SARIMAX mostró un excelente desempeño, con un MAE absoluto de apenas 2.27 DU y un MAE relativo de 0.93%, lo que indica una alta capacidad del modelo para capturar el comportamiento diario de esta variable, a pesar de un R^2 negativo (-0.11) que se explicó adecuadamente por la baja variabilidad natural de la serie. Para el índice UV, el modelo también fue capaz de generar predicciones útiles, con un MAE de 1.55 y un MAE relativo de 14.33%, reflejando un rendimiento razonable en una variable sujeta a variaciones más abruptas. Finalmente, aunque el modelo para la precipitación presentó un desempeño más modesto (MAE: 3.08, MAE relativo: 148.39%), esto es atribuible a la complejidad intrínseca de esta variable, que combina valores nulos frecuentes con picos intensos difíciles de predecir con modelos lineales.

Con base en estos resultados, se concluye que se cumplieron tanto el objetivo general como los objetivos específicos de la investigación. Se logró explorar y caracterizar las series temporales de las variables atmosféricas, identificar sus patrones principales y aplicar modelos SARIMAX que, si bien con distintos niveles de precisión, ofrecieron predicciones útiles y coherentes para los fines propuestos.

En términos generales, esta investigación no solo proporciona un marco estadístico robusto para la vigilancia atmosférica en zonas de alta montaña como Marcapomacocha, sino que también representa una base sólida para futuras investigaciones que integren enfoques híbridos o de aprendizaje automático, con miras a mejorar la capacidad de monitoreo y respuesta ante fenómenos atmosféricos cambiantes.

RECOMENDACIONES

A partir del análisis realizado mediante modelos de series temporales como SARIMA y SARIMAX, se recomienda considerar la implementación de técnicas más avanzadas como las redes neuronales LSTM (Long Short-Term Memory) para la predicción de variables meteorológicas complejas, como la precipitación. Estas redes son capaces de capturar dependencias temporales no lineales y patrones estacionales prolongados que los modelos tradicionales no logran modelar con precisión. Por ejemplo, Sarwade, Shinde y Patil (2024) demostraron que los modelos LSTM redujeron significativamente el error cuadrático medio (MSE) en la predicción de lluvias extremas, pasando de 374.24 en ARIMA a solo 92.18 en LSTM.

Asimismo, es indispensable mejorar la calidad y continuidad de los registros de datos atmosféricos. Se identificó durante el análisis que la falta de datos completos o con una periodicidad constante afecta negativamente el rendimiento de los modelos predictivos. Esta limitación también fue abordada por estudios recientes, donde la calidad del dataset (incluyendo su normalización, limpieza y estructuración en secuencias temporales) fue determinante para lograr resultados precisos con LSTM (International Journal of Innovative Science and Research Technology, 2024). Se recomienda, por tanto, fortalecer la infraestructura de monitoreo meteorológico, promoviendo bases de datos históricas completas y estandarizadas que incluyan variables clave como temperatura, humedad, presión y radiación UV.

También, se sugiere fomentar la capacitación del personal técnico y científico en metodologías de aprendizaje profundo aplicadas a la climatología. El dominio de estas técnicas no solo permitiría mejorar las predicciones, sino también fortalecer la toma de decisiones basadas en evidencia. Podría potenciarse el uso de estos modelos en sistemas de alerta temprana y políticas públicas de prevención climática. Un estudio del Rochester Institute of Technology (2024) confirmó que los modelos LSTM redujeron el error absoluto medio (MAE) en un 32 % en comparación con SARIMA, mostrando su utilidad práctica en la predicción meteorológica operativa.

Finalmente, se recomienda replicar este tipo de estudios en distintas regiones y observatorios del país. Esta estrategia permitiría comparar comportamientos climáticos en distintas zonas geográficas, mejorar la generalización de los modelos, y ajustar políticas regionales específicas frente al cambio climático.

BIBLIOGRAFÍA

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control*. Wiley.
<http://repo.darmajaya.ac.id/4781/1/Time%20Series%20Analysis%20Forecasting%20and%20Control%28%20PDFDrive%20%29.pdf>
- Colque Candia, W. (2020-09-11). *Aplicación de series temporales en el estudio de precipitaciones en la ciudad del Cusco, 1964 – 2018*. Universidad Andina del Cusco.
<https://hdl.handle.net/20.500.12557/3572>
- Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC). (1992).
Convención Marco de las Naciones Unidas sobre el Cambio Climático. Naciones Unidas.
<https://unfccc.int/resource/docs/convkp/convsp.pdf>
- Cort J., W., & Matsuura, K. (2005). *Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance*. (Vol. Vol. 30). Climate research. <https://www.int-res.com/articles/cr2005/30/c030p079.pdf>
- De la Cruz Córdoba, I. P., & Fernández Carrión, K. L. (2024). *Pronóstico de precipitaciones en Baños, Ecuador mediante técnicas de machine learning: análisis del período 2014 – 2023*. Universidad Central del Ecuador. <https://www.dspace.uce.edu.ec/handle/25000/34505>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
<https://otexts.com/fpp2/>
- International Journal of Innovative Science and Research Technology. (2024, julio). *Prediction of daily climate using long short-term memory (LSTM)*.
<https://ijisrt.com/assets/upload/files/IJISRT24JUL073.pdf>
- James, Witten, Hastie, & Tibshirani. (2017). *An Introduction to Statistical Learning: with Applications in R* (2nd ed. ed.). Springer. <https://www.statlearning.com/>
- Rochester Institute of Technology. (2024). *Rainfall prediction using machine learning methods*. RIT.
<https://repository.rit.edu/theses/12028/>

Sánchez San Román, F. J. (2020). *Precipitaciones*. Dpto. Geología. Univ. Salamanca.

<https://hidrologia.usal.es/temas/Precipitaciones.pdf>

Sánchez Vega, M. V. (2016). *La capa de ozono*. Universidad Estatal a Distancia, Costa Rica Rica.

<https://revistas.uned.ac.cr/index.php/biocenosis/article/view/1276>

Sarwade, M. P., Shinde, S. A., & Patil, V. S. (2024). Predictive modeling of extreme weather forecasting events: An LSTM approach. Current World Environment

<http://dx.doi.org/10.12944/CWE.19.1.17>

Senamhi. (2015). *El club de los protectores contra la radiación solar UV*. Senamhi PERÚ.

https://issuu.com/senamhi_peru/docs/03-historieta_uv

Senamhi. (2024, 25 de abril). *El Niño Costero 2023-2024 fue el más intenso de los últimos 20 años en el oeste de Sudamérica*. Gob.pe.

<https://www.gob.pe/institucion/senamhi/noticias/944204-el-nino-costero-2023-2024-fue-el-mas-intenso-de-los-ultimos-20-anos-en-el-oeste-de-sudamerica>

Senamhi. (2025, 3 de febrero). *El OVA Marcapomacocha del Senamhi recibe la categoría de estación “regional” y entra al Programa de Vigilancia Atmosférica Global de la OMM*. Gob.pe.

<https://www.gob.pe/institucion/senamhi/noticias/1101845-el-ova-marcapomacocha-del-senamhi-recibe-la-categoría-de-estacion-regional-y-entra-al-programa-de-vigilancia-atmosferica-global-de-la-omm>

Turkey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

<https://consoleflare.com/blog/wp-content/uploads/2022/09/Exploratory-Data-Analysis-1977-John-Tukey.pdf>

ANEXO

ANÁLISIS DESCRIPTIVO

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.impute import KNNImputer

import pandas as pd
from google.colab import drive
drive.mount('/content/drive')

df = pd.read_csv('/content/drive/MyDrive/Tesis/Variables.csv')

df['FECHA'] = pd.to_datetime(df['FECHA'], format='%Y%m%d')
df['FECHA']

import missingno as msno
msno.matrix(df[['O3', 'UV_MAX', 'PP']])

# ----- Imputación para UV_MAX con KNN -----
cols_for_knn = ['O3', 'UV_MAX', 'PP']
df_knn = df[cols_for_knn]

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaled_data = scaler.fit_transform(df_knn)

imputer = KNNImputer(n_neighbors=3)
imputed_data = imputer.fit_transform(scaled_data)

imputed_data = scaler.inverse_transform(imputed_data)
df_imputed = pd.DataFrame(imputed_data, columns=cols_for_knn, index=df.index)
df['O3'] = df_imputed['O3']
df['UV_MAX'] = df_imputed['UV_MAX']
df['PP'] = df_imputed['PP']

correlation_matrix = df[['O3', 'UV_MAX', 'PP']].corr()
print("\nMatriz de Correlación:")
print(correlation_matrix)

plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)
plt.title('Matriz de Correlación entre O3, UV_MAX y PP')
plt.show()

df.set_index('FECHA', inplace=True)

for var in ['O3', 'UV_MAX', 'PP']:
    plt.figure()
    sns.lineplot(data=df, x=df.index, y=var)
    plt.title(f'{var} a lo largo del tiempo')
    plt.xlabel('Fecha')
    plt.ylabel(var)
    plt.show()
```

```

variables = ['O3', 'UV_MAX', 'PP']

for var in variables:
    plt.figure(figsize=(10, 4))
    sns.histplot(df[var], kde=True, bins=30, color='skyblue')
    plt.title(f'Distribución de {var}')
    plt.xlabel(var)
    plt.ylabel('Frecuencia')
    plt.grid(True)
    plt.show()

for var in variables:
    plt.figure(figsize=(8, 3))
    sns.boxplot(x=df[var])
    plt.title(f'Boxplot de {var}')
    plt.grid(True)
    plt.show()

colors = {'O3': 'darkgreen', 'UV_MAX': 'darkorange', 'PP': 'royalblue'}

df[['O3', 'UV_MAX', 'PP']].plot(
    subplots=True,
    figsize=(14, 8),
    title='Evolución Temporal de Variables Ambientales',
    color=[colors[col] for col in ['O3', 'UV_MAX', 'PP']]
)

plt.xlabel("Fecha")
plt.tight_layout()

plt.show()

df['month'] = df.index.month

import calendar
df['month_name'] = df['month'].apply(lambda x: calendar.month_abbr[x])

sns.set(style="whitegrid")

fig, axes = plt.subplots(3, 1, figsize=(14, 12), sharex=True)

variables = ['O3', 'UV_MAX', 'PP']
titles = ['Distribución mensual del Ozono (O3)',
          'Distribución mensual del Índice UV Máximo (UV_MAX)',
          'Distribución mensual de la Precipitación (PP)']

for ax, var, title in zip(axes, variables, titles):
    sns.boxplot(x='month_name', y=var, data=df, ax=ax)
    ax.set_title(title, fontsize=14)
    ax.set_xlabel("Mes")
    ax.set_ylabel(var)

plt.tight_layout()
plt.show()

from statsmodels.tsa.seasonal import seasonal_decompose

for col in ['O3', 'UV_MAX', 'PP']:
    decomposition = seasonal_decompose(df[col], model='additive', period=365)

    decomposition.plot()
    plt.suptitle(f'Descomposición estacional de {col}', fontsize=14)
    plt.tight_layout()
    plt.show()

```

```

from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

adf_result_o3 = adfuller(df['O3'])
print(f"ADF Statistic: {adf_result_o3[0]}")
print(f"p-value: {adf_result_o3[1]}")

if adf_result_o3[1] < 0.05:
    print("La serie es estacionaria.")
else:
    print("La serie no es estacionaria.")

adf_result_uv = adfuller(df['UV_MAX'])
print(f"ADF Statistic: {adf_result_uv[0]}")
print(f"p-value: {adf_result_uv[1]}")

if adf_result_uv[1] < 0.05:
    print("La serie es estacionaria.")
else:
    print("La serie no es estacionaria.")

adf_result_pp = adfuller(df['PP'])
print(f"ADF Statistic: {adf_result_pp[0]}")
print(f"p-value: {adf_result_pp[1]}")

if adf_result_pp[1] < 0.05:
    print("La serie es estacionaria.")
else:
    print("La serie no es estacionaria.")

plot_acf(df['O3'], lags=3287)
plt.title("ACF - Autocorrelación")
plt.show()

plot_acf(df['UV_MAX'], lags=3287)
plt.title("ACF - Autocorrelación")
plt.show()

plot_acf(df['PP'], lags=3287)
plt.title("ACF - Autocorrelación")
plt.show()

df = df.sort_index()

start_date = '2020-06-01'
dfs = df.loc[start_date:]

dfs.index.freq = 'D'

dfs = dfs[['O3', 'UV_MAX', 'PP']].dropna()

from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

train_start = '2020-06-01'
train_end = '2022-05-31'
test_start = '2022-06-01'
test_end = '2023-05-31'

df_train = dfs.loc[train_start:train_end]
df_test = dfs.loc[test_start:test_end]

print(f"Train shape: {df_train.shape}")
print(f"Test shape: {df_test.shape}")

df_train.plot(figsize=(12, 5), title="Serie temporal - Entrenamiento (Train)", ylabel="Valor")
plt.grid(True)
plt.show()

result = seasonal_decompose(df_train['O3'], model='additive', period=365)

result.plot()
plt.show()

```

```

result = seasonal_decompose(df_train['UV_MAX'], model='additive', period=365)
result.plot()
plt.show()

result = seasonal_decompose(df_train['PP'], model='additive', period=365)
result.plot()
plt.show()

adf_result = adfuller(df_train['O3'])
print(f"ADF Statistic: {adf_result[0]}")
print(f"p-value: {adf_result[1]}")

if adf_result[1] < 0.05:
    print("La serie es estacionaria.")
else:
    print("La serie no es estacionaria.")

adf_result = adfuller(df_train['UV_MAX'])
print(f"ADF Statistic: {adf_result[0]}")
print(f"p-value: {adf_result[1]}")

if adf_result[1] < 0.05:
    print("La serie es estacionaria.")
else:
    print("La serie no es estacionaria.")

adf_result = adfuller(df_train['PP'])
print(f"ADF Statistic: {adf_result[0]}")
print(f"p-value: {adf_result[1]}")

if adf_result[1] < 0.05:
    print("La serie es estacionaria.")
else:
    print("La serie no es estacionaria.")

plot_acf(df_train['O3'], lags=50)
plt.title("ACF - Autocorrelación")
plt.show()

plot_pacf(df_train['O3'], lags=50)
plt.title("PACF - Autocorrelación parcial")
plt.show()

plot_acf(df_train['UV_MAX'], lags=50)
plt.title("ACF - Autocorrelación")
plt.show()

plot_pacf(df_train['UV_MAX'], lags=50)
plt.title("PACF - Autocorrelación parcial")
plt.show()

plot_acf(df_train['PP'], lags=50)
plt.title("ACF - Autocorrelación")
plt.show()

plot_pacf(df_train['PP'], lags=50)
plt.title("PACF - Autocorrelación parcial")
plt.show()

df_train.to_csv('df_trainv2.csv')
df_test.to_csv('df_testv2.csv')

```

MODELADO PREDICTIVO

```
import pandas as pd

df_train = pd.read_csv('df_trainv2.csv')
df_test = pd.read_csv('df_testv2.csv')
df_train = df_train[['O3']]

from statsmodels.tsa.statespace.sarimax import SARIMAX

modelo3 = SARIMAX(df_train['O3'],
                   order=(1,0,1),
                   seasonal_order=(1,1,1,91),
                   enforce_stationarity=False,
                   enforce_invertibility=False)

resulto3 = modelo3.fit()
print(resulto3.summary())
predicciones = resulto3.get_forecast(steps=365)

pred_mean = predicciones.predicted_mean
conf_int = predicciones.conf_int()

pred_mean.index = df_test.index
conf_int.index = df_test.index

df_resultado = df_test.copy()
df_resultado['O3_Pronosticado'] = pred_mean
df_resultado['IC_inf'] = conf_int.iloc[:, 0]
df_resultado['IC_sup'] = conf_int.iloc[:, 1]

df_resultado

import matplotlib.pyplot as plt

plt.figure(figsize=(12, 6))
plt.plot(df_test['O3'], label='O3 real', color='black')
plt.plot(df_resultado['O3_Pronosticado'], label='O3 predicho', color='blue')
plt.title('Predicción vs Real - O3')
plt.legend()
plt.show()

pred = resulto3.get_forecast(steps=91)
conf_int = pred.conf_int()

plt.figure(figsize=(10,5))
plt.plot(df_train.index, df_train['O3'], label='Observado')
plt.plot(pred.predicted_mean.index, pred.predicted_mean, label='Pronóstico', color='red')
plt.fill_between(conf_int.index, conf_int.iloc[:, 0], conf_int.iloc[:, 1], color='pink', alpha=0.3)
plt.legend()
plt.title('Pronóstico de O3 con SARIMAX')
plt.show()

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

y_pred = df_resultado['O3_Pronosticado']
y_true = df_test['O3']

mae = mean_absolute_error(y_true, y_pred)
rmse = np.sqrt(mean_squared_error(y_true, y_pred))
r2 = r2_score(y_true, y_pred)

mean_real = y_true.mean()
mae_rel = mae / mean_real * 100

print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R2: {r2:.2f}")
print(f"MAE Relativo: {mae_rel:.2f}%")
```

```

import pandas as pd

df_train = pd.read_csv('df_trainv2.csv')
df_test = pd.read_csv('df_testv2.csv')

df_train = df_train[['UV_MAX']]
from statsmodels.tsa.statespace.sarimax import SARIMAX

model = SARIMAX(df_train['UV_MAX'],
                 order=(1,0,1),
                 seasonal_order=(1,1,1,91),
                 enforce_stationarity=False,
                 enforce_invertibility=False)

result = model.fit()
print(result.summary())
predicciones = result.get_forecast(steps=365)

pred_mean = predicciones.predicted_mean
conf_int = predicciones.conf_int()

pred_mean.index = df_test.index
conf_int.index = df_test.index

df_resultado = df_test.copy()
df_resultado['UV_MAX_Pronosticado'] = pred_mean
df_resultado['IC_inf'] = conf_int.iloc[:, 0]
df_resultado['IC_sup'] = conf_int.iloc[:, 1]

df_resultado
import matplotlib.pyplot as plt

plt.figure(figsize=(12, 6))
plt.plot(df_test['UV_MAX'], label='UV_MAX real', color='black')
plt.plot(df_resultado['UV_MAX_Pronosticado'], label='UV_MAX predicho', color='blue')
plt.title('Predicción vs Real - UV_MAX')
plt.legend()
plt.show()

pred = result.get_forecast(steps=91)
conf_int = pred.conf_int()

plt.figure(figsize=(10,5))
plt.plot(df_train.index, df_train['UV_MAX'], label='Observado')
plt.plot(pred.predicted_mean.index, pred.predicted_mean, label='Pronóstico', color='red')
plt.fill_between(conf_int.index, conf_int.iloc[:, 0], conf_int.iloc[:, 1], color='pink', alpha=0.3)
plt.legend()
plt.title('Pronóstico de UV_MAX con SARIMAX')
plt.show()

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

y_pred = df_resultado['UV_MAX_Pronosticado']
y_true = df_test['UV_MAX']

mae = mean_absolute_error(y_true, y_pred)
rmse = np.sqrt(mean_squared_error(y_true, y_pred))
r2 = r2_score(y_true, y_pred)

mean_real = y_true.mean()
mae_rel = mae / mean_real * 100

print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R2: {r2:.2f}")
print(f"MAE Relativo: {mae_rel:.2f}%")

```

```

import pandas as pd

df_train = pd.read_csv('df_trainv2.csv')
df_test = pd.read_csv('df_testv2.csv')

df_train = df_train[['PP']]

from statsmodels.tsa.statespace.sarimax import SARIMAX

model = SARIMAX(df_train['PP'],
                 order=(1,0,1),
                 seasonal_order=(1,1,1,91),
                 enforce_stationarity=False,
                 enforce_invertibility=False)

results = model.fit()
print(results.summary())

predicciones = results.get_forecast(steps=365)

pred_mean = predicciones.predicted_mean.clip(lower=0)
conf_int = predicciones.conf_int()
conf_int.iloc[:, 0] = conf_int.iloc[:, 0].clip(lower=0)

pred_mean.index = df_test.index
conf_int.index = df_test.index

df_resultado = df_test.copy()
df_resultado['PP_Pronosticado'] = pred_mean
df_resultado['IC_inf'] = conf_int.iloc[:, 0]
df_resultado['IC_sup'] = conf_int.iloc[:, 1]

df_resultado

import matplotlib.pyplot as plt

plt.figure(figsize=(12, 6))
plt.plot(df_test['PP'], label='PP real', color='black')
plt.plot(df_resultado['PP_Pronosticado'], label='PP predicho', color='blue')
plt.title('Predicción vs Real - PP')
plt.legend()
plt.show()

import matplotlib.pyplot as plt

pred = results.get_forecast(steps=91)
conf_int = pred.conf_int()
conf_int.iloc[:, 0] = conf_int.iloc[:, 0].clip(lower=0)

plt.figure(figsize=(10,5))
plt.plot(df_train.index, df_train['PP'], label='Observado')
plt.plot(pred.predicted_mean.index, pred.predicted_mean.clip(lower=0), label='Pronóstico', color='red')
plt.fill_between(conf_int.index, conf_int.iloc[:, 0], conf_int.iloc[:, 1], color='pink', alpha=0.3)
plt.legend()
plt.title('Pronóstico de PP con SARIMAX')
plt.show()

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

y_pred = df_resultado['PP_Pronosticado']
y_true = df_test['PP']

mae = mean_absolute_error(y_true, y_pred)
rmse = np.sqrt(mean_squared_error(y_true, y_pred))
r2 = r2_score(y_true, y_pred)

mean_real = y_true.mean()
mae_rel = mae / mean_real * 100

print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R2: {r2:.2f}")
print(f"MAE Relativo: {mae_rel:.2f}%")

```